Introduction to Semidefinite Programming: Motivation and Duality

Henry Wolkowicz*

June 18, 2001

*Department of Combinatorics & Optimization, University of Waterloo

Outline

- Basic Properties and Notation
- Examples
- Duality

1 Basic Properties and Notation

Basic linear **Semidefinite Programming** looks just like Linear Programming

$$
\textbf{(PSDP)} \qquad \begin{array}{rlll} p^* = & \max & \operatorname{trace} CX & (\langle C, X \rangle) \\ & \text{s.t.} & \mathcal{A}X = b & \text{(linear)} \\ & & X \succeq 0, \quad (X \in \mathcal{P}) & \text{(nonneg)} \end{array}
$$

$$
\mathcal{A}(\alpha X_1 + \beta X_2) = \alpha \mathcal{A}(X_1) + \beta \mathcal{A}(X_2), \quad \forall X_i, \text{ and scalars } \alpha, \beta
$$

$\mathcal{S}^n$:= space of $n \times n$ real symmetric matrices, $(A = A^T)$
(Hermitian?)

$\preceq$ denotes the Löwner partial order
$A \preceq B$ if $B - A \succeq 0$ (positive semidefinite)

For $A \in \mathcal{S}^n$:
A is positive semidefinite (positive definite), (denoted $A \succeq 0$ $(A \succ 0)$,)
if $x^T A x \geq 0 (> 0), \ \forall x \neq 0$.

For $M, n \times n$, $Mx = \lambda x$ (eigenvector-eigenvalue equation, simplest linear operation is multiplication by scalar), equivalently,
$\det(M - \lambda I) = 0$.

TFAE:

1. $A \succeq 0 \quad (A \succ 0)$
2. the eigenvalues $\lambda(A) \geq 0 \quad (\lambda(A) > 0)$
3. $A = R^T R$, for some R (Cholesky factorization if R is upper triangular - Gram matrix representation of n columns of R, $a_{ij} = u_i^T u_j$, R nonsingular if $A \succ 0$)
4. $A = \sum_i \alpha_i v_i v_i^T$ for some nonnegative α_i (sum of rank ones, $A \succ 0$??)
5. all principal minors (determinants of principal submatrices) are nonnegative (positive for $A \succ 0$, leading princ. minors??) (principal submatrix: delete some rows and corresp. cols.; history-Sylvester?)
6. $S = (S^{\frac{1}{2}})^2$, for some $S^{\frac{1}{2}} \succeq 0 \quad (\succ 0)$

For a general square matrix $M, n \times n$, define $\operatorname{trace} M = \sum_i M_{ii}$ Then

$$\operatorname{trace} M = \sum_i \lambda_i \quad \text{(trace is sum of eigenvalues)}$$

Theorem 1.0.0.1 *(Interlacing/Majorization Eigenvalues, Fan, 1954) Let $H, \bar{H}$ be an $n \times n$ Hermitian matrices of the form*

$$H = \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix}, \qquad \bar{H} = \begin{pmatrix} H_{11} & 0 \\ 0 & H_{22} \end{pmatrix},$$

with H_{11} size $l \times l$. Then

$$(\lambda(H_{11}), \lambda(H_{22})) = \lambda(\bar{H}) \prec_M \lambda(H), \quad \text{on } \Re^n,$$

where $\prec_M$ denotes majorization, i.e. $\lambda \prec_M \mu$ if
$\sum_{i=1}^{k} \lambda_{[i]} \leq \sum_{i=1}^{k} \mu_{[i]}$, with equality for $k = n$. ([.] denotes nonincreasing order). In particular,
$\lambda_j(H) \geq \lambda_j(H_{11}), \quad \lambda_{n-j+1}(H) \leq \lambda_{l-j+1}(H_{11}).$ ■

1.0.1 Further Properties and Definitions

- For matrices P, Q compatible for multiplication,

$$\operatorname{trace} PQ = \operatorname{trace} QP.$$

- Every symmetric matrix S has an orthogonal decomposition $S = Q\Lambda Q^T$, where Λ is diagonal with the eigenvalues of S on the diagonal, and Q has orthonormal columns consisting of eigenvectors of S.

- $X \succeq 0 \Rightarrow P^T X P \succeq 0 \quad$ (congruence)

Further properties of $A \succeq 0$:

• $A = \sum_{i=1}^{r} \alpha_i v_i v_i^T$ for minimum number of nonnegative α_i characterizes $\operatorname{rank} A$;

• $\operatorname{diag}(A) \geq 0$; and $a_{ii} = 0$ implies i-th and i-th col are 0;

• $\operatorname{trace} A \geq 0$ and $\operatorname{trace} A = 0$ iff $A = 0$;

• $\mathcal{P} \subset \mathcal{S}^n$ is a closed convex cone, i.e. $\mathcal{P} \subset \mathcal{P} + \mathcal{P}; \alpha\mathcal{P} \subset \mathcal{P}, \forall \alpha \geq 0$;

• $A, B \succeq 0$ implies $\operatorname{trace} AB \geq 0$ with equality iff $AB = 0$ (proof?).

• geometry of $\mathcal{P}$ in $\mathcal{S}^n$ equipped with trace inner product: all vectors within 45 degrees of identity matrix;

• $\mathcal{P}$ is a *self-polar* cone (proof??), $\mathcal{P} = \mathcal{P}^*$, where

$$C^* = \{\phi \in \Re^n : \langle \phi, x \rangle \geq 0, \forall x \in C\}.$$

Lemma 1.0.1.1 (Schur Complement) *Let* $A = \begin{pmatrix} B & C^T \\ C & D \end{pmatrix}$ *be a symmetric matrix with the block* $B \succ 0$*. Then*

$$A \succ 0 \ (\succeq 0) \ \textit{iff} \ D - CB^{-1}C^T \succ 0 \ (\succeq 0).$$

Theorem 1.0.1.1 *(Perron-Frobenius) If an $n \times n$ matrix has nonnegative entries then it has a nonegative real eigenvalue λ which has maximum absolute value among all eigenvalues. This eigenvalue λ has a nonnegative real eigenvector. If, in addition, the matrix has no block-triangular decomposition (is irreducible - does not contain a $k \times (n-k)$ block of 0-s disjoint from the diagonal), then λ has multiplicity 1 and the corresponding eigenvector is positive.* ■

The linear operator $\mathcal{A}$ and its adjoint:

$$\mathcal{A} : \mathcal{S}^n \to \mathbb{R}^m, \qquad \mathcal{A}^* : \mathbb{R}^m \to \mathcal{S}^n$$

$$\text{for given } A_i \in \mathcal{S}^n, i = 1, \ldots m :$$

$$(\mathcal{A}X)_i = \operatorname{trace}(A_i X), \quad \mathcal{A}^* y = \textstyle\sum_{i=1}^m y_i A_i$$

where

$$\langle \mathcal{A}X, y \rangle = \langle X, \mathcal{A}^* y \rangle, \quad \forall X, y$$

$\mathcal{P}$ - cone of positive semidefinite matrices

replaces

$\mathbb{R}^n_+$ - nonnegative orthant

$\operatorname{trace} CX = \langle C, X \rangle = \sum_{i=1}^n \sum_{j=1}^n C_{ij} X_{ij}$

SDP is equivalent to:

$$
\textbf{(PSDP)} \qquad
\begin{array}{rcl}
p^* = & \max & \sum_{i=1}^n \sum_{j=1}^n C_{ij} X_{ij} \\
 & \text{s.t.} & \langle A_i, X \rangle = b_i, \qquad i = 1, \ldots m \\
 & & X \succeq 0, \quad (X \in \mathcal{P})
\end{array}
$$

1.0.2 Duality (like LP)

payoff function, player Y to player X (Lagrangian)

$$L(X,y) := \operatorname{trace}(CX) + y^T(b - \mathcal{A}X)$$

Optimal (worst case) strategy for player X:

$$p^* = \max_{X \succeq 0} \min_y L(X,y)$$

Using the *hidden constraint* $b - \mathcal{A}X = 0$, recovers primal problem (PSDP).

$$
\begin{aligned}
L(X,y) &= \operatorname{trace}(CX) + y^T(b - \mathcal{A}X) \\
&= b^T y + \operatorname{trace}\,(C - \mathcal{A}^* y)\, X
\end{aligned}
$$

adjoint operator, $\mathcal{A}^* y = \sum_i y_i A_i$

$$
\langle \mathcal{A}^* y, X \rangle = \langle y, \mathcal{A}X \rangle, \quad \forall X, y
$$

$$
p^* = \max_{X \succeq 0} \min_y L(X,y) \le d^* := \min_y \max_{X \succeq 0} L(X,y)
$$

For dual, use *hidden constraint* $g(y) := C - \mathcal{A}^* y \preceq 0$

$$p^* = \max_{X \succeq 0} \min_y L(X,y) \leq d^* := \min_y \max_{X \succeq 0} L(X,y)$$

dual obtained from optimal strategy of competing player Y;
use *hidden constraint* $g(y) = C - \mathcal{A}^* y \preceq 0$

$$(\textbf{DSDP}) \qquad \begin{array}{rcc} d^* = & \min & b^T y \\ & \text{s.t.} & \mathcal{A}^* y \succeq C \end{array}$$

for the primal

$$(\textbf{PSDP}) \qquad \begin{array}{rcc} p^* = & \max & \operatorname{trace} CX \\ & \text{s.t.} & \mathcal{A}X = b \\ & & X \succeq 0 \end{array}$$

1.0.3 Weak Duality - Optimality

Proposition 1.0.3.1 *(Weak Duality) If X feas. in (PSDP), y feas. in (DSDP), $Z = C - \mathcal{A}^* y \succeq 0$ is slack variable, then*

$$\textit{duality gap} \; := b^T y - \operatorname{trace} CX = \operatorname{trace} XZ \geq 0.$$

Proof. (Direct - using
$\operatorname{trace} ZX = \operatorname{trace} X^{1/2} X^{1/2} Z = \operatorname{trace} X^{1/2} Z X^{1/2} \geq 0$)

$$\begin{array}{rcl} \operatorname{trace} CX - b^T y & = & \operatorname{trace} (\mathcal{A}^* y - Z) X - b^T y \\ & = & \operatorname{trace} y^T \mathcal{A} X - \operatorname{trace} ZX - b^T y \\ & = & \operatorname{trace} y^T (\mathcal{A} X - b) - \operatorname{trace} ZX = -\operatorname{trace} ZX, \end{array}$$

■

Characterization of optimality
(in the case of a zero duality gap at optimality and attainment)
for the dual pair X, y (slack $Z \succeq 0$)

$$\mathcal{A}^* y - Z = C \qquad \text{dual feasibility}$$

$$\mathcal{A} X = b \qquad \text{primal feasibility}$$

$$ZX = 0 \qquad \text{complementary slackness (equiv. } \operatorname{trace} ZX = 0)$$

$$ZX = \mu I \qquad \text{perturbed}$$

Forms the basis for:

interior point methods
(primal simplex method, dual simplex method)

1.0.4 Preliminary Examples

Example 1.0.4.1 Minimizing the Maximum Eigenvalue

Arises in e.g. stability of differential equations

- *The mathematical problem:*
  - ◇ *given* $A(x)$ *depending linearly on vector* x
  - ◇ *Find* x *to minimize the maximum eigenvalue of* $A(x)$
- *SDP Model:*
  - ◇ *the largest eigenvalue* $\lambda_{\max}(A(x)) \leq \alpha$ *iff* $\lambda_{\max}(A(x) - \alpha I) \leq 0$ *iff* $A(x) - \alpha I \preceq 0.$
  - ◇ *The DSDP (in dual form) is:*

$$\max -\alpha \qquad s.t. \quad A(x) - \alpha I \preceq 0.$$

HOW DOES SDP arise from quadratic approximations?

$$\text{Let} \quad q_i(y) = \frac{1}{2} y^T Q_i y + y^T b_i + c_i, \; y \in \mathbb{R}^n$$

$$(\mathbf{QQP}) \qquad \begin{cases} q^* = & \min & q_0(y) \\ & \text{s.t.} & q_i(y) = 0 \\ & & i = 1, \ldots m \end{cases}$$

- Lagrangian (quadratic, linear, constant in y terms):

$$L(y,x) = \frac{1}{2} y^T (Q_0 - \sum_{i=1}^m x_i Q_i) y + y^T (b_0 - \sum_{i=1}^m x_i b_i) + (c_0 - \sum_{i=1}^m x_i c_i)$$

- Primal-Dual pair: $q^* = \min\limits_y \max\limits_x L(y,x) \geq d^* = \max\limits_x \min\limits_y L(y,x)$

  - homogenize (add y_0): $\quad y_0 y^T (b_0 - \sum_{i=1}^m x_i b_i), \quad y_0^2 = 1.$

$$\begin{array}{rcll} d^* & = & \max_x \min_y & L(y,x) \\ & = & \max_x \min\limits_{y_0^2=1} & \frac{1}{2} y^T (Q_0 - \sum_{i=1}^m x_i Q_i) y \quad (+t y_0^2) \\ & & & + y_0 y^T (b_0 - \sum_{i=1}^m x_i b_i) \\ & & & + (c_0 - \sum_{i=1}^m x_i c_i) \quad (-t) \end{array}$$

hidden semidefinite constraint yields SDP constraint

$$(\mathcal{A} : \mathbb{R}^{m+1} \to \mathcal{S}_{n+1}) \qquad B - \mathcal{A} \begin{pmatrix} t \\ x \end{pmatrix} \succeq 0.$$

$$B = \begin{pmatrix} 0 & b_0^T \\ b_0 & Q_0 \end{pmatrix}, \quad \mathcal{A} \begin{pmatrix} t \\ x \end{pmatrix} = \begin{bmatrix} -t & \sum_{i=1}^m x_i b_i^T \\ \sum_{i=1}^m x_i b_i & \sum_{i=1}^m x_i Q_i \end{bmatrix}$$

The dual program is equivalent to the SDP (with $c_0 = 0$)

$$
(\mathbf{D}) \qquad \begin{array}{rll} d^* = & \sup & -\sum_{i=1}^m x_i c_i - t \\ & \text{s.t.} & \mathcal{A}\begin{pmatrix} t \\ x \end{pmatrix} \preceq B \\ & & x \in \mathbb{R}^m, t \in \mathbb{R} \end{array}
$$

As in linear programming, the dual is obtained from the optimal strategy of the competing player:

$$
(\mathbf{DD}) \qquad \begin{array}{rll} d^* = & \inf & \operatorname{trace} BU \\ & \text{s.t.} & \mathcal{A}^* U = \begin{pmatrix} -1 \\ -c \end{pmatrix} \\ & & U \succeq 0. \end{array}
$$

Example 1.0.4.2 Pseudoconvex (Nonlinear) Optimization Problem

$$
(\mathbf{PCP}) \qquad \begin{array}{rll} d^* = & \min & \frac{(c^T x)^2}{d^T x} \\ & s.t. & Ax + b \geq 0, \end{array}
$$

(where $Ax + b \geq 0 \Rightarrow d^T x > 0$*) is equivalent to*

$$
\begin{array}{rll} d^* = & \min & t \\ & s.t. & \begin{bmatrix} \mathrm{Diag}\,(Ax+b) & 0 & 0 \\ 0 & t & c^T x \\ 0 & c^T x & d^T x \end{bmatrix} \succeq 0 \end{array}
$$

• ***Alternate Name - Linear Matrix Inequality***

$$g(y) = C - \mathcal{A}^* y \preceq 0$$

adjoint operator $\qquad \mathcal{A}^* y = \sum_i y_i A_i, \quad A_i = A_i^T$

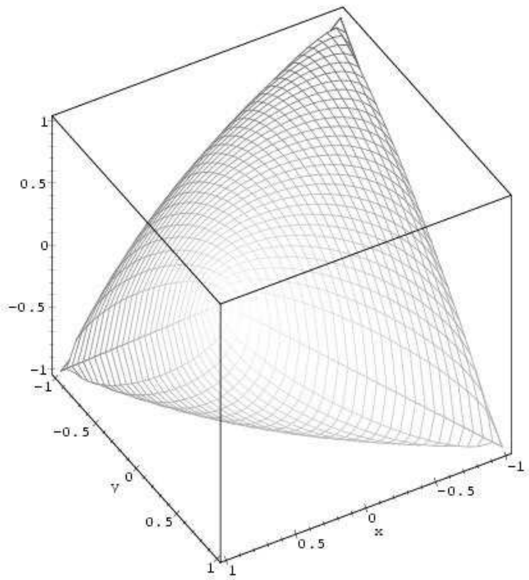
• $g(y)$ is a **convex constraint**, i.e.

$$g(\lambda x + (1-\lambda)y) \preceq \lambda g(x) + (1-\lambda) g(y), \quad \forall 0 \leq \lambda \leq 1, \forall x, y \in \mathbb{R}^m$$

equivalently, for every $P \succeq 0$, the real valued function $f_P(x) := \operatorname{trace} P g(x)$ is a convex function, i.e.

$$f_P(\lambda x + (1-\lambda)y) \leq \lambda f_P(x) + (1-\lambda) f_P(y), \quad \forall 0 \leq \lambda \leq 1, \forall x, y \in \mathbb{R}^m$$

- *feasible set* $\{x \in \mathbb{R}^m : g(x) \preceq 0\}$ is convex set
- *optimum point* is on the boundary ($g(x)$ is singular)
- *boundary of feasible set* is *not* smooth (piecewise smooth, algebraic surface)



1.1 Introduction to Convex Cones and the Löwner Partial Order

Definition 1.1.0.1 *Let* $\alpha \in \mathbb{R}$ *and* $S, T \subset \mathbb{R}^n$*. Then*
$\alpha S = \{y : y = \alpha s, \text{ for some } s \in S\}$ *and*
$S + T = \{y : y = s + t, \text{ for some } s \in S, t \in T\}$

Definition 1.1.0.2 $\mathcal{K} \subset \mathbb{R}^n$ *is a* <u>*cone*</u> *if* $\alpha\mathcal{K} \subset K, \quad \forall \alpha > 0.$

Definition 1.1.0.3 *the cone* $\mathcal{K}$ *is a* <u>*convex cone*</u> *if* $\mathcal{K} + \mathcal{K} \subset \mathcal{K}$.

Definition 1.1.0.4 *A cone* $\mathcal{K}$ *is a* <u>*pointed cone*</u> *if* $\mathcal{K} \cap (-\mathcal{K}) = \{0\}$.

Definition 1.1.0.5 *A cone* $\mathcal{K} \subset \mathbb{R}^n$ *is a* <u>*proper cone*</u> *if it is closed, pointed, and convex and has nonempty interior.*

Examples of Cones

Example 1.1.0.3 *open half line:* $\{x \in \mathbb{R} : x > 0\}$.

Example 1.1.0.4 *closed half line:* $\{x \in \mathbb{R} : x \geq 0\}$.

Example 1.1.0.5 *psd matrices, $\mathcal{P}$:* $\{X \in \mathcal{S}^n : X \succeq 0\}$.

Example 1.1.0.6 *Lorentz cone*
(ice-cream cone, second-order cone):
$L^m = \{x = (x_1, \ldots, x_m) \in \mathbb{R}^m : x_m \geq \sqrt{x_1^2 + \ldots + x_m^2}$.

And, an important closed convex cone is:

Definition 1.1.0.6 *Polar (dual, conjugate) of set $\mathcal{S}$:*
$\mathcal{S}^+ := \{z : \langle x, z\rangle \geq 0, \forall x \in \mathcal{S}\} \quad (= S^*)$.

Example 1.1.0.7 *nonnegative orthant, psd cone $\mathcal{P}$, and Lorentz cone L^m, are all self-polar, i.e. $\mathcal{K} = \mathcal{K}^+$.*

Definition 1.1.0.7 *Direct sum of two cones*
$\mathcal{K} \oplus \mathcal{L} := \{(k\ l) : k \in \mathcal{K}, l \in \mathcal{L}\}$.

Example 1.1.0.8 *Let* L *denote the half line in* $\mathbb{R}$, *then* $\underbrace{L \oplus L \cdots \oplus L}_{n}$ *is the* n*-dimension nonnegative orthant.*

Suppose $\mathcal{K}$, $\mathcal{K}_1$, $\mathcal{K}_2$ are proper cones, then:

- $\mathcal{K}^+$ is proper.
- $(\mathcal{K}^+)^+ = \mathcal{K}$.
- $(\mathcal{K}_1 \cap \mathcal{K}_2)^+ = \overline{\mathcal{K}_1^+ + \mathcal{K}_2^+}$.
- $(\mathcal{K}_1 + \mathcal{K}_2)^+ = \mathcal{K}_1^+ \cap \mathcal{K}_2^+$.
- $\mathcal{K}_1 \oplus \mathcal{K}_2$ is proper.
- $(\mathcal{K}_1 \oplus \mathcal{K}_2)^+ = \mathcal{K}_1^+ \oplus \mathcal{K}_2^+$.

Useful Lemma, e.g. to Prove Farkas' Lemma

Lemma 1.1.0.1

$$\mathcal{K} \textit{ is a closed convex cone} \iff \mathcal{K} = \mathcal{K}^{++}.$$

Partial Orders:

Definition 1.1.0.8 $x \succeq_{\mathcal{K}} y$ *(respectively* $x \succ_{\mathcal{K}} y$*) if* $x - y \in \mathcal{K}$ *(respectively* $x - y \in \operatorname{int} \mathcal{K}$*).*

Remark 1.1.0.1 *If* $\mathcal{K}$ *is a pointed, convex cone, then "*$\succeq_{\mathcal{K}}$*" is a* <u>*(linear) partial order*</u> *(reflexive, transitive, and antisymmetric):*

- $0 \in \mathcal{K} \Rightarrow x \succeq_{\mathcal{K}} x$ *(reflexive);*
- $\mathcal{K}$ *is convex* $\Rightarrow$ *if* $x \succeq_{\mathcal{K}} y$ *and* $y \succeq_{\mathcal{K}} z$*, then* $x \succeq_{\mathcal{K}} z$ *(transitive);*
- $\mathcal{K}$ *is pointed* $\Rightarrow$ *if* $x \succeq_{\mathcal{K}} y$ *and* $y \succeq_{\mathcal{K}} x$*, then* $x = y$ *(antisymmetric);*
- $\mathcal{K}$ *is convex cone* $\Rightarrow$
  if $a, b \geq 0, u \succeq_{\mathcal{K}} x$ *and* $v \succeq_{\mathcal{K}} y$*, then* $au + bv \succeq_{\mathcal{K}} ax + by$ *(linear - homogeneous, additive);*

1.2 Convex Cone Program

Let: $\mathcal{K}, \mathcal{L}$ be convex cones;

f real valued convex function, g is $\mathcal{K}$-convex, i.e.

$g(\alpha u + (1-\alpha)v) \preceq_{\mathcal{K}} \alpha g(u) + (1-\alpha)g(v), \forall 0 \leq \alpha \leq 1, \forall u, v$

$$(CP) \qquad \begin{array}{rc} \mu^* = \min & f(x) \\ \text{s.t.} & g(x) \preceq_{\mathcal{K}} 0 \\ & x \succeq_{\mathcal{L}} 0 \end{array}$$

with Lagrangian dual (weak duality)

$$(DCP) \qquad \mu^* \geq \nu^* = \max_{y \in \mathcal{K}^+} \min_{x \in \mathcal{L}} \quad f(x) + \langle y, g(x) \rangle$$

Example 1.2.0.9 *If* $f(x) = \langle c, x \rangle$, $g(x) = b - \mathcal{A}x$, *then we have a linear cone programming problem, (LCP). The dual*

$$(DLCP) \qquad \begin{aligned} \mu^* \geq \nu^* &= \max_{y \in \mathcal{K}^+} \min_{x \in \mathcal{L}} \quad \langle c, x \rangle + \langle y, b - \mathcal{A}x \rangle \\ &= \max_{y \in \mathcal{K}^+} \min_{x \in \mathcal{L}} \quad \langle c - \mathcal{A}^* y, x \rangle + \langle y, b \rangle \\ &= \max_{\substack{y \in \mathcal{K}^+ \\ c - \mathcal{A}^* y \in \mathcal{L}^+}} \min_{x \in \mathcal{L}} \quad \langle c - \mathcal{A}^* y, x \rangle + \langle y, b \rangle \end{aligned}$$

reduces to the elegant (LP or SDP type) form

$$(DLCP) \qquad \begin{array}{rl} \mu^* \geq \nu^* = \max & \langle y, b \rangle \\ s.t. & \mathcal{A}^* y \preceq_{\mathcal{L}^+} c \\ & y \succeq_{\mathcal{K}^+} 0 \end{array}$$

$$
(LCP) \qquad \begin{array}{rl} \mu^* = \min & \langle c, x \rangle \\ \text{s.t.} & \mathcal{A}x \succeq_{\mathcal{K}} b \\ & x \succeq_{\mathcal{L}} 0 \end{array}
$$

$$
(DLCP) \qquad \begin{array}{rl} \mu^* \geq \nu^* = \max & \langle y, b \rangle \\ \text{s.t.} & \mathcal{A}^* y \preceq_{\mathcal{L}^+} c \\ & y \succeq_{\mathcal{K}^+} 0 \end{array}
$$

Theorem 1.2.0.1 *Suppose that Slater's CQ holds for (LCP), i.e.* $\exists \hat{x}$ *such that* $\mathcal{A}\hat{x} \succ_{\mathcal{K}} b, \quad \hat{x} \succ_{\mathcal{L}} 0$. *Then* strong duality *holds for (LCP),(DLCP), i.e. there exists* y^* *feasible for (DLCP) such that*

$$
\mu^* = \nu^* = \langle y^*, b \rangle .
$$

Example 1.2.0.10 *If the primal is*

$$(\mathbf{P}) \qquad \begin{array}{rll} p^* = & \sup & x_2 \\ & s.t. & \begin{bmatrix} x_2 & 0 & 0 \\ 0 & x_1 & x_2 \\ 0 & x_2 & 0 \end{bmatrix} \preceq \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{array}$$

$$S^f = \\ cone \left\{ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right\} \quad \textit{(minimal cone containing feasible set)}$$

Then the dual is

$$(\mathbf{D}) \qquad \begin{array}{rlc} d^* = & \inf & U_{11} \\ & s.t. & U_{22} = 0 \\ & & U_{11} + 2U_{23} = 1 \\ & & U \succeq 0. \end{array}$$

Then $p^* = 0 < d^* = 1$.

————————————–

(But regularized dual (below) has $U \succeq_{(S^f)^+} 0$, *i.e. only constraint is* $U_{11} \geq 0$. *So new dual optimal value is 0. (Attained.))*

1.3 Optimality Conditions without constraint qualifications

1.3.1 Facial Structure

Definition 1.3.1.1 *$\mathcal{F}$ is a* face *of a convex cone $\mathcal{K}$ (denoted $\mathcal{F} \lhd \mathcal{K}$) if $\mathcal{F} \subset \mathcal{K}$ and*

$$\frac{1}{2}(x_1 + x_2) \in \mathcal{F} \Rightarrow x_1, x_2 \in \mathcal{F}, \qquad \forall x_1, x_2 \in \mathcal{K}$$

Definition 1.3.1.2 *A face $\mathcal{F} \lhd \mathcal{K}$ is called* exposed *if*

$$\mathcal{F} = \mathcal{K} \cap \phi^{\perp}, \qquad \text{for some } \phi \in \mathcal{K}^{+}$$

Definition 1.3.1.3 *A face $\mathcal{F} \lhd \mathcal{P}$ is called* projectionally exposed *if*

$$\mathcal{F} = Q\mathcal{P}Q^{t} \qquad \text{for some matrix } Q$$

1.3.2 Weak Duality

The weak duality: $c^T x \geq b^T y$ holds in Cone-LP:

$$c^T x - b^T y = c^T x - (Ax)^T y = x^T (c - A^T y) = x^T s \geq 0$$

1.3.3 Extended Farkas' Lemma

Theorem 1.3.3.1 ***(Seperation Theorem)*** *If $K \subseteq \mathbb{R}^n$ is closed and convex, $b \in \mathbb{R}^n$, $b \notin K$, then $\exists a \in \mathbb{R}^n$, such that* $\forall x \in K : a^T b < 0$ *and* $a^T x \geq 0$.

Lemma 1.3.3.1 ***(Farkas' Lemma)*** *Let $\mathcal{K} \subseteq \mathbb{R}^n$ be a closed, convex cone, $A \in \mathbb{R}^{m \times n}, A(\mathcal{K})$ closed, and $b \in \mathbb{R}^m$. Then:*

$$\{\exists x \succeq_{\mathcal{K}} 0 : Ax = b\} \iff \left\{A^T y \succeq_{\mathcal{K}^+} 0 \Rightarrow b^T y \geq 0\right\}.$$

Proof. (Necessity:) Suppose $\exists x \in \mathcal{K} : Ax = b$. And suppose $A^T y \succeq_{\mathcal{K}^+} 0$. Then $b^T y = (Ax)^T y = x^T (A^T y) \geq 0$, since $x \succeq_K 0$.

(Sufficiency:) Suppose that the RHS of (1.3.3.1) holds but the LHS fails, i.e. $b \notin A(\mathcal{K})$. Then separate, i.e. $\exists y \in \mathbb{R}_m$ such that $b^T y < 0$, and $\forall x \in \mathcal{K}, (Ax)^T y \geq 0$. Therefore, $A^T y \in K^+$ which implies $b^T y \geq 0$, contradiction. ∎

Cone programming: S a closed convex cone induces a linear partial order ([11])

$$\min\{f(x) : g(x) \preceq_S 0,\ x \in \Omega\}$$

feasible set A

——————————

rewrite using minimal face $g(A) \subset -S^f$

$$\min\{f(x) : g(x) \preceq_{S^f} 0,\ x \in \Omega^f\}$$

where $\Omega^f = \Omega \cap g^{-1}(S^f - S^f) = \Omega \cap g^{-1}(S^f - S)$

——————————

optimality conditions for some $\Lambda \in (S^f)^+$:

$$f(x) + \Lambda g(x) \geq \mu^*,\ \forall x \in \Omega^f$$

Special case of linear cone programming, $\mathcal{K}, \mathcal{L}$ closed convex cones ([85]):

primal:

$$\mu^* = \min\{cx : \mathcal{A}x \succeq_{\mathcal{K}} 0,\ x \succeq_{\mathcal{L}} 0\}$$

minimal cones $\mathcal{K}^f, \mathcal{L}^f$

dual:

$$\mu^* = \nu^* = \max\{by : \mathcal{A}^* y \preceq_{(\mathcal{L}^f)^+} b,\ y \succeq_{(\mathcal{K}^f)^+} 0\}$$

Linear Programming ($\mathcal{A} : \mathbb{R}^m \to \mathbb{R}^n$)

$\mathbb{R}^n_+$ is a closed convex cone

$$
(\mathbf{P}) \qquad \begin{array}{rll} p^* = & \sup & c^T x \\ & \text{s.t.} & \mathcal{A}x \preceq b \quad (b - \mathcal{A}x \in \mathbb{R}^n_+) \\ & & x \in \mathbb{R}^m \end{array}
$$

Lagrangian (payoff function):
$L(x, U) = c^T x + \langle U, b - \mathcal{A}x \rangle$

$$
p^* = \max_{x} \min_{U \succeq 0} L(x, U)
$$

(the constraint $U \succeq 0$ is needed to recover the hidden constraint $\mathcal{A}x \preceq b$.)

The dual is obtained from the optimal strategy of the competing player

$$p^* \leq d^* = \min_{U \succeq 0} \max_x L(x, U) = \langle U, b \rangle + x^T (c - \mathcal{A}^* U)$$

The hidden constraint $c - \mathcal{A}^* U = 0$ yields the dual

$$(\mathbf{D}) \qquad \begin{array}{rll} d^* = & \inf & \operatorname{trace} bU \\ & \text{s.t.} & \mathcal{A}^* U = c \\ & & U \succeq 0. \end{array}$$

for the primal

$$(\mathbf{P}) \qquad \begin{array}{rll} p^* = & \sup & c^T x \\ & \text{s.t.} & \mathcal{A} x \preceq b \\ & & x \in \mathbb{R}^m \end{array}$$

If Slater's condition fails for the primal LP, then there are an infinite number of different dual programs.

The implicit equality constraints are:

$$\mathcal{A}_e x = b_e$$

where $\mathcal{A} = \begin{bmatrix} \mathcal{A}_l \\ \mathcal{A}_e \end{bmatrix}$

$$\begin{array}{rlc} d^* = & \inf & \operatorname{trace} bU \\ & \text{s.t.} & \mathcal{A}_l^* U_l + \mathcal{A}_e^* U_e = c \\ (\mathbf{D}) & & U \in \mathcal{U} \\ & & \{U : U \succeq 0\} \subset \mathcal{U} \\ & & \mathcal{U} \subset \{U : U_l \succeq 0, U_e \text{ free}\}. \end{array}$$

for the equivalent primal program

$$
(\mathbf{P}) \qquad \begin{array}{rcc} p^* = & \sup & c^T x \\ & \text{s.t.} & \mathcal{A}_l x \preceq b_l \\ & & \mathcal{A}_e x = b_e \\ & & x \in \mathbb{R}^m \end{array}
$$

DUALITY THEOREM

1. If one of the problems is inconsistent, then the other is inconsistent or unbounded.

2. **WEAK DUALITY**

   Let the two problems be consistent, and let x^0 be a feasible solution for P and U^0 be a feasible solution for D. Then

$$c^T x^0 \leq \langle b, U^0 \rangle .$$

3. **STRONG DUALITY**

   If both P and D are consistent, then they have optimal solutions and their optimal values are equal.

4. **COMPLEMENTARY SLACKNESS**

Let x^0 and U^0 be feasible solutions of P and D, respectively. Then x^0 and U^0 are optimal if and only if

$$\left\langle U^0, (b - \mathcal{A}x^0) \right\rangle = 0.$$

if and only if

$$U^0 \circ (b - \mathcal{A}x^0) = 0.$$

5. **SADDLE POINT**

The vectors x^0, U^0 are optimal solutions of P and D, respectively, if and only if (x^0, U^0) is a saddle point of the Lagrangian $L(x, U)$ for all (x,U),

$$L(x, U^0) \leq L(x^0, U^0) \leq L(x^0, U)$$

and then

$$L(x^0, U^0) = c^T x^0 = \left\langle b, U^0 \right\rangle.$$

Characterization of optimality for the dual pair x, U

$$\mathcal{A}x \preceq b \qquad \text{primal feasibility}$$

$$\mathcal{A}^*U = c \qquad \text{dual feasibility}$$

$$U \circ (\mathcal{A}x - b) = 0e \quad \text{complementary slackness}$$

$$U \circ (\mathcal{A}x - b) = \mu e \qquad \text{perturbed}$$

Forms the basis for:

primal simplex method
dual simplex method
interior point methods

What is a proper duality theory?

Do duality gaps occur in practice?

Are there an infinite number of duals if Slater's condition fails?

Facial Structure for $\mathcal{P}$

the cone $T \subset \mathcal{K}$ is a *face* of the cone $\mathcal{K}$, denoted $T \lhd \mathcal{K}$, if

$$x, y \in \mathcal{K}, \;\; x + y \in T \Rightarrow x, y \in T.$$

Each face, $\mathcal{K} \lhd \mathcal{P}$, is characterized by a subspace, $S \subset \mathbb{R}^n$.

$$\mathcal{K} = \{X \in \mathcal{P} : \mathcal{N}(X) \supset S\}.$$

$$\operatorname{relint} \mathcal{K} = \{X \in \mathcal{P} : \mathcal{N}(X) = S\}.$$

The complementary face of $\mathcal{K}$ is $\mathcal{K}^c = \mathcal{K}^\perp \cap \mathcal{P}$

$$\mathcal{K}^c = \{X \in \mathcal{P} : \mathcal{N}(X) \supset S^\perp\}.$$

$$\operatorname{relint} \mathcal{K}^c = \{X \in \mathcal{P} : \mathcal{N}(X) = S^\perp\}.$$

the face $\mathcal{K}$ (respectively, $\mathcal{K}^c$) is determined by the supporting hyperplane corresponding to any $X \in \operatorname{relint} \mathcal{K}^c$ (respectively, $\operatorname{relint} \mathcal{K}$);

and

$$XY = 0, \quad \forall X \in \mathcal{K}, \forall Y \in \mathcal{K}^c$$

An SDP:

$$
(\mathbf{P}) \qquad \begin{array}{rll} p^* = & \max & c^T x \\ & \text{s.t.} & \mathcal{A}x \preceq_{\mathcal{P}} b \\ & & x \in \mathbb{R}^m. \end{array}
$$

Optimality conditions: $\quad c \in \mathcal{A}^*(\mathcal{P}^+)$ (closed?)

dual program

$$
(\mathbf{D}) \qquad \begin{array}{rll} p^* = & \min & \operatorname{trace} bU \\ & \text{s.t.} & \mathcal{A}^*U = c \\ & & U \succeq_{\mathcal{P}^+} 0. \end{array}
$$

Optimality conditions: $\quad b \in \mathcal{R}(\mathcal{A}) + \mathcal{P}^+$ (closed?)

The *minimal cone* of P is defined as

$$\mathcal{P}^f = \cap\{\text{faces of } \mathcal{P} \text{ containing } (b - \mathcal{A}(F))\}.$$

Therefore, an equivalent program is the *regularized P program*

$$(\mathbf{RP}) \qquad \begin{array}{rcc} p^* = & \max & c^T x \\ & \text{s.t.} & \mathcal{A}x \preceq_{\mathcal{P}^f} b \\ & & x \in \mathbb{R}^m. \end{array}$$

there exists x such that $b - \mathcal{A}x \in \operatorname{relint} \mathcal{P}^f$. (generalized Slater's constraint qualification) Strong duality pair is RP and

$$(\mathbf{DRP}) \qquad \begin{array}{rcc} p^* = & \min & \operatorname{trace} bU \\ & \text{s.t.} & \mathcal{A}^* U = c \\ & & U \succeq_{(\mathcal{P}^f)^+} 0. \end{array}$$

Find $\mathcal{P}^f$? Properties?

LEMMA 1

Suppose $\mathcal{P}^f \lhd \mathcal{K} \lhd \mathcal{P}$. Then the system

$$\mathcal{A}^* U = 0, U \succeq_{\mathcal{K}^+} 0, \operatorname{trace} Ub = 0$$

is consistent only if

$$\text{the minimal cone } \mathcal{P}^f \subset \{U\}^\perp \cap \mathcal{K}.$$

PROOF

Since $\operatorname{trace} U(\mathcal{A}x - b) = 0$, for all x, we get $\mathcal{A}(F) - b \subset U^\perp$, i.e. $\mathcal{P}^f \subset \{U\}^\perp$.

$\square$

LEMMA 2 (surprising)
Suppose that $0 \neq \mathcal{K} \lhd \mathcal{P}$ (proper face). Then

$$\mathcal{P}^+ + \mathcal{K}^\perp = cl(\mathcal{P}^+ + span\ \mathcal{K}^c) \text{ is always closed.}$$

But

$$\mathcal{P}^+ + span\ \mathcal{K} \text{ is never closed.}$$

$\square$

Define: $\mathcal{P}_0 := \mathcal{P}$ and

$$\mathcal{U}_1 := \{U \succeq_{(\mathcal{P}_0)^+} 0 : \mathcal{A}^* U = 0, \operatorname{trace} Ub = 0\}$$

Choose $U_1 \in \mathcal{U}_1 \cap \operatorname{relint} \mathcal{U}_1^f$ (if $U_1 = 0$ - **STOP**)

$$\mathcal{P}_1 := \mathcal{U}_1^c = \{U_1\}^\perp \cap \mathcal{P}_0 \lhd \mathcal{P}_0$$

$$(\mathbf{RP_1}) \qquad \begin{array}{rcc} p^* = & \max & c^T x \\ & \text{s.t.} & \mathcal{A}x \preceq_{\mathcal{P}_1} b \\ & & x \in \mathbb{R}^m. \end{array}$$

We get: $\qquad p^* \leq d_1^* \leq d^*$

$$(\mathbf{DRP_1}) \qquad \begin{array}{rcc} d_1^* = & \min & \operatorname{trace} bU \\ & \text{s.t.} & \mathcal{A}^* U = c \\ & & U \succeq_{(\mathcal{P}_1)^+} 0. \end{array}$$

$$(\mathcal{P}_1)^+ = (\mathcal{P} \cap \mathcal{P}_1)^+ = \mathcal{P} + (\mathcal{P}_1)^\perp$$

$$
\begin{array}{rcl}
d_1^* = & \min & \operatorname{trace} b(U + (W + W^T)) \\
& \text{s.t.} & \mathcal{A}^*(U + (W + W^T)) = c \\
(\mathbf{ELSD_1}) & & \mathcal{A}^* U_1 = 0, \operatorname{trace} U_1 b = 0 \\
& & U \succeq 0, \quad \begin{bmatrix} I & W^T \\ W & U_1 \end{bmatrix} \succeq 0.
\end{array}
$$

We used

$$\begin{aligned}\mathcal{P}_1^{\perp} = \quad & \Big\{(W + W^T) : \mathcal{A}^* U_1 = 0, \operatorname{trace} U_1 b = 0, \\ & \qquad\qquad \begin{bmatrix} I & W^T \\ W & U_1 \end{bmatrix} \succeq 0 \Big\}\end{aligned}$$

and

$$U_1 \succeq W W^T \qquad \text{iff} \qquad \begin{bmatrix} I & W^T \\ W & U_1 \end{bmatrix} \succeq 0$$

$$\text{implies} \quad W = U_1 H, \text{for some matrix H}$$

$$
\begin{aligned}
\mathcal{U}_2 &:= \{U \succeq_{(\mathcal{P}_1)^+} 0 : \mathcal{A}^* U = 0, \operatorname{trace} Ub = 0\} \\
&= \{U + Z : \mathcal{A}^*(U + Z) = 0, \operatorname{trace} Ub = 0, \\
&\qquad U \succeq_{(\mathcal{P}_0)^+}, Z \in (\mathcal{P}_1)^\perp\}
\end{aligned}
$$

Choose $U_2 \in \mathcal{U}_2 \cap \operatorname{relint} \mathcal{U}_2^f$ (if $U_2 = 0$ - **STOP**)

$$
\mathcal{P}_2 := \mathcal{U}_2^c = \{U_2\}^\perp \cap \mathcal{P}_1 \lhd \mathcal{P}_1
$$

$$
(\mathbf{RP_2}) \qquad \begin{array}{rll} p^* = & \max & c^T x \\ & \text{s.t.} & \mathcal{A}x \preceq_{\mathcal{P}_2} b \\ & & x \in \mathbb{R}^m. \end{array}
$$

$p^* \le d_2^* \le d_1^* \le d^*$

$$
(\mathbf{DRP_2}) \qquad \begin{array}{rll} d_2^* = & \min & \operatorname{trace} bU \\ & \text{s.t.} & \mathcal{A}^* U = c \\ & & U \succeq_{(\mathcal{P}_2)^+} 0. \end{array}
$$

$(\mathcal{P}_2)^+ = (\mathcal{P} \cap \mathcal{P}_2)^+ = \mathcal{P} + (\mathcal{P}_2)^\perp$

$$
(\mathbf{ELSD_2}) \qquad
\begin{array}{rll}
d_1^* = & \min & \operatorname{trace} b(U + (W + W^T)) \\
& \text{s.t.} & \mathcal{A}^*(U + (W + W^T)) = c \\
& & \mathcal{A}^* U_1 = 0, \operatorname{trace} U_1 b = 0 \\
& & \mathcal{A}^*(U_2 + (W_1 + W_1^T)) = 0, \\
& & \operatorname{trace}(U_2 + (W_1 + W_1^T))b = 0 \\
& & U \succeq 0, \quad \begin{bmatrix} I & W_1^T \\ W_1 & U_1 \end{bmatrix} \succeq 0 \\
& & \begin{bmatrix} I & W^T \\ W & U_2 \end{bmatrix} \succeq 0.
\end{array}
$$

HOMOGENIZATION (Alternate view of optim. cond.)

$$
\begin{array}{llll}
 & 0 = & \max & c^T x + t(-p^*) \\
(\mathbf{HP}) & & \text{subject to} & \mathcal{A}x + t(-b) + Z = 0 \\
 & & & w \in \mathcal{K} = \mathbb{R}^m \otimes \mathbb{R}_+ \otimes \mathcal{P}
\end{array}
$$

This defines the objective, constraints, and variables:

$$(= \langle a, w \rangle)$$

$$(\mathcal{B}w = 0)$$

$$\left(w = \begin{pmatrix} x \\ t \\ Z \end{pmatrix} \right)$$

the feasible set is

$$F_H = \mathcal{N}(\mathcal{B}) \cap \mathcal{K}, \quad (\mathcal{B}w = 0, w \in \mathcal{K} \text{ implies } \langle a, w \rangle \leq 0)$$

Optimality conditions:

$$a = \begin{pmatrix} c \\ -p \\ 0 \end{pmatrix} \in -(\mathcal{N}(\mathcal{B}) \cap \mathcal{K})^+.$$

$$\begin{pmatrix} -c \\ p \\ 0 \end{pmatrix} \in \overline{\mathcal{R}(\mathcal{B}^*) + \mathcal{K}^+},$$

WCQ - Weakest Constraint Qualification:

CLOSURE HOLDS

Conditions for closure:

If C, D are closed convex sets and the intersection of their recession cones is empty, then $D - C$ is closed.

$cone(F_H - \mathcal{K})$ is the whole space

(Slater's)

$$\exists \hat{x} \in F \text{ such that } \mathcal{A}\hat{x} \prec b.$$

FIX: Find sets, T, to add to attain the closure. Equivalently, find sets, C, $C^+ = T$, to intersect with $\mathcal{K}$ to attain the closure since

$$(\mathcal{N}(\mathcal{A}) \cap (\mathcal{K} \cap C))^+ = \overline{\mathcal{R}(\mathcal{A}^*) + \mathcal{K}^+ + C^+}.$$

1.4 Logarithmic Barrier Function and the Central Path

References:
•Fiacco and McCormick [19] 1968, barrier techniques (SUMT)
•Nesterov and Nemirovskii [60] 1990, self-concordant barriers
• $\log\det(X)$ is a strictly concave barrier on pos def cone (e.g. Bellman 1960, [7])

family of strictly convex primal-dual (e.g. [44]) pairs, parameterized by the scalar $\mu > 0$,

$$\min\left\{\operatorname{trace} CX - \mu \log\det(X) : \mathcal{A}(X) = b, X \succ 0\right\},$$
$$\max\left\{by + \mu \log\det(Z) : \mathcal{A}^*(y) + Z = C, Z \succ 0\right\}.$$

$$(P_\mu) \qquad \inf \left\{ \langle C, X\rangle - \mu \log\det(X) \;\middle|\; \mathcal{A}(X) = b \right\},$$

$$(D_\mu) \qquad \sup \left\{ \langle b, y\rangle + \mu \log\det(Z) \;\middle|\; \mathcal{A}^*(y) + Z = C \right\}.$$

log-barrier pair parametrized by μ

objective functions are strictly convex/concave

under Slater's (strict feasibility), there exists unique (positive definite) solutions X_μ, y_μ, Z_μ

Lagrangians:

$$\begin{aligned}\mathcal{L}_P(X,y) &:= \langle C, X\rangle - \mu \log\det(X) + \langle y, b - \mathcal{A}(X)\rangle,\\ \mathcal{L}_D(X,y,Z) &:= \langle b, y\rangle + \mu \log\det(Z) + \langle X, C - Z - \mathcal{A}^*(y)\rangle.\end{aligned}$$

stationarity:

$$0 = \nabla \mathcal{L}_P(X,y) = \begin{pmatrix} C - \mu X^{-1} - \mathcal{A}^*(y) \\ b - \mathcal{A}(X) \end{pmatrix}$$

add $Z := \mu X^{-1}$

$$0 = \nabla \mathcal{L}_D(X,y,Z) = \begin{pmatrix} C - Z - \mathcal{A}^*(y) \\ b - \mathcal{A}(X) \\ \mu Z^{-1} - X \end{pmatrix}$$

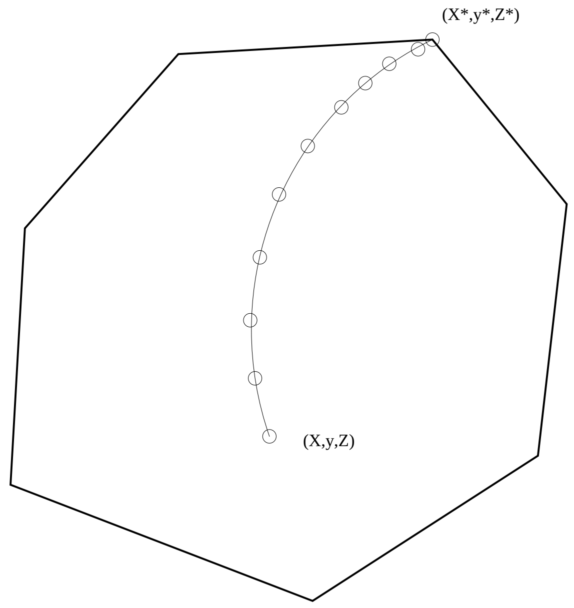
$$
\text{optimality conditions} \quad \left\{ \begin{array}{rcl} \mathcal{A}^*(y) + Z & = & C, \\ \mathcal{A}(X) & = & b, \\ \multicolumn{3}{c}{\boxed{\star\star \quad Z = \mu X^{-1}.}} \end{array} \right.
$$

For all $\mu > 0$, the set of unique solutions, denoted (X_μ, y_μ, Z_μ) is called the *primal-dual central path.*

Note complementarity:

$$\langle Z_\mu, X_\mu \rangle = \langle \mu X_\mu^{-1}, X_\mu \rangle = \mu n.$$

Central Path



1.4.1 Nonlinearity and Smoothing

$$Z - \mu X^{-1} = 0 \quad \text{is highly nonlinear}$$

Replace with (multiplication by X)

$$ZX = \mu I. \tag{1.1}$$

to reduce the nonlinearity of the system with the aim of accelerating Newton-like methods.

Consider Newton-like method to find v^* such that $F(v^*) = 0$.
(From e.g. [15] Theorems 5.2.1 and 10.2.1) radius (r) for quadratic convergence is bounded by a measure of *relative nonlinearity* of F

$$r \propto \frac{k}{\beta\gamma}.$$

k small constant that depends on method (Newton or Gauss-Newton),

$$\|\mathcal{F}'(v^*)^{-1}\| \leq \beta, \quad \text{(bound)}$$

γ is Lipschitz continuity constant for $\mathcal{F}'(v)$ near v^*.

Equivalently: (condition number of Jacobian inverse)

$$r \propto k \frac{\sigma_{\min}\mathcal{F}'(v^*)}{\sigma_{\max}\mathcal{F}'(v^*)}.$$

for $F(X,Z) := Z - \mu X^{-1}$: we have

$$\mathcal{F}'(X,Z)(dX,dZ) = dZ + \mu X^{-1} dX X^{-1}.$$

- optimal solution X^* is almost always rank deficient
- condition number is arbitrarily large
- proven radius of quadratic convergence is exceedingly small

BUT for $F(X,Z) = ZX - \mu I$:

$$\mathcal{F}'(X,Z)(dX,dZ) = ZdX + dZX,$$

- $\sigma_{\max}$ can be bounded by $||Z|| + ||X||$

(Jacobian is full-rank from standard assumptions)

- we obtain a nonzero radius of quadratic convergence.

Summary - optimality conditions (look like for LP):

$$\mathcal{A}^*(y) + Z = C, \tag{1.1}$$

$$\mathcal{A}(X) = b, \tag{1.1}$$

$$ZX = \mu I. \tag{1.1}$$

1.5 Differences with LP

1. Duality gaps can exist for SDP in the absence of strictly feasible solutions (Slater's constraint qualification, CQ).
2. Strict complementarity can fail at the optimum.
3. The perturbed optimality conditions involve nondiagonal matrices and map between different spaces (overdetermined -does it require symmetrization?).

1.5.1 Duality Gaps

Example 1.5.1.1 *If the primal is*

$$
(\mathbf{P}) \qquad
\begin{array}{rll}
p^* = & \max & x_2 \\
& s.\ t. & \begin{bmatrix} x_2 & 0 & 0 \\ 0 & x_1 & x_2 \\ 0 & x_2 & 0 \end{bmatrix} \preceq \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}
\end{array}
$$

then the dual is

$$
(\mathbf{D}) \qquad
\begin{array}{rll}
d^* = & \min & \operatorname{trace} U_{11} \\
& s.\ t. & U_{22} = 0, \quad U_{11} + 2U_{23} = 1 \\
& & U \succeq 0.
\end{array}
$$

Then $p^* = 0 < d^* = 1$.

1.5.2 Strict Complementarity

Example 1.5.2.1 *If the primal is*

$$
(\mathbf{P}) \qquad \begin{array}{rll} p^* = & \max & x_1 \\ & s.\ t. & \begin{bmatrix} x_1 & x_3 & x_2 \\ x_3 & x_2 & 0 \\ x_2 & 0 & x_3 \end{bmatrix} \preceq \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{array}
$$

then the dual is

$$
(\mathbf{D}) \qquad \begin{array}{rll} d^* = & \min & \operatorname{trace} U_{33} \\ & s.\ t. & U_{11} = 1 \\ & & U_{22} + 2U_{13} = 0 \\ & & U_{33} + 2U_{12} = 0 \\ & & U \succeq 0. \end{array}
$$

Then the (unique) optimum pair is:

$$U = \mathrm{Diag}\,(1, 0, 0), \quad x = [0, 0, 0], Z = \mathrm{Diag}\,(0, 0, 1)$$

But: $U + Z$ *is not positive definite*

1.5.3 Overdetermined System of Optimality Conditions

$$
\begin{aligned}
\mathcal{A}^*(y) + Z &= C, \\
\mathcal{A}(X) &= b, \\
ZX &= \mu I.
\end{aligned}
$$

unfortunate consequence:
residual $(ZX - \mu I)$ is not symmetric unless X and Z commute
therefore Newton step impossible

due to success in linear programming codes,

current philosophy - symmetrize?? But - correct???

E.g. AHO direction: project onto subspace of symmetric matrices

$$H(M) := \frac{1}{2}[M + M^T].$$

Unscaled Symmetric System,

$$\begin{array}{rcll} \mathcal{A}^*(dy) + dZ & = & -(\mathcal{A}^*(y_y) + Z_k - C) & =: -F_d, \\ \mathcal{A}(dX) & = & -(\mathcal{A}(X_k) - b) & =: -f_p, \\ H(Z_k dX + dZ X_k) & = & -H(Z_k X_k - \mu I) & =: -H(F_c). \end{array}$$

Monteiro-Zhang family of directions [51]:
scale the problem, nonsingular matrix P

$$X := PXP^T, \quad A_i := P^{-t}A_iP^{-1}, \quad C := P^{-t}CP^{-1}. \tag{1.2}$$

scaled primal problem

$$\min\left\{\langle C, X\rangle \mid \mathcal{A}(X) = b, X \succeq 0\right\}, \tag{1.3}$$

scaled dual problem using $y := y, Z := P^{-t}ZP^{-1}$

$$\max\left\{\langle b, y\rangle \mid \mathcal{A}^*(y) + Z = C, Z \succeq 0\right\}. \tag{1.4}$$

symmetric direction for family of parameterized programs

$$
\begin{array}{rcll}
\mathcal{A}^*(dy) + dZ & = & -(\mathcal{A}^*(y_k) + Z_k - C) & =: -F_d, \\
\mathcal{A}(dX) & = & -(\mathcal{A}(X_k) - b) & =: -f_p, \\
H(Z_k dX + dZ X_k) & = & -H(Z_k X_k - \mu I) & =: -H(F_c),
\end{array}
$$

where

$$
\begin{array}{rclcl}
F_d & = & \mathcal{A}^*(y_k) + Z_k - C & = & P^{-t} F_d P^{-1}, \\
f_p & = & \mathcal{A}(X_k) - b & = & f_p, \\
F_c & = & Z_k X_k - \mu I & = & P^{-t} F_c P^{-1}.
\end{array}
$$

P	Direction	Solvers
I	AHO [2]	SDPPack,SDPA,SDPT3
$Z^{\frac{1}{2}}$	HKM [31, 40, 49]	CSDP,SDPA,SDPT3
$[X^{\frac{1}{2}}(X^{\frac{1}{2}}ZX^{\frac{1}{2}})^{-\frac{1}{2}}X^{\frac{1}{2}}]^{\frac{1}{2}}$	NT [61]	SDPA,SDPT3,SeDuMi

Table 1: Instances of Monteiro-Zhang scaling matrix P of equation (1.2).

CSDP	http://www.nmt.edu/~borchers
SDPA	http://www-neos.mcs.anl.gov/neos/solvers/SDP:SDPA
SDPPACK	http://cs.nyu.edu/cs/faculty/overton/sdppack/sdppack.html
SDPT3	http://www.math.cmu.edu/~reha/sdpt3.html
SeDuMi	http://www2.unimaas.nl/~sturm/research.html

Table 2: Major Semidefinite Solvers

1.5.4 Historical Notes

• currently most active area in optimization (see HANDBOOK OF SEMIDEFINITE PROGRAMMING: Theory, Algorithms, and Applications, 2000, [86], for comprehensive results, history, references, ...)

• Lyapunov over 100 years ago on stability analysis of differential equations

• Bohnenblust 1948 on geometry of the cone of SDPs, [10]

• Yakubovitch in the 1960's and Boyd and others on convex optimization in control in the 1980's, e.g. solving Ricatti Equations (called LMIs), e.g. [13],[84, 83]

• matrix completion problems (another name for SDP) started early 1980's, continues to be a very active area of research, [18],[27], surveys: [38, 41]

• combinatorial optimization applications 1980's: Lovász *theta function* [45], more recently the strong approximation results for the max-cut problem by Goemans-Williamson, e.g. [26], survey papers: [24, 25],[70]

• linear complementarity problems can be extended to problems over the cone of semidefinite matrices, e.g. [20, 37, 40, 39, 50].

• Complexity, Distance to Ill-Posedness, and Condition Numbers SDP is a convex program and it can be solved to any desired accuracy in polynomial time, see seminal work of Nesterov and Nemirovski e.g. [55, 56, 60, 58, 54, 57, 59]. Another measure of complexity is the distance to ill-posedness: e.g. work by Renegar [72, 76, 75, 74, 73].

• Cone Programming this is a generalization of SDP, also called *generalized linear programming*, in paper by Bellman and Fan 1963, [8]. Other books deal with problems over cones date back to 60s, e.g. [62],[32, 46, 36, 35, 68]. More recently, generalization of SDP to more general cones, e.g. Güler and Tuncel, [29, 82] and also Hauser [30]

- Other Related Areas e.g.: Eigenvalue Functions, e.g. [14], [64, 65]; Financial Applications; Generalized Convexity, e.g. [77, 48]; Statistics; Nonlinear Programming; ...

1.6 Motivation/Examples/Applications

- Quadratic constrained quadratic programs
- Lovász theta function
- Statistics
- minimizing the L_2-operator norm of a matrix
- linear programming
- robust mathematical programming
- engineering, e.g. control theory
- Combinatorial Problems, e.g. the Max-Cut Problem

1.6.1 Quadratic Constrained Quadratic Programs

What is SEMIDEFINITE PROGRAMMING?
Why use it?

- Quadratic approximations are better than linear approximations. (For example, model $x \in \{0, 1\}$ using $x^2 - x = 0$.)

And, we can solve relaxations of quadratic approximations efficiently using semidefinite programming.

1.6.2 Quadratic Matrix Inequalities

Let X be $k \times l$ and
$F(X) = (AXB)AXB)^T + CXD + (CXD)^T + E$. Then

$$\{(X,Y) : F(X) \preceq Y\} =$$
$$\left\{(X,Y) : \begin{pmatrix} I & (AXB)^T \\ (AXB) & \left(Y - E - CXD - (CXD)^T\right) \end{pmatrix} \succeq 0\right\}$$

(by Schur Complement Lemma)

1.6.3 Nonnegative Polynomial Approximation

(e.g. [53, 23]) best polynomial $p(t)$ approximation to a given function $f(t)$, where $p(t) \in \mathbb{P}^+_{2k}(\mathbb{R})$, i.e. nonnegative, degree $\leq 2k$

$$p(t) = \sum_{i=0}^{k} p_i t^i \equiv (p_0, p_1, \ldots p_{2k})^T$$

Then

$$p \in \mathbb{P}^+_{2k}(\mathbb{R}) \quad \underline{\text{ iff }} \quad p(t) = e(t)^T X e(t), \text{ for some } X \succeq 0,$$

where $e(t) = \left(1, t, t^2, \ldots, t^k\right)$

1.6.4 Lovász theta function, [45]

Let $G = (V, E)$ be a ***simple undirected graph***
(node set V, edge set E)

a ***clique*** - is subgraph of G, $\exists$ edge between every pair of nodes

number of colours to colour clique $\geq$ size of the clique

$\chi(G)$ **-** ***chromatic number***, least number of colours for G

size of maximum clique $\omega(G)$

$$\omega(G) \leq \chi(G) \quad \text{two } \textbf{hard} \text{ to compute integers}$$

strict inequality, for example: G simple circle with $\|V\| = 5$, then: $2 = \omega(G) < \chi(G) = 3$.

Let $A = A(G)$ be ***adjacency matrix*** of G:
i.e. $A_{ij} = 1$ if $(i,j) \in E$ and 0 otherwise.

The largest eigenvalue of A, $\lambda_{\max}(A) = \max_{||x||=1} x^T A x$ is a convex function of its elements, A_{ij}. Fix $A_{ij} = 1$ if $(i,j) \in E$, otherwise leave the elements free, i.e. define the *Lovász theta number*

$$\vartheta(G) := \min \left\{ \lambda_{\max}(X) : X = X^T, X_{ij} = 1 \text{ if } (i,j) \in E \right\}$$

Lemma 1.6.4.1

size of max clique $\omega(G) \le \vartheta(G)$ *Lovász theta number*

Proof. Rename vertices of G, such that $1, 2, \ldots, k$ is the maximum clique of G. This is a permutation of the rows/columns of the adjacency matrix X, so the eigenvalues of X are unchanged. We get

$$X = \begin{pmatrix} J_k & X_2 \\ X_3 & X_4 \end{pmatrix}, \qquad J_k = E \in \mathbb{R}^{k \times k}.$$

Therefore

$$\lambda_{\max}(J_k) = \operatorname{trace}(J_K) = \operatorname{trace}(ee^T) = \operatorname{trace}(e^T \cdot e) = k$$

Now use interlacing of eigenvalues

$$\lambda_{max}(J_k) \le \lambda_{max}(X)$$

. ■

In addition, it can be shown that

size of max clique $\quad \omega(G) \leq \vartheta(G) \leq \chi(G) \quad$ size of chromatic number

$\vartheta(G)$ gives a lower bound of $\chi(G)$ and an upper bound of $\omega(G)$. (bounds for two NP-hard numbers)

And, as shown in Section 1.0.4, we can use SDP to solve the min-max eigenvalue problem (efficiently).

1.6.5 Generalized Eigenvalue Problems for $X = X^T$

1. **The Generalized Eigenvalue Problem**

   Let M, A be two $n \times n$ symmetric matrices, $M \succ 0$.

   The set of eigenvalues of the *matrix pencil*, denoted $[M, A]$, is $\{\lambda \in \mathbb{R} : \lambda M - A \text{ is singular}\}$.

   $$\lambda_{\max}([M, X]) \le t \quad \text{is equivalent to} \quad tM - X \succeq 0.$$

2. **Spectral Norm of Symmetric X**

   $$\{|\lambda_i(X)| \le t, \ \forall i\} \text{ iff } \{tI - X \succeq 0, \ tI + X \succeq 0\}$$

3. **Sum of k largest eigenvalues of Symmetric X**

   Let $S_k(X)$ denote the sum of the largest k eigenvalues of X.

   $$S_k(X) \le t \underline{\text{ iff }} \ t - ks - \operatorname{trace} Z \succeq 0, Z \succeq 0, Z - X + sI \succeq 0.$$

1.6.6 SDP Application in Statistics

If $m_0, m_1, m_2, \ldots, m_{2n}$ are moments of some distribution, then

$$H(m_0, m_1, \ldots, m_{2n}) = \begin{bmatrix} m_0 & m_1 & \cdots & m_n \\ m_1 & \cdots & \cdots & m_{n+1} \\ \cdots & \cdots & \cdots & \cdots \\ m_n & m_{n+1} & \cdots & m_{2n} \end{bmatrix} \succeq 0$$

Find distribution with maximal variance and $l_i \le m_i \le u_i$:

$$\begin{array}{ll} \max & y \\ \text{s.t.} & \begin{bmatrix} m_2 - y & m_1 \\ m_1 & 1 \end{bmatrix} \succeq 0 \\ & l_i \le m_i \le u_i \, (i = 1, \ldots 2n) \end{array}$$

1.6.7 Minimizing the L_2-operator Norm of a Matrix

1.6.8 Linear Programming

1.6.9 Robust Mathematical Programming

1.6.10 Control Theory

1.6.11 Max-Cut Problem

(MC) is a combinatorial optimization problem on undirected graphs with weights on the edges.

Problem 1.6.11.1 *Find a partition of the set of vertices into two parts that maximizes the sum of the weights on the edges that have one end in each part of the partition.*

Quadratic Model of MC

given graph G with vertex set $\{1,\ldots,n\}$
weighted adjacency matrix $A(G)=(a_{ij})$, weight on edge ij
Laplacian matrix $L := \mathrm{Diag}\,(A(G)\cdot e) - A(G)$
Let $v \in \{\pm 1\}^n$ represent any cut in the graph

$$\text{(MC)} \qquad \begin{array}{rll} \mu^* = & \max & \frac{1}{4} v^T L v \\ & \text{s.t.} & v_i^2 = 1, \quad i = 1,\ldots,n, \end{array}$$

where μ^* denotes the optimal value of MC.

Some applications of MC

- Statistical physics: Finding the ground state of a spin glass according to the Ising model
- VLSI: Minimizing the number of vias in a two-sided circuit board
- Network design: Solving the separation problem in a cutting plane approach

Spin glasses

A *spin glass* is an alloy of magnetic impurities diluted in a nonmagnetic metal.

Suppose we have a spin glass subjected to an exterior magnetic field. At a very low temperature (close to 0^{o}K) the spin glass attains a minimum energy configuration, called the ground state.

This state can be found by *minimizing* the hamiltonian representing the total energy of the system.

Ising model

Suppose the spin glass has n impurities, or atoms.

Let $v_i \in \{\pm 1\}$ denote the magnetic orientation of atom i ("north/south"), and v_0 denote the magnetic orientation of the exterior magnetic field which has strength h.

For any pair i, j of atoms, J_{ij} denotes the interaction between them, e.g.

$$J_{ij} = A \frac{\cos(D r_{ij})}{B^3 r_{ij}^3},$$

where r_{ij} is the distance between the atoms, and A, B, D are given constants.

Then the energy of the system is modelled by the hamiltonian

$$H := -\sum_{1\le i<j\le n} J_{ij} v_i v_j - h \sum_{j=1}^{n} v_0 v_j .$$

To find the ground state, we want to minimize H over the 2^n possible configurations of the Ising spins (assuming $v_0 = +1$ wlog).

Equivalently, we can maximize $-H$.

MC Formulation

Let $G = (V, E)$ be a complete graph with $V = \{0, 1, \ldots, n\}$ and

$w_{ij} = -J_{ij}$, for $1 \leq i < j \leq n$,

$w_{0j} = -h$, for $1 \leq j \leq n$.

Then

$$\begin{aligned} H &= \sum_{0 \leq i < j \leq n} w_{ij} v_i v_j \\ &= \sum_{0 \leq i < j \leq n : v_i = v_j} w_{ij} v_i v_j + \sum_{0 \leq i < j \leq n : v_i \neq v_j} w_{ij} v_i v_j \\ &= \sum_{ij \notin \delta(S)} w_{ij} - \sum_{ij \in \delta(S)} w_{ij}, \end{aligned}$$

where $S = \{i \in V : v_i = +1\}$ and $\delta(S) = \{ij \in E : i \in S, j \in V \backslash S\}$.

Now define the constant $K := \sum_{ij \in E} w_{ij}$.

Then

$$H - K = -2 \sum_{ij \in \delta(S)} w_{ij}$$

and so $-H = 2 \sum_{ij \in \delta(S)} w_{ij} - K$. So our equivalent optimization problem is:

$$\begin{array}{ll} \max & \sum_{ij \in \delta(S)} w_{ij} \\ \text{s.t.} & S \subseteq V \end{array}$$

which is a MC problem.

Approaches to MC

The general MC problem is NP-hard, even though it is tractable for some classes of graphs, e.g. planar graphs. In fact, solving within relative error of 1 percent is NP-hard.

Techniques used for general MC include:

- Heuristics;
- Integer programming enumerative techniques (Branch-and-Bound);
- Approximation algorithms.

We seek *SDP relaxations* of MC (solvable in polynomial-time) that yield tight (upper) bounds on the optimal value of MC.

One Formulation of MC

Let $v \in \{\pm 1\}^n$, $n = |V|$, represent a cut in the graph G via $S = \{i : v_i = +1\}$ and $V \backslash S = \{i : v_i = -1\}$.

Then we can formulate MC as:

$$\begin{array}{rll} \mu^* = & \max & \sum\limits_{1 \leq i < j \leq n} w_{ij} \left(\frac{1 - v_i v_j}{2} \right) \\ & \text{s.t.} & v \in \{\pm 1\}^n. \end{array}$$

Equivalently,

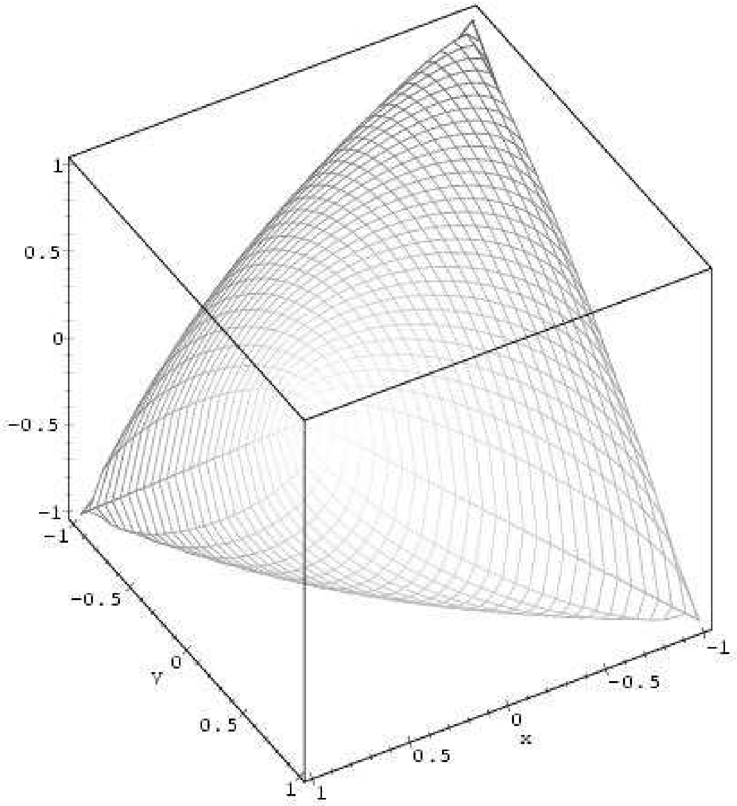
$$(\mathrm{MC1}) \qquad \begin{array}{rll} \mu^* = & \max & v^T Q v \\ & \text{s.t.} & v_i^2 = 1, \qquad i = 1, \ldots, n, \end{array}$$

where $Q = \frac{1}{4}(\mathrm{Diag}\,(A\,e) - A)$ (Laplacian L), and $A = (w_{ij})$ is the weighted adjacency matrix of G.

With $Q := \frac{1}{4}L$, $X := vv^T, v \in \{\pm 1\}^n$,
then $v^T Q v = \operatorname{trace} QX$ and equivalent formulation is:

$$
\text{(MC1)} \qquad
\begin{array}{rll}
\mu^* = & \max & \operatorname{trace} QX \\
 & \text{s.t.} & \operatorname{diag}(X) = e \\
 & & \operatorname{rank}(X) = 1 \\
 & & X \succeq 0, X \in \mathcal{S}^n,
\end{array}
$$

relax by deleting the *hard* constraint $\operatorname{rank}(X) = 1$ (get *elliptope*)



Elliptope for n=3, [42]

Goemans & Williamson proved that

if $w_{ij} \geq 0 \; \forall \, i, j$, then

$$\mu^* \geq \alpha \, \nu_1^*,$$

where $\alpha = \min_{0 \leq \theta \leq \pi} \frac{2}{\pi} \frac{\theta}{1 - \cos\theta} \approx 0.87856.$

Since $\frac{1}{\alpha} \approx 1.13823 \leq 1.14$, this implies

$$\mu^* \leq \nu_1^* \leq 1.14 \, \mu^*.$$

Other such convex relaxations have been studied before.

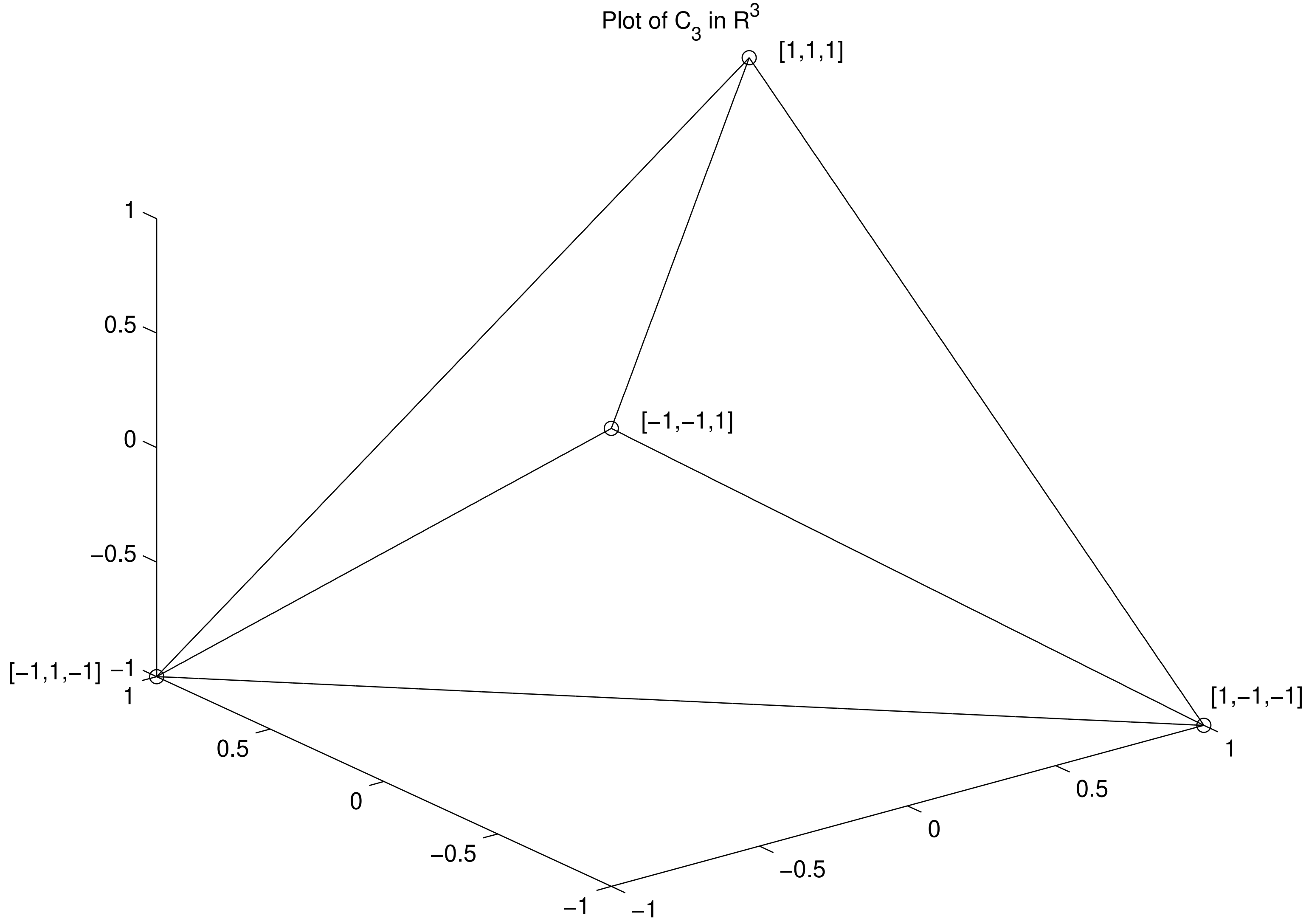
The smallest convex set containing all the rank-one matrices X corresponding to cuts is the *cut polytope*:

$$C_n := \mathrm{Conv}\{X : X = vv^T, v \in \{\pm 1\}^n\}.$$

In fact,

$$\begin{array}{rll} \mu^* = & \max & \mathrm{trace}\, QX \\ & \text{s.t.} & X \in C_n. \end{array}$$

However, it is not known how to optimize in polynomial-time over C_n.



Cut polytope for n=3

Another convex relaxation is the *metric polytope* M_n, defined by

$$
\begin{aligned}
M_n := \{ & X \in \mathcal{S}^n : \operatorname{diag}(X) = e, \text{ and} \\
& X_{ij} + X_{ik} + X_{jk} \geq -1, X_{ij} - X_{ik} - X_{jk} \geq -1, \\
& - X_{ij} + X_{ik} - X_{jk} \geq -1, -X_{ij} - X_{ik} + X_{jk} \geq -1, \\
& \forall\, 1 \leq i < j < k \leq n \}.
\end{aligned}
$$

These inequalities model the fact that for any three mutually connected vertices in the graph, it is only possible to cut either zero or two of the edges joining them.

It is known that:

$C_n = M_n$ for $n \leq 4$, but $C_n \subsetneq M_n$ for $n \geq 5$.

2 Algorithms

2.1 Outline

- Primal-dual framework
- Gauss-Newton methods
- Exploiting Sparsity

2.2 Primal-Dual Framework

Generic Interior-Point Algorithm for Monteiro-Zhang family

- Given $\epsilon > 0$ *Tolerance*
- Given X, y, Z *Must satisfy some condition*
- $\mu = \frac{\langle Z,X \rangle}{n}$ *Initial barrier parameter*
- **WHILE** $\mu > \epsilon$
  - •• Choose scaling P *Possibly dependent on X, Z*
  - •• Choose centrality $0 < \tau < 1$ *According to some condition*
  - •• $\mu \leftarrow \tau \frac{\langle Z,X \rangle}{n}$ *Update target*
  - •• Solve (*scaled ahosystem*) *Scaled AHO direction*
  - •• Choose steplength α *To maintain positive definiteness*
  - •• $X = X + dX, y = y + dy, Z = Z + dZ$ *Update iterate*
- **ENDWHILE**

2.2.1 Common LP,SDP Framework

For LPs, Loréntz cone problems, SDPs: use Newton's method to get step at each iteration.

Common system of equations at each step:

$$\begin{pmatrix} \mathcal{A} & 0 & 0 \\ 0 & \mathcal{A}^* & I \\ \mathcal{E} & 0 & \mathcal{F} \end{pmatrix} \begin{pmatrix} \Delta X \\ \Delta y \\ \Delta Z \end{pmatrix} = \begin{pmatrix} \mathcal{F}_p \\ \mathcal{F}_d \\ \mathcal{F}_c \end{pmatrix} \tag{2.1}$$

Δx, Δy, ΔZ step; $\mathcal{F}_p$, $\mathcal{F}_d$, and $\mathcal{F}_c$ residuals primal, dual, complementarity

Assume $\mathcal{E}$ and $\mathcal{A}\mathcal{E}^{-1}\mathcal{F}\mathcal{A}^*$ are invertible. Then the solution to (2.1) is:

$$\begin{aligned}
\Delta y &= (\mathcal{A}\mathcal{E}^{-1}\mathcal{F}\mathcal{A}^T)^{-1}(\mathbf{F_p} - \mathcal{A}\mathcal{E}^{-1}(\mathbf{F_c} - \mathcal{F}\mathbf{F_d})) \\
\Delta Z &= \mathbf{F_d} - \mathcal{A}^*\Delta y \\
\Delta X &= \mathcal{E}^{-1}(\mathbf{F_c} - \mathcal{F}(\mathbf{F_d} - \mathcal{A}^*\Delta y))
\end{aligned} \tag{2.2}$$

LP Case

central path satisfies

$$\begin{aligned} \mathcal{A}X &= b \\ \mathcal{A}^* y + Z &= C \\ X_i Z_i &= \mu \; i = 1, \ldots, n, \end{aligned} \tag{2.3}$$

where $X = x, Z = z$. Newton's method on (2.3) yields (2.1), where $\mathcal{A} = A$, $\mathcal{E} = \mathrm{Diag}\,(z)$, $\mathcal{F} = \mathrm{Diag}\,(x)$, $\mathbf{F_p} = b - Ax$, $\mathbf{F_d} = c - A^T y - z$, $\mathbf{F_c}_i = \mu - x_i z_i$, $i = 1, \ldots, n$.

Remark 2.2.1.1 *The main work is the computation of (the positive definite Schur complement)* $(\mathcal{A}\mathcal{E}^{-1}\mathcal{F}\mathcal{A}^T)^{-1}$.

SDP Case

Use $(-\ln \det X)$ for barrier function. Then its gradient is $-X^{-1}$, its Hessian is $X^{-1} \otimes X^{-1}$. (So convex function on $\mathcal{P}$.) The log-barrier problem is:

$$\begin{array}{ll} \min\limits_{X} & \operatorname{trace} CX - \mu \ln \det X \\ \text{s.t.} & \operatorname{trace} A_i X = b_i \;\; i = 1, \cdots, m \end{array}$$

Let y be the Lagrange multiplier; Lagrangian is

$$\mathcal{L}(X, y) = \operatorname{trace} CX - \mu \ln \det X + y^T (b - \mathcal{A}X).$$

.

KKT system:

$$\begin{aligned} \nabla_X \mathcal{L} &= C - \mu X^{-1} - \mathcal{A}^* y = 0 \\ \nabla_y \mathcal{L} &= b - \mathcal{A}X = 0, \; i = 1, \ldots, m \end{aligned}$$

Define $Z := \mu X^{-1}$. The system of equations on the central path is now:

$$
\begin{array}{rcl}
A_i \bullet X & = & b_i \ (i = 1, \cdots m) \\
\sum_{i=1}^m y_i A_i + S & = & C \\
\text{(one of)} & & \left\{ \begin{array}{rl} S & = \mu X^{-1} \\ X & = \mu S^{-1} \\ XS & = \mu I \end{array} \right.
\end{array}
\tag{2.4}
$$

Use $XZ = \mu I$ and apply Newton's method to the system:

$$\begin{aligned} \mathcal{A}\Delta X &= b - \mathcal{A}X && (= (\mathbf{F_p})_i) \\ \mathcal{A}\Delta y + \Delta Z &= C - S - \mathcal{A}^* y && (= F_d) \\ X\Delta Z + \Delta X Z &= \mu I - XZ && (= F_c) \end{aligned} \tag{2.5}$$

To write (2.5) into the common form of (2.1), one should first turn matrix into vector. For a $p \times q$ matrix B, define $\operatorname{vec} B$ to be a pq column vector made up of columns of B stacked on each other. Another way is to use svec. Since here the matrices X and Z are symmetric, only the lower triangle of the matrices need to be saved. Let svec denote such an operator that maps $\mathbb{R}^{n\times n}$ to $\mathbb{R}^{\frac{n(n+1)}{2}}$. For the reason of simplicity, we use *kvec* for now.

Definition 2.2.1.1 *Kronecker product($\otimes$): If $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{p\times q}$, then $A \otimes B \in \mathbb{R}^{mp\times nq}$,*

$$A \otimes B := \begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{pmatrix}$$

Example 2.2.1.1 *If*

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \text{ and } B = \begin{pmatrix} 5 & 6 \end{pmatrix}$$

then

$$A \otimes B = \begin{pmatrix} 5 & 6 & 10 & 12 \\ 15 & 18 & 20 & 24 \end{pmatrix}$$

Fundamental properties of Kronocker product:

1. $$\text{vec}\,(ABC) = (C^T \otimes A)\text{vec}\,(B) \text{ and } (A \otimes B)(C \otimes D) = (AC) \otimes (BD)$$

2. $(A \otimes B)^T = A^T \otimes B^T$

3. $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$

4. $\text{Diag}\,(x) \otimes \text{Diag}\,(y) = \text{Diag}\,(x \otimes y)$

5. Suppose $A = P^{-1}(\text{Diag}\,(\lambda_i))P$, and $B = Q^{-1}(Diag(\omega_j))Q$, then $A \otimes B = (P^{-1} \otimes Q^{-1})(\text{Diag}\,(\lambda_i \omega_j))(P \otimes Q)$

6. Let λ_i be the i-th eigenvalue of A, and ω_j be the j-th eigenvalue of B, then the ij-th eigenvalue of $A \otimes I + I \otimes B$ is $\lambda_i + \omega_j$

$X\Delta Z + \Delta X Z = \mu I - XZ$, can be rewritten using Kronocker products:

$$(Z \otimes I)\mathrm{vec}\,\Delta X + (I \otimes X)\mathrm{vec}\,\Delta Z = \mathrm{vec}\,(\mu I - XZ) = \mathrm{vec}\,(F_c)$$

$XZ + ZX$ Method to guarantee symmetry (AHO Direction)

$$\text{replace } XZ = \mu I \text{ with } \frac{XZ + ZX}{2} = \mu I.$$

To justify this method, one can show that for $X \succeq 0, Z \succeq 0$,

$$XZ = \mu I \iff \frac{XZ + ZX}{2} = \mu I.$$

2.3 Gauss-Newton Method

2.3.1 Outline

- Background for GN
- Properties of the Gauss-Newton Direction
- Conditioning of the Jacobian
- Invariance
- Convergence of Gauss-Newton Method
- Numerical Tests

2.3.2 Background for GN Methods

H_1, H_2 TWO real Hilbert spaces

$F : H_1 \to H_2$ a nonlinear operator

solve the equation:

$$F(x) = 0. \tag{2.6}$$

Assumption: Problem (2.6) has a solution $\hat{x}$

Newton's method ([63]):

$$x_{k+1} = x_k - F'(x_k)^{-1}F(x_k), \tag{2.7}$$

where $F'(x) : H_1 \to H_2$ - Fréchet derivative of F (linear operator)

needs: bounded invertibility of $F'(x_k)$

(existence of a bounded linear operator $[F'(x)]^{-1}$ near $\hat{x}$)

to avoid the restrictions use Gauss-Newton procedure ([63]):

$$\begin{aligned} x_{k+1} &= x_k - [F'^*(x_k)]^\dagger F(x_k) \\ &= x_k - [F'^*(x_k)F'(x_k)]^{-1} F'^*(x_k)F(x_k), \end{aligned} \tag{2.8}$$

where $\cdot^\dagger$ denotes Moore-Penrose generalized inverse

2.3.3 Gauss-Newton Methods for SDP

In numerical analysis, how do we solve a nonlinear system of equations?

The standard approach is to solve the equivalent nonlinear least squares problem using "Gauss-Newton" method.

$$\min \frac{1}{2} ||F_\mu(X, y, Z)||_2^2$$

subject to $X, Z > 0$.

In the LP case, when we differentiate we get a square system and Gauss-Newton reduces to Newton's method.

i.e. solving

$$\left(F'_\mu\right)^T F'_\mu(\Delta s) = -\left(F'_\mu\right)^T F_\mu$$

(where F'_μ is the Jacobian)
is equivalent to solving

$$F'_\mu(\Delta s) = -F_\mu$$

framework p-d i-p:

Given $(X^0, y^0, Z^0) \in \mathcal{F}^0$

for $k = 0, 1, 2 \ldots$

solve for the search direction

$$F'_\mu(X^k, y^k, Z^k) \begin{pmatrix} \Delta X^k \\ \Delta y^k \\ \Delta Z^k \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -X^k Z^k + \sigma_k \mu_k I \end{pmatrix}$$

where σ_k centering, $\mu_k = \operatorname{trace} X^k Z^k / n$

$$\begin{aligned} &(X^{k+1}, y^{k+1}, Z^{k+1}) = \\ &\qquad (X^k, y^k, Z^k) + \alpha_k(\Delta X^k, \Delta y^k, \Delta Z^k) \end{aligned}$$

so that $(X^{k+1}, Z^{k+1}) \succ 0$

end (for).

Shrinking the Overdetermined System

We can find redundant equations and solve a smaller system. We can remove the lower triangular part from the nonlinear equation and get the equivalent equation

$$\mathcal{T}(ZX - \mu I) = 0,$$

where the linear operator $\mathcal{T} : \mathcal{M}^n \to \mathcal{S}^n$ by ignoring the strictly lower triangular part of the matrix, i.e. the i, j components are

$$(\mathcal{T}(W))_{ij} = \begin{cases} W_{ij} & \text{if } i \leq j \\ W_{ji} & \text{if } i > j \end{cases}$$

The resulting system is now square.

$\mathcal{T}$ is an orthogonal projection

The optimality conditions become

$$\begin{pmatrix} Z + C - \mathcal{A}^* y \\ b - \mathcal{A}(X) \\ \mathcal{T}(ZX - \mu I) \end{pmatrix} = 0.$$

This does not introduce new nonlinearities but does guarantee that we map between the same spaces. In fact, we can prove that we do not lose information in the optimality conditions when we only consider the upper triangular parts.

LEMMA
Suppose that $Z \succ 0$. Then the linear operator $\mathcal{T}_Z(\cdot) := \mathcal{T}(Z\cdot)$ is a one-one mapping on $\mathcal{S}^n$. In addition, if $\mathcal{T}_Z(X) = \mu I$, for some $X = X^T$, then $ZX = \mu I$.

The Newton direction is obtained from solving the system

$$\begin{aligned} \Delta Z - \mathcal{A}^*(\Delta y) &= -F_d \\ -\mathcal{A}(\Delta X) &= -F_p \\ \mathcal{T}(Z\Delta X + \Delta Z X) &= -\mathcal{T}(ZF_{Z,X}). \end{aligned}$$

Equivalently,

$$\begin{bmatrix} 0 & -\mathcal{A}^* & I \\ -\mathcal{A} & 0 & 0 \\ \mathcal{T}_Z(\cdot) & 0 & \mathcal{T}(\cdot X) \end{bmatrix} \begin{pmatrix} \Delta X \\ \Delta y \\ \Delta Z \end{pmatrix} = \begin{pmatrix} -F_d \\ -F_p \\ -\mathcal{T}_Z(X) + \mu I \end{pmatrix}.$$

MORE LINEARITY
SPEEDS UP NEWTON'S METHOD
but
results in overdetermined system

2.3.4 Properties of the Gauss-Newton Directions

Well-Defined

Lemma 2.3.4.1 *If $\mathcal{A}$ is surjective, the Gauss-Newton direction exists and is unique for all $X \succ 0$, $Z \succ 0$. (The Jacobian is full-rank.)*

result requires only positive definite X and Z in contrast to other directions

nonsingularity of Jacobian holds in the limit

Lemma 2.3.4.2 *If $\mathcal{A}$ is surjective and the optimal primal-dual solution (X^*, y^*, Z^*) is unique and strictly complementary, then $\mathcal{F}_\mu'(v^*)$ at $\mu = 0$ is nonsingular.*

Merit function - Descent direction

$$\varphi(v) := \frac{1}{2}\langle F_\mu(v), F_\mu(v)\rangle \tag{2.8}$$

$$= \frac{1}{2}\langle F_d, F_d\rangle + \frac{1}{2}\langle f_p, f_p\rangle + \frac{1}{2}\langle F_c, F_c\rangle \tag{2.8}$$

$$=: \varphi_d(v) + \varphi_p(v) + \varphi_c(v). \tag{2.8}$$

Lemma 2.3.4.3 *The Gauss-Newton direction dv, is a strict descent direction for the merit function $f_\mu(v) = \frac{1}{2}\langle F_\mu(v), F_\mu(v)\rangle$ if and only if $F_\mu(v)$ is not perpendicular to the range of $\mathcal{F}'_\mu(v)$.*

Summary

Well-Defined
nonsingularity of Jacobian holds in the limit
Merit function - Descent direction

if $\mathcal{A}$ is surjective (and we may assume it is), the Gauss-Newton direction is a strict descent direction until stationarity of the merit function is attained.
Moreover, under uniqueness of the optimal solution and strict complementarity, (conditions that hold generically [67]), the Gauss-Newton system minimizes the risk of ill-conditioning as we approach the optimal solution.

2.3.5 Conditioning of the Jacobian

equivalent form of the (overdetermined) optimality system

$$\text{AHO} \quad \left\{ \begin{array}{rcl} \mathcal{A}^*(dy) + dZ & = & -(\mathcal{A}^*(y) + Z - X), \\ \mathcal{A}(dX) & = & -(\mathcal{A}(X) - b), \\ H(ZdX + dZX) & = & -H(ZX - \mu I), \end{array} \right. \tag{2.8}$$

$$\text{GN uses extra info} \quad K(ZdX + dZX) = -K(ZX - \mu I), \tag{2.8}$$

where

$$\begin{array}{rcll} H(M) & := & \frac{1}{2}[M + M^T], & \text{(symmetric part);} \\ K(M) & := & \frac{1}{2}[M - M^T], & \text{(skew-symmetric part).} \end{array}$$

matrix formulation

$$J_{gn}d = P \begin{bmatrix} J_{aho} \\ J_k \end{bmatrix} d = - \begin{bmatrix} F_s \\ F_k \end{bmatrix}, \tag{2.9}$$

permutation of the rows P

J_{gn} - Jacobian of the Gauss-Newton system

J_{aho} - Jacobian of the symmetric (AHO) system

J_k - skew-symmetric component

$$\begin{aligned} \bar{n} &:= t(n) + m + t(n), & (2.9) \\ \bar{m} &:= t(n) + m + n^2, & (2.9) \\ r &:= \bar{m} - \bar{n} = t(n) - t(n-1). & (2.9) \end{aligned}$$

J_{gn} is $(\bar{m} \times \bar{n})$, J_{aho} is $(\bar{n} \times \bar{n})$, and J_k is $(r \times \bar{n})$.

ordering of the singular values

$$\sigma_{\max}(J) := \sigma_1(J) \geq \sigma_2(J) \geq \ldots \geq \sigma_n(J) =: \sigma_{\min}(J).$$

Lemma 2.3.5.1 *The singular values of J_{gn} and of J_{aho} are related by $\sigma_k(J_{gn}) \geq \sigma_k(J_{aho}) \geq \sigma_{k+t(n-1)}(J_{gn})$ for $1 \leq k \leq \bar{n}$.*

Proof. Follows from *interlacing* of singular values. [34]. ■

Gauss-Newton Jacobian is no closer to singularity than the AHO Jacobian is.

Lemma 2.3.5.2 *The largest singular value of the Gauss-Newton Jacobian is bounded by $\sigma_{\max}(J_{gn}) \leq \sqrt{\sigma^2_{\max}(J_{aho}) + \sigma^2_{\max}(J_k)}$.*

Gauss-Newton system is nonsingular and has a bounded condition number as we approach the optimal solution. (as with AHO direction; but not with NT nor HKM nor original Newton system)

(assuming backward-stable algorithm) the (numerical) direction $\widehat{dv}$ is the exact solution to a nearby problem [28]

$$(J_s - \delta J_s)\widehat{dv} = -(f + \delta f),$$

with bounded relative error

$$\frac{\|\widehat{dv} - dv\|}{\|dv\|} \leq \frac{\kappa(J_s)}{1 - \kappa(J_s)\frac{\|\delta J_s\|}{\|J_s\|}} \left(\frac{\|\delta J_s\|}{\|J_s\|} + \frac{\|\delta f\|}{\|f\|} \right). \tag{2.10}$$

Lemma 2.3.5.3 *Suppose that $\mathcal{A}$ is surjective, that X, y, Z is on the central path with $F_d = 0, F_p = 0$ and $ZX - \mu I = 0, \ \mu > 0$. Suppose that the new target for the barrier parameter is $\tau\mu$, where $0 < \tau < 1$. Then the Gauss-Newton direction from coincides with the HKM direction.*

This is the only case where the Gauss-Newton direction coincides with other directions of the Monteiro-Zhang family.

2.3.6 Invariance

Lemma 2.3.6.1 *(Q-scale invariance) Let (dX, dy, dZ) be the Gauss-Newton direction obtained at point $(X, y, Z) \in \mathbb{S}^n_{++} \times \mathbb{R}^m \times \mathbb{S}^n_{++}$. Consider the scaled primal-dual pair obtained using orthogonal P. Then the scaled vector $(PdXP^T, dy, P^{-t}dZT^{-1})$ is the Gauss-Newton direction at the iterate $(\widetilde{X}, \widetilde{y}, \widetilde{Z})$ for the scaled problem.*

Theorem 2.3.6.1 *The Gauss-Newton direction, at any point $v_c \in \mathbb{V}$ is invariant under affine transformation of the variables $w = H(v) + h$ where $H : \mathbb{V} \to \mathbb{V}$ is nonsingular.*

2.3.7 Convergence of Gauss-Newton Method

semidefinite program pair

$$\min\left\{\langle C, X\rangle \mid \mathcal{A}(X) = b, X \succeq 0\right\},$$

$$\max\left\{\langle b, y\rangle \mid \mathcal{A}^*(y) + Z = C, Z \succeq 0\right\}.$$

merit function: (combined norms infeasibility and complementarity)

$$\varphi(v, \mu) := \frac{1}{2}\langle F_\mu(v), F_\mu(v)\rangle. \tag{2.11}$$

Generic Gauss-Newton based interior-point code

1. • Given $\epsilon > 0$ *Tolerance*

   • $k := 1; X^{(k)} := I; Z^{(k)} := I; y^{(k)} := 0; \mu^{(k)} := \frac{\langle Z^{(k)}, X^{(k)} \rangle}{n} = 1;$

   • **WHILE** $\max \left\{ \varphi(v^{(k)}, \mu^{(k)} \right\} > \epsilon$

   •• $d^{(k)} = -F'_\mu(v^{(k)}) F_\mu(v^{(k)});$ *Gauss-Newton direction*

   •• $\alpha^{(k)} := \text{LineSearch}(v^{(k)}, d^{(k)}, \mu^{(k)});$ *Decrease φ*

   •• $\mu^{(k+1)} := \text{TargetSelect}(v^{(k)}, d^{(k)}, \alpha^{(k)}, \mu^{(k)});$

   $\mu^{(k+1)} = \tau \mu^{(k)}, 0 \leq \tau < 1$

   •• $v^{(k+1)} := v^{(k)} + \alpha^{(k)} d^{(k)};$ *New iterate*

   •• $k := k + 1;$

   ENDWHILE

Now: assumptions for convergence analysis

- $\exists v^0$ satisfying centrality condition
- operator $\mathcal{A}$ is surjective.
- optimal primal-dual pair is unique and satisfies strict complementarity (i.e. $Z + X \succ 0$).

-algorithm approximately follows the central path by attempting to solve $F_\mu(X, y, Z) = 0$ for decreasing values of μ

-major difference: the scalar μ is not updated using the iterates, but is reduced by a factor $\tau < 1$ at every step ($\mu \leftarrow \tau\mu$);
no attempt is made to dampen the step to maintain the iterates within the cone of positive definite matrices. (The algorithm only maintains the weaker full rank condition on the Jacobian.)

Lemma 2.3.7.1 *The operator $F'_{\tau\mu}(s)$ is Lipschitz continuous with constant 1.* ∎

Theorem 2.3.7.1 *Let $\sigma_{\min}$ and $\sigma_{\max}$ be, respectively, the smallest and largest singular value of $F'_{\tau\mu}(s_{\tau\mu})$. There is a $\delta > 0$ such that for all s_c such that $\|s_c - s_{\tau\mu}\| < \delta$, the Gauss-Newton step*

$$s_+ = s_c - F'_{\tau\mu}(s_c)^\dagger F_{\tau\mu}(s_c)$$

is well-defined and converges to $s_{\tau\mu}$ at a rate such that

$$\|s_+ - s_{\tau\mu}\| \leq \frac{1}{\sigma_{\min}} \|s_c - s_{\tau\mu}\|^2.$$

Moreover, we can choose δ as long as $\delta < \frac{\sigma_{\min}}{2}$. ∎

Corollary 2.3.7.1 *Let $\sigma_{\min}$ and $\sigma_{\max}$ be, respectively, the smallest and largest singular value of $F'_{\tau\mu}(s_{\tau\mu})$. There is a $\delta > 0$ where for all s_c such that $\|s_c - s_{\tau\mu}\| < \delta$,*

$$\|F_{\tau\mu}(s_+)\| \leq \frac{1}{2}\|F_{\tau\mu}(s_c)\|.$$

Moreover, we can choose any δ such that

$$\delta < \frac{\sigma_{\min}^2}{8\sigma_{\max}}. \tag{2.12}$$

$\blacksquare$

general idea:
from an iterate s_k , "close enough" to s_μ, we can choose a target on the central path $s_{\tau\mu}$ in such a way that the next iterate s_{k+1}, obtained from the Gauss-Newton direction, is now "close enough" to $s_{\tau\mu}$ for the process to be repeated.

First we estimate the distance between two points on the central paths in terms of the required radius of convergence.

Lemma 2.3.7.2 *Let $\sigma_{\min}$ and $\sigma_{\max}$ be, respectively, the smallest and largest singular value of $F'_\tau(s_{\tau\mu})$. Let s_μ and $s_{\tau\mu}$ be the optima on the central path.*

1. *If we choose $0 < \tau < 1$ such that*

$$1 - \tau \leq \frac{\sigma^2_{\min}}{8\sqrt{n}\mu}, \tag{2.13}$$

   then

$$\|s_\mu - s_{\tau\mu}\| \leq \frac{1}{2}\left(\frac{\sigma_{\min}}{2}\right), \tag{2.14}$$

   which implies s_μ is within half of the radius of quadratic convergence of $s_{\tau\mu}$.

2. *If we choose $0 < \tau < 1$ such that*

$$1 - \tau \leq \frac{\sigma_{\min}^3}{32\sqrt{n}\mu\sigma_{\max}}, \tag{2.15}$$

then

$$\|s_\mu - s_{\tau\mu}\| \leq \frac{1}{2}\left(\frac{\sigma_{\min}^2}{8\sigma_{\max}}\right). \tag{2.16}$$

In this case s_μ is within half of the radius of guaranteed constant decrease of the merit function in (2.12) in Corollary 2.3.7.1.

■

We now estimate the distance to the new target after a Gauss-Newton step.

Lemma 2.3.7.3 *Let $\sigma_{\min}$ and $\sigma_{\max}$ be, respectively, the smallest and largest singular value of $F'_\tau(s_{\tau\mu})$. Let s_μ and $s_{\tau\mu}$ be optima on the central path. Suppose that the point s_c is well-centered in the sense that*

$$\|s_\mu - s_c\| \quad \leq \quad \min\left\{\frac{\sigma_{\min}}{4}, \frac{\sigma_{\min}}{16\sigma_{\max}}\right\},$$

and we choose τ to satisfy

$$0 < \tau < 1, \quad 1 - \tau \leq \min\left\{\frac{\sigma_{\min}^2}{8\sqrt{n}\mu}, \frac{\sigma_{\min}^3}{32\sqrt{n}\mu}\right\}, \tag{2.16}$$

as in Lemma 2.3.7.2. Then, after one Gauss-Newton step, the new point s_+ will be within half the radius of convergence of $s_{\tau\mu}$, i.e.

$$\|s_{\tau\mu} - s_+\| \quad \leq \quad \frac{\sigma_{\min}}{4}. \tag{2.16}$$

Moreover, the merit function is reduced

$$\|F_\tau(s_+)\| \leq \frac{1}{2}\|F_\tau(s_c)\|. \tag{2.16}$$

$\blacksquare$

Main Convergence Theorem:

Theorem 2.3.7.2 *Suppose that we are given a tolerance* $\epsilon > 0$*, an initial barrier parameter* $\mu_0 > \epsilon$*, and* $Z_0, X_0 \in \mathbb{S}^n_{++}$ *such that* $s_0 = (X_0, y_0, Z_0)$ *is a well-centered starting point:* s_0 *is within half the quadratic convergence radius of* s_{μ_0} *in Theorem 2.3.7.1,*

$$\|s_{\mu_0} - s_0\| \leq \frac{1}{2}\left(\frac{\sigma_{\min}}{2}\right). \tag{2.16}$$

Suppose moreover that s_0 *is within half the radius for guaranteed constant decrease of the merit function given in Corollary 2.3.7.1,*

$$\|s_{\mu_0} - s_0\| \leq \frac{1}{2}\left(\frac{\sigma_{\min}^2}{8\sigma_{\max}}\right),$$

where $0 < \sigma_{\min}$ *(respectively* $\sigma_{\max}$*) is smaller than the smallest (respectively larger than the largest) singular value of* $F'_{\omega\mu_0}(s_{\omega\mu_0})$*, for all* $\frac{\epsilon}{\mu_0} < \omega < 1$*.*

If we choose τ (small) satisfying (2.16) in Lemma 2.3.7.3, i.e.

$$\alpha = \min\left\{\frac{\sigma_{\min}^2}{8\sqrt{n}\mu_0}, \frac{\sigma_{\min}^3}{32\sqrt{n}\mu_0}\right\},$$

and $\tau \geq \max\{0, 1-\alpha\}$, $0 < \tau < 1$, then the GN Algorithm converges to $\bar{s}$, ϵ-optimal in the following sense,

$$\tau^k \mu_0 \leq \epsilon, \quad \|F_{\tau^k \mu_0}(\bar{s})\| \leq \epsilon, \quad \|\bar{s} - s_{\tau^k \mu_0}\| \leq 2\sigma_{\min}\epsilon;$$

and the number of iterations, k, depends on τ:

1. $$\mathcal{O}\left(\max\left\{\log\left(\frac{\|F_{\tau\mu_0}(s_0)\|}{\epsilon}\right), \ \log\left(\frac{\mu_0\sqrt{n}}{\epsilon}\right)\right\}\right) \qquad (2.16)$$

   iterations, if $0 < \tau \leq \frac{1}{2}$;

2.

$$\mathcal{O}\left(\max\left\{\log\left(\frac{\|F_{\tau\mu_0}(s_0)\|}{\epsilon}\right),\ \left(\frac{\log\frac{(2\tau-1)\epsilon}{(1-\tau)2\mu_0\sqrt{n}}}{\log\tau}\right),\ \left(\frac{\log\frac{\epsilon}{\mu_0}}{\log\tau}\right)\right\}\right) \tag{2.16}$$

iterations, if $\frac{1}{2}<\tau<1$.

■

Asymptotic Convergence We are also able to specialize a standard result pertaining to the asymptotic convergence rate of the Gauss-Newton method.

Theorem 2.3.7.3 *(Superlinear Convergence) Assume a primal-dual pair with a unique, strictly complementary, optimal solution, denoted v^*. Then for each $c \in (1, \infty)$, there is an $\epsilon > 0$ such that, from $v^{(0)}$ satisfying $\|v^{(0)} - v^*\| < \epsilon$, the sequence generated by the Gauss-Newton method converges to v^* and obeys*

$$\|v^{(k+1)} - v^*\| \leq \frac{c\alpha}{2\sigma_{\min}^2}\|v^{(k)} - v^*\|^2. \tag{2.16}$$

where $\|J(v)\| \leq \alpha$ in the ϵ-neighborhood of v^* and where $\sigma_{\min}$ represent the smallest singular value of $J(v^*)$.

This last result suggests that, whatever upper-level algorithm we use, if we use the Gauss-Newton as the search direction, then as soon as we can estimate $\sigma_{\min}$ accurately, we compute the radius of superlinear convergence and we can set the barrier parameter to zero, ignore the cone constraint and let the algorithm converge to the optimal solution.

2.3.8 Numerical Tests with GN

Iter	$\|\|F_d\|\|$	$\|\|f_p\|\|$	$\langle Z, X\rangle/n$
1	+8.900e+01	+4.175e+01	+8.749e-01
2	+5.433e+01	+3.545e+01	+8.580e-01
3	+1.503e+01	+1.759e+01	+6.269e-01
4	+3.066e+00	+3.465e-03	+3.841e-01
5	+2.131e-02	+4.408e-04	+1.278e-01
6	+5.573e-03	+8.302e-05	+2.873e-02
7	+1.042e-04	+3.153e-06	+1.367e-02
8	+1.611e-05	+2.913e-07	+7.862e-04
9	+7.275e-08	+1.620e-09	+8.680e-05
10	+4.473e-13	+2.041e-14	+1.812e-06
11	+1.017e-14	+2.800e-15	+3.784e-08
12	+1.030e-14	+3.426e-15	+7.902e-10
13	+1.318e-14	+5.184e-15	+1.650e-11
14	+1.138e-14	+4.835e-15	+4.964e-14

One instance of well-conditioned problem. $n = 15, m = 30$

	$-\log_{10}\max\{\|F_d\|, \|f_p\|\}$					$-\log_{10}\langle Z, X\rangle/n$				
	AHO	HKM	NT	GT	GN	AHO	HKM	NT	GT	GN
random	14.7	9.1	10.6	9.5	13.3	12.0	10.7	5.5	12.8	14.4
norm min.	15.0	9.9	10.9	15.3	14.8	13.8	12.7	10.1	14.4	14.9
Maxcut	15.7	9.6	10.6	14.8	15.8	14.4	12.4	9.3	14.4	15.5
Lovàsz θ	14.7	9.2	9.2	13.9	14.8	14.2	13.7	12.8	14.1	15.3

Accuracy of solutions on SDPT3 test problems. Average of one hundred random instances.

r, n, m	AHO			GN		
	iter.	Infeas.	Gap	iter.	Infeas.	Gap
3,10,9	18	14.2	15.1	13	14.3	15.4
6,20,24	22	12.1	14.6	17	13.8	16.3

Ill-conditioned problems' solutions. Average of fifty random instances.

n	m	iter	$\|F_d\|$	$\|f_p\|$	$\langle Z, X\rangle$
10	11	24	9.296152e-13	5.594466e-11	5.842689e-07
20	21	25	2.260052e-12	4.882512e-15	7.906645e-07
30	31	22	4.211995e-12	9.367561e-15	2.190712e-06
40	41	23	1.816565e-12	7.334760e-15	2.301071e-06

Gruber-Rendl ill-posed Problems with $\alpha = 10^{-7}$ and accuracy set to 10^{-5}.

(average iterations for G-R was 115; $\alpha = 0$ means no primal Slater point)

n	m	iter	$\|F_d\|$	$\|f_p\|$	$\langle Z, X\rangle$
10	11	23	2.377587e-12	4.076899e-08	9.096929e-06
20	21	22	2.479800e-12	1.516398e-08	1.050569e-06
30	31	25	5.153522e-12	5.729514e-08	1.416319e-06

with $\alpha = 0$ and accuracy set to 10^{-5}.

n	m	iter	$\|F_d\|$	$\|f_p\|$	$\langle Z, X\rangle$
50	51	32	3.467449e-12	6.469784e-15	-1.233752e-14

with $\alpha = 0$ and increased accuracy.

n	m	iter	$\|F_d\|$	$\|f_p\|$	$\|\langle Z, X\rangle\|$
10	9	7	8.015765e-11	8.009760e-11	5.080324e-06
20	19	8	9.000070e-11	9.001461e-11	1.743560e-06
30	28	8	1.272965e-10	9.000636e-11	2.615344e-06
40	15	5	2.593612e-09	9.999999e-11	1.651607e-07
40	30	5	2.642280e-09	1.000001e-10	3.821364e-07
40	39	5	2.231644e-09	9.999997e-11	5.396936e-07
50	49	5	2.928152e-09	9.999996e-11	6.840107e-07

no primal or dual Slater points: accuracy set to 10^{-5}.

(G-R needed average of 52 iterations)

2.3.9 Solving SDPs using NLP

Though the current popular approaches for solving SDP use extensions of interior-point methods from LP (successful approaches), there are several that use NLP techniques, e.g. Monteiro-Zhang... or Vanderbei... details????

However, we use the primal-dual framework with known results from NLP to derive the Gauss-Newton method.

Provide convergence results based on standard NLP approaches (redo KW paper using $\cos\theta$ bounded below and Wolfe conditions)

2.4 Exploiting Sparsity

2.4.1 Special Structure

2.4.2 Bundle Trust Methods

2.4.3 Potential Function Methods

3 Applications

3.1 Outline

- Quadratic Boolean Programming
- Quadratically Constrained Quadratic Programs
- Trust Region Algorithms for NLP
- Sequential Quadratic Programming for NLP
- Historical Notes

3.2 Quadratic Boolean Programming

3.2.1 Outline

- Tractable Relaxations
- Strengthened SDP relaxations
- Strengthened Bounds for Max-Cut
- Historical Notes

3.2.2 Tractable Relaxations of Max-Cut

(quadratic boolean programming)

$$(MCQ) \qquad \mu^* := \max_{x\in\mathcal{F}} q_0(x) \quad (:= x^TQx - 2c^Tx).$$

where $\mathcal{F} = \{\pm 1\}^n$

---

perturbing diagonal of Q on $\mathcal{F}$:

$$\begin{aligned} q_u(x) \quad &:= \quad x^T(Q + \mathrm{Diag}\,(u))x - 2c^Tx - u^Te \\ &= \quad q_0(x), \quad \forall x \in \mathcal{F}, \end{aligned}$$

where e vector of ones.

Bound 0: trivial bound from diagonal perturbations:

$$\mu^* \leq f_0(u) := \max_x q_u(x).$$

The function f_0 can take on the value $+\infty$. Let

$$S := \left\{u : u^T e = 0, Q + \operatorname{Diag}(u) \preceq 0\right\}.$$

Then:

$$\boxed{\begin{array}{c} \mu^* \leq B_0 := \min_u f_0(u) \\ \left(= \min_{u^T e=0} f_0(u), \text{ if } S \neq \emptyset\right). \end{array}}$$

Using the hidden semidefinite constraint:

$$\mu^* \leq B_0 = \min_{Q+\operatorname{Diag}(u) \preceq 0} f_0(u).$$

Bound 1: relax the feasible set to the sphere of radius $\sqrt{n}$ (tractable trust region subproblem, TRS):

$$\mu^* \leq f_1(u) := \max_{||x||^2=n} \; q_u(x)$$

and

$$\boxed{\mu^* \leq B_1 := \min_u f_1(u).}$$

The inner maximization problem is called a trust region subproblem and is tractable. This bound provides the central tool in the proofs of equivalence.

Bound 2: box constraint:

$$\mu^* \leq f_2(u) := \max_{|x_i| \leq 1} \; q_u(x).$$

add the semidefinite constraint to make bound tractable.

$$\mu^* \leq \min_u f_2(u)$$

and

$$\boxed{\mu^* \leq B_2 := \min_{Q + \mathrm{Diag}\,(u) \preceq 0} f_2(u).}$$

Bound B_1^c: lift to eigenvalue bound:

$$Q^c := \begin{bmatrix} 0 & -c^T \\ -c & Q \end{bmatrix}$$

$$q_u^c(y) := y^T(Q^c + \operatorname{diag}(u))y - u^T e$$

$$\mu^* \leq f_1^c(u) := \max_{||y||^2=n+1} q_u^c(y)$$

where

$$\max_{||y||^2=n+1} q_u^c(y) = (n+1)\lambda_{\max}(Q^c + \operatorname{diag}(u)) - u^T e$$

$$\boxed{\mu^* \leq B_1^c := \min_u f_1^c(u).}$$

Similarly, equivalent bnds B_0^c and homog. bnds for other models.

Bound B_3: SDP bound:

After homogenization ($c = 0$), use

$$x^T Q x = \operatorname{trace} x^T Q x = \operatorname{trace} Q x x^T$$

and, for $x \in \mathcal{F}$, $y_{ij} = x_i x_j$ defines a symmetric, rank one, positive semidefinite matrix Y with diagonal elements 1. Relax the rank one condition.

$$\boxed{\begin{array}{lcc} B_3 := & \max & \operatorname{trace} QY \\ & \text{subject to} & \operatorname{diag}(Y) = e \\ & & Y \succeq 0. \end{array}}$$

Summary: ref. Polyak-Rendl-Wolkowicz (PRW) 1995. (without restrictions e.g. $u^T e = 0$)

$$
\begin{aligned}
B_0 &= \min_u \max_x q_u(x) \\
B_1 &= \min_u \max_{x^T x = n} q_u(x) \\
B_2 &= \min_u \max_{-1 \leq x_i \leq 1} q_u(x) \\
B_3 &= \max\{\operatorname{trace} Q^c Y : \operatorname{diag}(Y) = e,\ Y \succeq 0.\} \\
B_1^c &= \min_u \max_{y^T y = n+1} q_u^c(y)
\end{aligned}
$$

Now replace ± 1 constraints with $x_i^2 = 1, \forall i$.

$$(P_E) \qquad \max \qquad q_0(x) = x^T Q x - 2c^T x$$
$$\text{subject to} \qquad x_i^2 = 1, \quad i = 1, \cdots, n.$$

B_L denotes Lagrangian relaxation bound.

Following the theme: "Lagrangian relaxation is best"

$\boxed{\textbf{Theorem} \quad B_L \text{ equals all above bounds.}}$

ref PR, PRW/95.
(The proofs come from exploiting strong Lagrangian duality of TRS.)

3.2.3 Strengthened SDP Bounds for Max-Cut Problem

(ref Anjos-Wolkowicz/99 -Illustration of the recipe for SDP relaxation. based on a second lifting - motivated by strong duality results that follow from adding redundant constraints of type XX^T, ref. Anstreicher-Wolkowicz/98 and below)

first lifting procedure

$$X = xx^T, \quad \operatorname{Diag} X = e$$

implies second lifting procedure

$$X^2 = xx^T xx^T = nX.$$

and provides an equivalent quadratic matrix model for MC

MCQ is equivalent to:

$$\begin{array}{rll} \mu^* := & \max & \operatorname{trace} QX \\ & \text{s.t.} & \operatorname{diag}(X) = e \\ & & X \succeq 0 \\ & & X \text{ is rank } 1. \end{array}$$

which is equivalent to

$$\begin{array}{rll} \mu^* := & \max & \operatorname{trace} QX \\ & \text{s.t.} & \operatorname{diag}(X) = e \\ & & X^2 - nX = 0 \end{array}$$

Since X and X^2 can be mutually diagonalized.

Now add redundant constraints and get:

$$\begin{array}{rll} \mu^* := & \max & \operatorname{trace} QX \\ & \text{s.t.} & \operatorname{diag}(X) = e \\ & & X^2 - nX = 0 \\ & & X \circ X = E \end{array}$$

where E is the matrix of ones.

Note: $\left(X^2 = nX,\ \operatorname{trace} X = n\right) \Rightarrow (\operatorname{rank} X = 1)$. But, nonconvex problem too hard to solve in general. Need Lagrangian relaxation with vectors dimension $t(n) = n(n+1)/2$.

Recipe for SDP relaxations:

1. add redundant constraints
2. take Lagrangian dual
3. homogenize
4. use hidden semidefinite constraint to obtain SDP equivalent
   (check Slater's constraint qualification - strict feasibility)
5. take Lagrangian dual again
6. check Slater's CQ again - project if it fails
7. delete redundant constraints

Notation:

(Interesting operators and adjoints –fun for me, if for no one else)

- $S \in \mathcal{S}^n, \quad t(n) = \frac{n(n+1)}{2}$
- $s = \mathrm{svec}\,(S) \in \mathbb{R}^{t(n)}$ vector formed (columnwise) from S ignoring strictly lower triang.
- $S = \mathrm{sMat}\,(s)$ inverse of svec
- $\mathrm{hMat}\,(v)$ is adjoint of svec, off-diagonal terms are multiplied by a half
- $\mathrm{dsvec}\,(S)$ is adjoint of sMat, operator like svec but off diagonal elements are multiplied by 2
- $\mathrm{sdiag}\,(s) := \mathrm{diag}\,(\mathrm{sMat}\,(s))$
- $\mathrm{vsMat}\,(s) := \mathrm{vec}\,(\mathrm{sMat}\,(s))$
- $$\mathrm{vsMat}^*(s) = \mathrm{dsvec}\,\left(\left(\mathrm{Mat}\,(v) + \mathrm{Mat}\,(v)^T\right)/2\right).$$

start with equivalent program to MCQ:

$$
\text{MC2} \qquad \begin{array}{rll} \mu^* = & \max & \operatorname{trace} QX \\ & \text{s.t.} & \operatorname{diag}(X) = e \\ & & X \circ X = E \\ & & X^2 - nX = 0, \end{array} \tag{3.1}
$$

to efficiently apply Lagrangian relaxation and not lose information from the linear constraint, replace the constraint with the norm constraint $||\operatorname{diag}(X) - e||^2 = 0$. Add $1 - y_0^2 = 0$.

We use: $X = \operatorname{sMat}(x)$ is a symmetric matrix.

$$
\begin{array}{ll}
\max & \operatorname{trace}\left(Q\,\operatorname{sMat}(x)\right) y_0 \\
\text{s.t.} & \operatorname{sdiag}(x)^T \operatorname{sdiag}(x) - 2e^T \operatorname{sdiag}(x) y_0 + n = 0 \\
 & \operatorname{sMat}(x) \circ \operatorname{sMat}(x) = E \\
 & \operatorname{sMat}(x)^2 - n\,\operatorname{sMat}(x) y_0 = 0 \\
 & 1 - y_0^2 = 0.
\end{array}
$$

take the Lagrangian dual; Lagrange multipliers w, T, S

$$
\begin{aligned}
\mu^* \leq & \\
\nu_2^* := & \min_{w,T,S} \max_{x, y_0^2=1} \operatorname{trace}\left(Q \operatorname{sMat}(x)\right) y_0 \\
& \quad + w(\operatorname{sdiag}(x)^T \operatorname{sdiag}(x) - 2e^T \operatorname{sdiag}(x) y_0 + n) \\
& \qquad + \operatorname{trace} T(E - \operatorname{sMat}(x) \circ \operatorname{sMat}(x)) \\
& \qquad\quad + \operatorname{trace} S((\operatorname{sMat}(x))^2 - n \operatorname{sMat}(x) y_0).
\end{aligned}
$$

move variable y_0 into the Lagrangian without increasing the duality gap since this is a trust region subproblem

$$
\begin{aligned}
\nu_2^* = \quad & \min_{t,w,S} \max_{x,y_0} \operatorname{trace}\,(Q\operatorname{sMat}(x))\, y_0 \\
& + w(\operatorname{sdiag}(x)^T \operatorname{sdiag}(x) - 2e^T \operatorname{sdiag}(x) y_0 + n) \\
& + \operatorname{trace} T(\operatorname{sMat}(x) \circ \operatorname{sMat}(x) - E) \\
& + \operatorname{trace} S((\operatorname{sMat}(x))^2 - n \operatorname{sMat}(x) y_0) \\
& + t(1 - y_0^2).
\end{aligned}
$$

The inner maximization of the above relaxation is an unconstrained pure quadratic maximization, i.e. the optimal value is infinity unless the Hessian is negative semidefinite **(hidden constraint)** is which case $x = 0$ is optimal.

Therefore we need to evaluate the Hessian of the Lagrangian.

Using $Q\mathrm{sMat}\,(x) = x^T \mathrm{dsvec}\,(Q)$, and adding a 2 for convenience, we get the constant part (no Lagrange multipliers) of the Hessian:

$$2H_c := 2 \begin{pmatrix} 0 & \frac{1}{2}\mathrm{dsvec}\,(Q)^T \\ \frac{1}{2}\mathrm{dsvec}\,(Q) & 0 \end{pmatrix}.$$

nonconstant part:

Use:

$$\operatorname{dsvec}\operatorname{Diag}\operatorname{diag}\operatorname{sMat} = \operatorname{sdiag}^*\operatorname{sdiag} = \operatorname{Diag}\operatorname{svec}(I)$$

rewrite the quadratic forms as follows:

$$\operatorname{sdiag}(x)^T \operatorname{sdiag}(x) = x^T \left(\operatorname{dsvec} \operatorname{Diag} \operatorname{diag} \operatorname{sMat}\right) x;$$

$$e^T \operatorname{sdiag}(x) = (\operatorname{dsvec} \operatorname{Diag} e)\, x;$$

$$\begin{aligned} \operatorname{trace} S(\operatorname{sMat}(x))^2 &= \operatorname{trace} \operatorname{sMat}(x)\, S \operatorname{sMat}(x) \\ &= x^T \operatorname{dsvec}\left(S \operatorname{sMat}(x)\right) \\ &= x^T \left(\operatorname{dsvec} S \operatorname{sMat}\right) x; \end{aligned}$$

$$\begin{aligned} &\operatorname{trace} T(\operatorname{sMat}(x) \circ \operatorname{sMat}(x)) \\ &\quad = x^T \left\{\operatorname{dsvec}\left(T \circ \operatorname{sMat}(x)\right)\right\} \\ &\quad = x^T \left(\operatorname{dsvec}\left(T \circ \operatorname{sMat}\right)\right) x. \end{aligned}$$

use the *negative* of the Hessian and split it into four linear operators with the factor 2:

$$
\begin{aligned}
2\mathcal{H} &:= 2\mathcal{H}(w) + 2\mathcal{H}(T) + 2\mathcal{H}(S) + 2\mathcal{H}(t) \\
&:= 2w \begin{pmatrix} 0 & (\operatorname{dsvec} \operatorname{Diag} e)^T \\ (\operatorname{dsvec} \operatorname{Diag} e) & -\operatorname{sdiag}^* \operatorname{sdiag} \end{pmatrix} \\
&\quad + 2 \begin{pmatrix} 0 & 0 \\ 0 & \operatorname{dsvec}\,(T \circ \operatorname{sMat}\,) \end{pmatrix} \\
&\quad + 2 \begin{pmatrix} 0 & \frac{1}{n}\operatorname{dsvec}\,(S)^T \\ \frac{1}{n}\operatorname{dsvec}\,(S) & -\operatorname{dsvec} S \operatorname{sMat} \end{pmatrix} \\
&\quad + 2t \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.
\end{aligned}
$$

The matrix $\operatorname{sdiag}^*\operatorname{sdiag} \in \mathcal{S}^{t(n)}$ is diagonal with elements determined using

$$\begin{aligned} e_i^T (\operatorname{sdiag}^*\operatorname{sdiag})\, e_j &= \operatorname{sdiag}(e_i)^T \operatorname{sdiag}(e_j) \\ &= \begin{cases} 1 & \text{if i=j=t(k)} \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

Similarly, we find that, for $T = \sum_{ij} t_{ij} E_{ij}$, where the matrices E_{ij} are the elementary matrices $e_i e_j^T + e_j e_i^T$, we have

$$\operatorname{dsvec}(T \circ \operatorname{sMat}) = \sum_{ij} t_{ij} \operatorname{dsvec}(E_{ij} \circ \operatorname{sMat}).$$

cancel the 2; get (equivalent to the Lagrangian dual)

$$\text{MCDSDP2} \qquad \begin{array}{rll} \nu_2^* = & \min & nw + \operatorname{trace} ET + \operatorname{trace} 0S + t \\ & \text{s.t.} & \mathcal{H}(w,T,S,t) \succeq H_c \end{array}$$

take T sufficiently positive definite and t sufficiently large; get Slater's constraint qualification!

take dual; get strengthened SDP relaxation of MC:

$$\text{MCPSDP2} \qquad \begin{array}{rll} \nu_2^* = & \max & \operatorname{trace} H_c Y \\ & \text{s.t.} & \mathcal{H}^*(Y) = n \\ & & \mathcal{H}^*(Y) = E \\ & & \mathcal{H}^*(Y) = 0 \\ & & \mathcal{H}^*(Y) = 1 \\ & & Y \succeq 0. \end{array}$$

We need to calculate the adjoint operators and remove redundant constraints in MCDSDP2. Use:

$$Y \cong \begin{pmatrix} y_0 \\ x \end{pmatrix} \begin{pmatrix} y_0 & x^T \end{pmatrix}, \quad X = \mathrm{sMat}\,(x)$$

simplified SDP relaxation MCPSDP2

$$\begin{array}{ll} \max & \operatorname{trace} H_c Y \\ \text{s.t.} & \operatorname{diag}(Y) = e \\ & Y_{0,t(i)} = 1, \quad \forall i = 1, \ldots, n \\ & \sum\limits_{k=1}^{i} Y_{t(i-1)+k,t(j-1)+k} \\ & \quad + \sum\limits_{k=i+1}^{j} Y_{t(k-1)+i,t(j-1)+i} \\ & \qquad + \sum_{k=j+1}^{n} Y_{t(i-1)+i,t(k-1)+j} \\ & \qquad\quad - nY_{0,t(j-1)+i} = 0 \\ & \qquad\qquad \forall 1 \leq i < j \leq n \\ & Y \succeq 0, Y \in S^{t(n)+1}. \end{array}$$

This problem has $2t(n) - 1$ constraints.

surprise result

LEMMA

Suppose that Y is feasible in MCPSDP2. Then the first row

$$\mathrm{sMat}\,\left(Y_{0,1:t(n)}\right) \succeq 0.$$

Proof: Let $x = Y_{0,2:t(n)}$. $X^2 = nX$ constraint:

$$n\mathrm{sMat}\,(x) = \mathrm{sMat}\,(x)\mathrm{sMat}\,(x) = \mathrm{sMat}\,(x)x^T\mathrm{sMat}^{\,*}.$$

identify xx^T with lower right block of Y; get congruence of a positive semidefinite matrix. Alternatively: using $\mathcal{H}_3^*$, and $\mathrm{dsvec}\,(\cdot)\mathrm{sMat}$ is self-adjoint operator

$$n\mathrm{dsvec}^{\,*}(x) = \mathrm{dsvec}\,\bar{Y}\mathrm{sMat} = \left(\mathrm{dsvec}\,\bar{Y}\mathrm{sMat}\,\right)^*,$$

where $\bar{Y}$ is the bottom right block of Y. Again congruence. ■

Second Strengthened SDP Relaxation

add more redundant quadratic constraints to MC2; dual provides tighter relaxation

change of variable $X = vv^T$ gives $X_{ij} = v_i v_j$ and $v_k^2 = 1$ for $k = 1, \dots, n$; therefore

$$X_{ij} = v_i v_j = v_i v_k^2 v_j = v_i v_k \cdot v_k v_j = X_{ik} \cdot X_{kj}$$

interesting connection between these constraints and the metric polytope.

Add these constraints to MC2

$$
\text{(MC3)} \qquad
\begin{array}{rll}
\mu^* = & \max & \operatorname{trace} QX \\
 & \text{s.t.} & \operatorname{diag}(X) = e \\
 & & X \circ X = E \\
 & & X^2 - nX = 0 \\
 & & X_{ij} = X_{ik}X_{kj}, \quad \forall\, 1 \leq i, j, k \leq n.
\end{array}
$$

Taking the dual of the dual of MC3 (and removing redundant constraints in the resulting SDP) yields the relaxation SDP3 defined below.

Alternatively, for rank-one matrices $X = vv^T, v \in \{\pm 1\}^n$. have all their entries equal to ± 1. Hence Y feasible for SDP2 have all their entries in the first row and column equal to ± 1. Now consider the following constraints from SDP2:

$$Y_{0,T(i,j)} = \frac{1}{n} \sum_{k=1}^{n} Y_{T(i,k),T(k,j)}, \quad \forall\, 1 \le i < j \le n,$$

for $Y = \begin{pmatrix} 1 & x^T \\ x & \bar{Y} \end{pmatrix}$ and $x = \operatorname{svec}(vv^T)$. The entry $Y_{0,T(i,j)}$ is in the first row of Y and therefore it is equal to 1 in magnitude; and so they must all have magnitude equal to 1.

Either approach yields the relaxation SDP3:

$$
\text{(SDP3)} \qquad
\begin{array}{rll}
\nu_3^* = & \max & \operatorname{trace} H_Q Z \\
& \text{s.t.} & \operatorname{diag}(Z) = e \\
& & Z_{0,t(i)} = 1, i = 1, \ldots, n \\
& & Z_{0,T(i,j)} = Z_{T(i,k),T(k,j)}, \quad \forall\, k, \forall\, 1 \leq i < j \leq n \\
& & Z \succeq 0, Z \in \mathcal{S}^{t(n)+1}.
\end{array}
$$

Properties of the Second Strengthened Relaxation

Let us define the projection of the feasible set of SDP3 onto $\mathcal{S}^n$ as

$$F_n := \{X \in \mathcal{S}^n : X = \mathrm{sMat}\,(Z_{1:t(n),0}), Z \text{ feasible for SDP3}\}.$$

Corollary 3.2.3.1 $C_n \subseteq F_n \subseteq \mathcal{E}_n \cap M_n$. ■

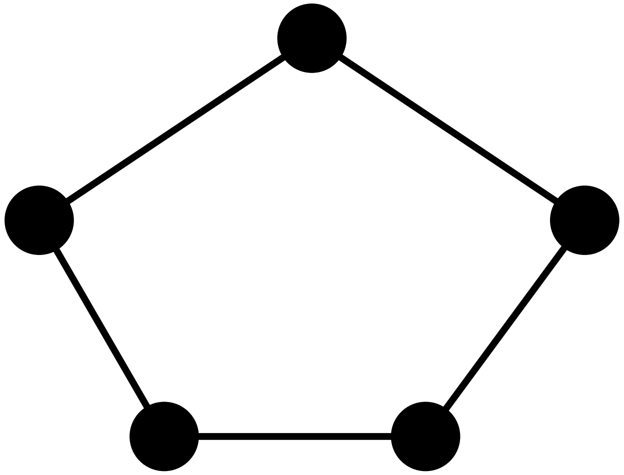
(cut polytope, SDP3, elliptope, metric polytope)

containment is *strict*

Theorem 3.2.3.1 $C_n \subsetneq F_n \subsetneq \mathcal{E}_n \cap M_n$ *for* $n \geq 5$. ■

Numerical Example

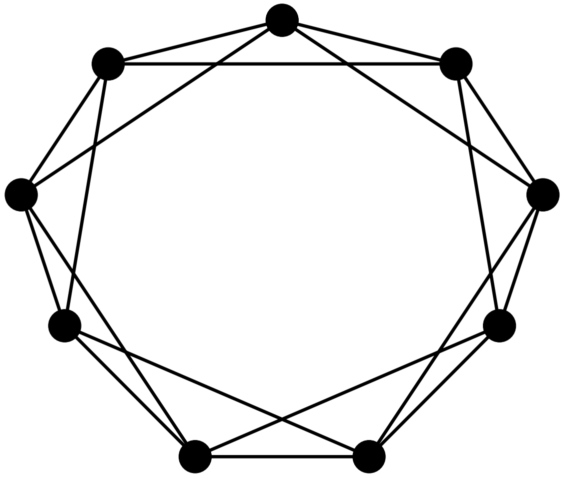
Graph:



μ^*	SDP1	SDP2
4	4.5225 R.E.: 13.06%	4.0000 R.E.: 0%

Numerical Example

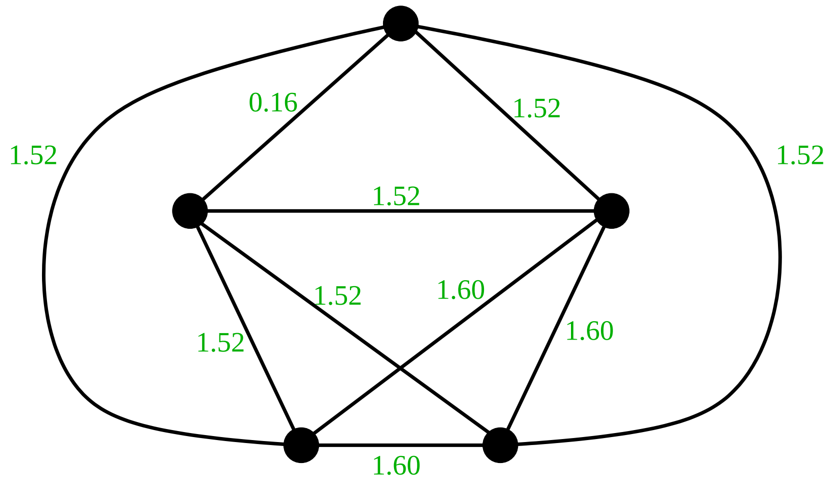
Graph:



μ^*	SDP1	SDP2
12	13.5000 R.E.: 12.50%	12.4967 R.E.: 4.14%

Numerical Example

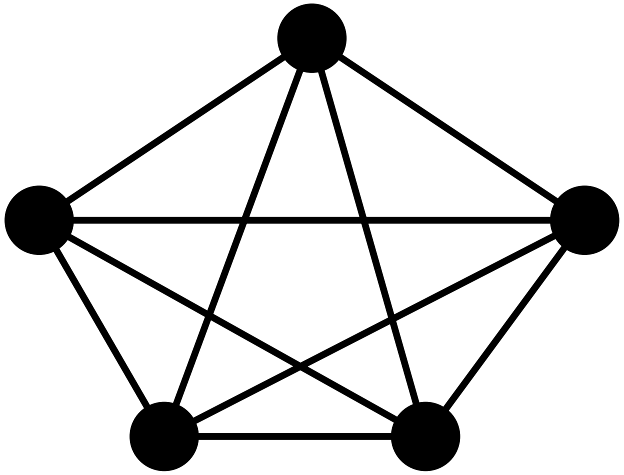
Graph:



μ^*	SDP1	SDP2
9.28	9.6040 R.E.: 3.49%	9.2800 R.E.: 0%

Numerical Example

Graph:



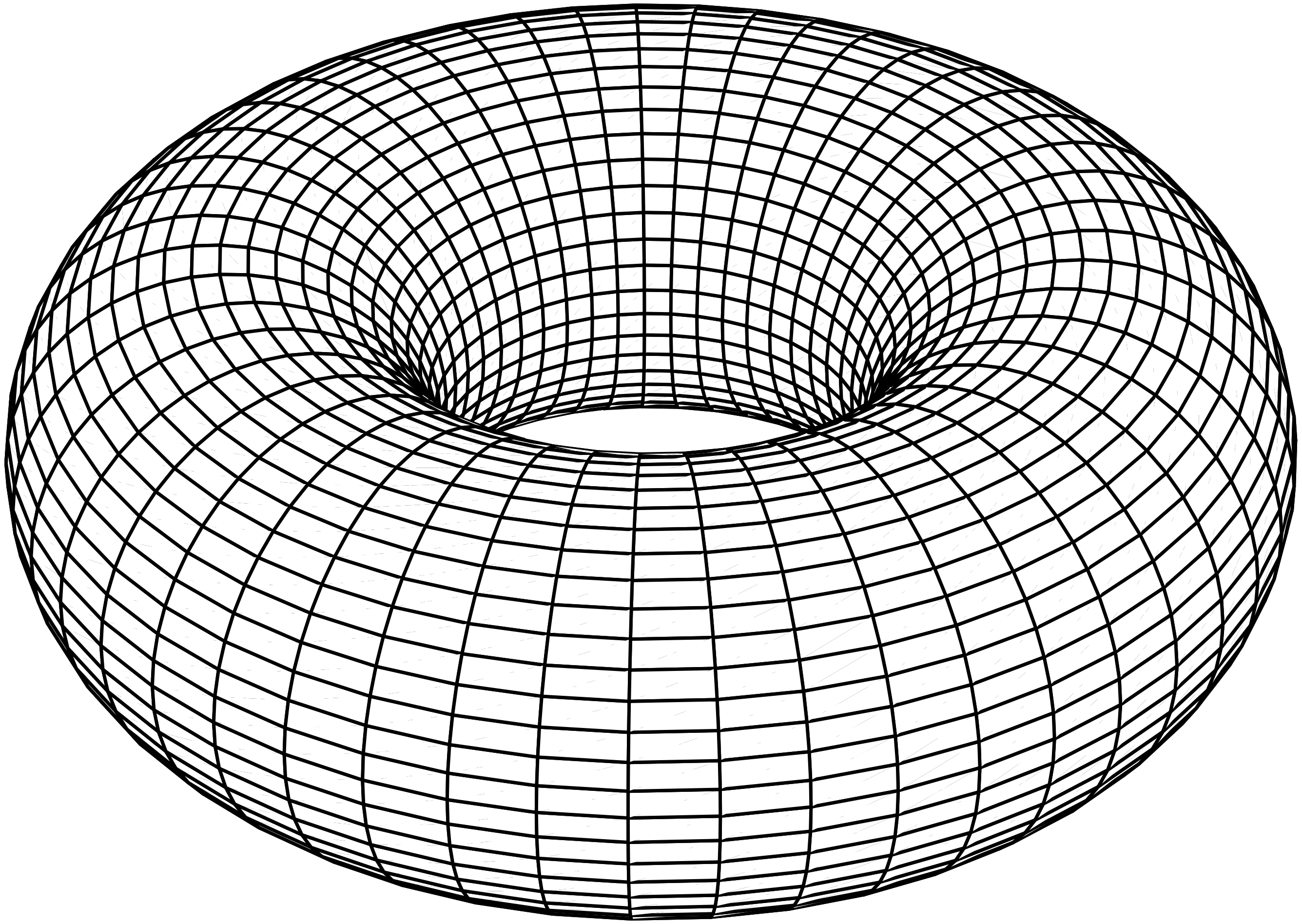
μ^*	SDP1	SDP2
6	6.2500 R.E.: 4.17%	6.2500 R.E.: 4.17%

Ising spin glass problems

All problems have short range interactions only,

periodic boundary conditions

and were generated using the `torusgen` codes from Jünger et al.



Problem type	Number of particles	SDP1 $\frac{\mu^*}{\nu_1^*}$	SDP2 $\frac{\mu^*}{\nu_2^*}$	SDP1 found rank-1 optimum	SDP2 found rank-1 optimum
2-D Gaussian	12	0.9825	1.0000	No	Yes
	20	0.9775	1.0000	No	Yes
	20	0.9666	1.0000	No	Yes
	24	0.9288	1.0000	No	Yes
	16	0.9985	1.0000	No	Yes
3-D Gaussian	16	0.9999	1.0000	No	Yes
	18	0.9720	1.0000	No	Yes
	12	0.9730	1.0000	No	Yes
	16	0.9192	1.0000	No	Yes
	20	0.9942	1.0000	No	Yes
2-D ± 1	12	0.8323	1.0000	No	No
	16	0.9205	1.0000	No	No
	16	0.7379	1.0000	No	No
	16	0.9859	1.0000	No	Yes
	20	0.8863	1.0000	No	No
3-D ± 1	12	0.9951	1.0000	No	Yes
	18	0.9470	1.0000	No	Yes
	18	0.8711	1.0000	No	No
	18	0.9075	1.0000	No	No
	18	0.9642	1.0000	No	Yes
	18	0.8428	1.0000	No	Yes

Observations:

1. SDP2 *always* found the weight of the optimal cut, for all problems.

2. Furthermore, SDP2 seems to often find an optimal rank-one matrix and hence an optimal cut. In fact, SDP2 found an optimal cut for all 10 problems with Gaussian distributed couplings, and also for almost half (5 out of 11) of the problems with $\pm$ 1 couplings.

3. *Why?* Perhaps looking into the structure of the specific test problems will help us to understand this phenomenon.

n	Wt opt.	MCSDP (% rel. err)	MCPSDP2 (% rel. err)	Num. rank
5	4	4.5225 (13.06%)	4.2890 (7.22%)	2
7	56	56.4055 (0.72%)	56.0954 (0.17%)	3
8	30	30.2015 (0.67%)	30. (e-10%)	1
9	58	58.9361 (1.61%)	58.1182 (0.20%)	3
10	64	64.08 (0.1268%)	64 (e-08%)	3
12	88	90.3919 (2.72%)	89.5733 (1.79%)	4

The first line of results corresponds to solving both MC relaxations for a 5-cycle with unit edge-weights; the others come from randomly generated weighted graphs.

Graph	μ^*	SDP1 bound	$\mathrm{SDP2_P}$ bound	M_n bound	$\mathcal{E}_n \cap M_n$ bound	$\mathrm{SDP3_P}$ bound
C_5	4	4.5225 $\rho = 0.8845$ R.E.: 13.06%	4.2889 $\rho = 0.9326$ R.E.: 7.22%	4.0000 $\rho = 1.0000$ R.E.: 0%	4.0000 $\rho = 1.0000$ R.E.: 0%	4.0000 $\rho = 1.0000$ R.E.: 0%
$K_5 \backslash e$	6	6.2500 $\rho = 0.9600$ R.E.: 4.17%	6.1160 $\rho = 0.9810$ R.E.: 1.93%	6.0000 $\rho = 1.0000$ R.E.: 0%	6.0000 $\rho = 1.0000$ R.E.: 0%	6.0000 $\rho = 1.0000$ R.E.: 0%
K_5	6	6.2500 $\rho = 0.9600$ R.E.: 4.17%	6.2500 $\rho = 0.9600$ R.E.: 4.17%	6.6667 $\rho = 0.9000$ R.E.: 11.11%	6.2500 $\rho = 0.9600$ R.E.: 4.17%	6.2500 $\rho = 0.9600$ R.E.: 4.17%
Given by $A(G)$ $(n = 5)$	9.28	9.6040 $\rho = 0.9663$ R.E.: 3.49%	9.4056 $\rho = 0.9866$ R.E.: 1.35%	9.3867 $\rho = 0.9886$ R.E.: 1.15%	9.2961 $\rho = 0.9983$ R.E.: 0.17%	9.2800 $\rho = 1.0000$ R.E.: 0%
AW_9^2	12	13.5 $\rho = 0.8889$ R.E.: 12.50%	12.9827 $\rho = 0.9243$ R.E.: 8.19%	12.8571 $\rho = 0.9333$ R.E.: 7.14%	12.6114 $\rho = 0.9515$ R.E.: 5.10%	12.4967 $\rho = 0.9603$ R.E.: 4.14%
Pet. $(n = 10)$	12	12.5 $\rho = 0.9600$ R.E.: 4.17%	12.3781 $\rho = 0.9695$ R.E.: 3.15%	12.0000 $\rho = 1.0000$ R.E.: 0%	12.0000 $\rho = 1.0000$ R.E.: 0%	12.0000 $\rho = 1.0000$ R.E.: 0%
Given in [3] $(n = 12)$	88	90.3919 $\rho = 0.9735$ R.E.: 2.72%	89.5733 $\rho = 0.9824$ R.E.: 1.79%	89.3333 $\rho = 0.9851$ R.E.: 1.52%	88.0029 $\rho = 1.0000$ R.E.: $3.3E-5$	88.0000 $\rho = 1.0000$ R.E.: $9.9E-7$

Numerical comparison of all MC relaxations for selected test problems

Number of vertices	μ^*	SDP1 bound	SDP2_P bound	SDP3_P bound
10	648	666.428 $\rho = 0.9723$ R.E.: 2.84%	656.8020 $\rho = 0.9866$ R.E.: 1.36%	648.000 $\rho = 1.0000$ R.E.: 0%
11	1060	1084.345 $\rho = 0.9775$ R.E.: 2.30%	1072.352 $\rho = 0.9885$ R.E.: 1.17%	1060.000 $\rho = 1.0000$ R.E.: 0%
15	2290	2317.354 $\rho = 0.9882$ R.E.: 1.19%	2301.634 $\rho = 0.9949$ R.E.: 0.51%	2290.000 $\rho = 1.0000$ R.E.: 0%
16	2270	2318.867 $\rho = 0.9789$ R.E.: 2.15%	2300.354 $\rho = 0.9868$ R.E.: 1.34%	2270.000 $\rho = 1.0000$ R.E.: 0%
25	380	385.4737 $\rho = 0.9858$ R.E.: 1.44%	383.6503 $\rho = 0.9905$ R.E.: 0.96%	380.000 $\rho = 1.0000$ R.E.: 0%
30	1705.5	1751.600 $\rho = 0.9737$ R.E.: 2.70%	1743.205 $\rho = 0.9784$ R.E.: 2.21%	1705.578 $\rho = 1.0000$ R.E.: $4.6E-5$
33	1888.5	1932.968 $\rho = 0.9770$ R.E.: 2.35%	1926.119 $\rho = 0.9805$ R.E.: 1.99%	1888.564 $\rho = 1.0000$ R.E.: $3.4E-5$
36	27108.55	28305.28 $\rho = 0.9577$ R.E.: 4.41%	27944.30 $\rho = 0.9701$ R.E.: 3.08%	27108.81 $\rho = 1.0000$ R.E.: $9.6E-6$

Numerical comparison of SDP1, $SDP2_P$ and $SDP3_P$ on randomly generated graphs with non-negative edge weights

3.3 Quadratically Constrained Quadratic Programs and SDP Relaxations

Outline

- Theme: Lagrangian Relaxation is Best
- Examples of QQPs with Zero Duality Gaps
- Advantages of Redundant Constraints (more in Section 3.2 on Quadratic Boolean Programming)
- Applications to QAP, GP

3.3.1 Quadratically Constrained Programs with Zero Duality Gaps

$$\text{Let} \quad q_i(y) = \frac{1}{2} y^T Q_i y + y^T b_i + c_i, \; y \in \mathbb{R}^n$$

Consider the NP-hard

$$(\mathbf{QQP}) \qquad \begin{cases} q^* = & \min & q_0(y) \\ & \text{s.t.} & q_i(y) \le 0 \\ & & i = 1, \dots m \end{cases}$$

- Lagrangian is $L(y, \lambda) = q_0(y) + \sum_{i=1}^m \lambda_i q_i(y)$
- Primal-Dual pair: $q^* = \min_y \max_{\lambda \ge 0} L(y, \lambda) \ge d^* := \max_{\lambda \ge 0} \min_y L(y, \lambda)$

Using the hidden constraint (a quadratic is bounded below only if the Hessian is positive semidefinite), we get:
the Lagrangian dual is equivalent to an SDP - thus tractable!

When do duality gaps exist? Zero duality gaps exist?

> Theme: Lagrangian duality is *best* relaxation. (Tractable relaxations can be obtained using Lagrangian relaxation.)

Example: **Rayleigh Quotient,** $A = A^T$
(Finding the eigenvalues is a tractable problem - well known.)

smallest eigenvalue λ_1 of A

$$\lambda_1 = \min\{x^T Ax : x^T x = 1\}. \tag{3.2}$$

A is not necessarily positive semidefinite, so this is the minimization of a nonconvex function on a nonconvex set. There is no duality gap for this nonconvex problem, i.e.

$$\lambda_1 = \max_{\lambda} \min_{x} \; x^T Ax - \lambda(x^T x - 1) = \max_{A-\lambda I \succeq 0} \lambda. \tag{3.3}$$

use the hidden semidefinite constraint

$$A - \lambda I \succeq 0,$$

Example: **Trust Region Subproblem**

$$\begin{array}{rl} \mu^* := \min & q_0(x) \\ \text{s.t.} & x^T x - \delta^2 \leq 0 \ (\text{or } \ = 0). \end{array}$$

This is a tractable problem, e.g. [22, 52]

(allow more general nonconvex constraints $\alpha \leq q(x) \leq \beta$.) for "$\leq$," the Lagrangian dual is:

$$\text{DTRS} \qquad \nu^* := \max_{\lambda \geq 0} \ \min_x \ q_0(x) + \lambda(x^T x - \delta^2).$$

strong duality holds ([79]), a dual is the (concave) nonlinear semidefinite program

$$\begin{array}{lrl} \text{DTRS} & \nu^* := \max & g_0^T (Q + \lambda I)^\dagger g_0 - \lambda \delta^2 \\ & \text{s.t.} & Q + \lambda I \succeq 0 \\ & & \lambda \geq 0. \end{array}$$

short proof of strong duality (e.g. [43])
WLOG TRS is nonconvex; smallest eigenvalue of Q_0, denoted γ, is negative:

$$
\begin{array}{rll}
 & \mu^* = & \\
= & \min\limits_{x^Tx\le\delta^2} & x^T(Q_0-\gamma I)x - 2c_0^t x + \gamma x^T x \\
= & \min\limits_{x^Tx=\delta^2} & x^T(Q_0-\gamma I)x - 2c_0^t x + \gamma x^t x, \quad (Q_0 \text{ is indefinite}) \\
= & \min\limits_{x^Tx=\delta^2} & x^T(Q_0-\gamma I)x - 2c_0^t x + \gamma \delta^2 \\
= & \min\limits_{x^Tx\le\delta^2} & x^T(Q_0-\gamma I)x - 2c_0^t x + \gamma \delta^2, \ (Q_0-\gamma I \text{ singular}) \\
= & \max\limits_{\lambda\ge 0}\min\limits_{x} & x^T(Q_0-\gamma I)x - 2c_0^t x + \lambda(x^Tx-\delta^2) + \gamma\delta^2, \ (\text{convex case}) \\
= & \max\limits_{\lambda\ge 0}\min\limits_{x} & x^TQ_0x - 2c_0^t x + (\lambda-\gamma)(x^Tx-\delta^2) \\
\le & \max\limits_{\lambda\ge \gamma}\min\limits_{x} & x^TQ_0x - 2c_0^t x + (\lambda-\gamma)(x^Tx-\delta^2) \quad (\gamma<0) \\
= & \nu^* \le \mu^*. &
\end{array}
$$

Two Trust Region Subproblem TTRS consists in minimizing a (possibly nonconvex) quadratic function subject to a norm and a least squares constraint.

ref in SQP methods by Celis-Dennis-Tapia.

TTRS can have a nonzero duality gap, ref Peng-Yuan.

if objective not convex, then the primal may not be attained, ref Luo-Zhang.

TRS can have at most one local and nonglobal optimum, ref Martinez.

Still an open problem whether TTRS is an NP-hard or a polynomial time problem.

Note: a quadratic constraint can be written as

$$y^t P y \leq \delta$$

after the homogenization. This is lifted to

$$\operatorname{trace} PY \leq \delta.$$

Second lifting:

$$YPY \preceq \delta Y.$$

This strictly strengthens the SDP relaxation.

Orthogonal Contraints

permutation matrices are a subset of orthogonal matrices

$$\Pi \subset \mathcal{O} := \{X : XX^T = I\}$$

(Stiefel manifold ref e.g. Edelman,Arias,Smith)

A and B $n \times n$ symmetric matrices

$$\begin{array}{llll} \mathrm{QQP_O} & \mu^* := & \min & \operatorname{trace} AXBX^T \\ & & \text{s.t.} & XX^T = I. \end{array}$$

(*looks* like TRS)

QQP_O is tractable; solved using e.g. classical Hoffman-Wielandt inequality; provides bounds for e.g. QAP.

Proof. Apply Lagrange multipliers, $S = S^T$. Lagrangian is:

$$\begin{array}{rcl} L(X,S) & = & \operatorname{trace}(AXBX^T) + \langle S, (XX^T - I)\rangle \\ & = & \operatorname{trace}(AXBX^T) + \operatorname{trace} S(XX^T - I) \end{array}$$

$$0 = \langle \nabla L(X,S), h\rangle = 2\operatorname{trace} AXBh^T + SXIh^T \quad \forall h.$$

Therefore $AXBX^T = -S = -S^T$, i.e. A and XBX^T are mutually diagonalizable. The optimal value is then

$$\mu^* = \sum_i \lambda_i(A)\lambda_{n-i+1}(B) \qquad (3.4)$$

(minimal scalar product of eigenvalues) ■

But, Lagrangian dual can have a duality gap, [89]

Lagrangian of (P)

$$L(X,S) = \operatorname{trace} AXBX^T + \operatorname{trace} SXX^T - \operatorname{trace} S.$$

Lagrangian dual is

$$\nu^* = \max_S \min_X L(X,S).$$

The hidden constraint (Hessian is psd) yields the dual

$$(D) \quad \begin{array}{rcc} \mu^D = & \max\limits_{\hat{S}=\hat{S}^T} & -\operatorname{trace} \hat{S} \\ & \text{subject to} & (B \otimes A + I \otimes \hat{S}) \succeq 0. \end{array}$$

Example 3.3.1.1

$$A := \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \qquad B := \begin{pmatrix} 3 & 0 \\ 0 & 4 \end{pmatrix}.$$

$\mu^* = 10$

$$B \otimes A = \begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 6 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 8 \end{pmatrix}.$$

But $s_{11} \geq -3$ *and* $s_{22} \geq -6$*; to maximize* (D)*, equality must hold, and therefore* $-\mathrm{trace}\, \hat{S} = 9$ *in the optimum.*

However, add redundant constraint $X^T X = I$*. Get* $T \otimes I$ *in Hessian and* $-\mathrm{trace}\, T$ *in objective function.* $s_{11} \geq -3 - t_{11}$ $s_{11} \geq -4 - t_{22}$ $s_{22} \geq -6 - t_{11}$ $s_{22} \geq -8 - t_{22}$ *So* $t_{22} = -1$*. Duality gap is closed.*

Summary: constraints $XX^T = I$ and $X^TX = I$ are equivalent.
Add redundant constraints.

$$\text{QQP}_{\text{OO}} \qquad \mu^O := \begin{array}{ll} \min & \operatorname{trace} AXBX^T \\ \text{s.t.} & XX^T = I, \; X^TX = I. \end{array}$$

dual problem is

$$\text{DQQP}_{\text{OO}} \qquad \mu^O \geq \mu^D := \begin{array}{ll} \max & \operatorname{trace} S + \operatorname{trace} T \\ \text{s.t.} & (I \otimes S) + (T \otimes I) \preceq (B \otimes A) \\ & S = S^T, \; T = T^T. \end{array}$$

Theorem 3.3.1.1 *([6]) Strong duality holds for* QQP_{OO} *and* DQQP_{OO}*, i.e.* $\mu^D = \mu^O$ *and both primal and dual are attained.*

Theme again holds.

Proof. (of Theorem 3.3.1.1) Let $A = V\Sigma V^T$, $B = U\Lambda U^T$, where V and U are orthonormal matrices whose columns are the eigenvectors of A and B, respectively, σ and λ are the corresponding vectors of eigenvalues, and $\Sigma = \mathrm{Diag}\,(\sigma)$, $\Lambda = \mathrm{Diag}\,(\lambda)$. Then for any S and T,

$$\begin{aligned}(B \otimes A) - (I \otimes S) - (T \otimes I) = \\ (U \otimes V)\left[(\Lambda \otimes \Sigma) - (I \otimes \bar{S}) - (\bar{T} \otimes I)\right](U^T \otimes V^T),\end{aligned}$$

where $\bar{S} = V^T S V$, $\bar{T} = U^T T U$. Since $U \otimes V$ is nonsingular, $\mathrm{trace}\, S = \mathrm{trace}\, \bar{S}$ and $\mathrm{trace}\, T = \mathrm{trace}\, \bar{T}$, the dual problem $\mathrm{DQQP}_{\mathrm{OO}}$ is equivalent to

$$\begin{array}{rll}\mu^D = & \max & \mathrm{trace}\, S + \mathrm{trace}\, T \\ & \text{s.t.} & (\Lambda \otimes \Sigma) - (I \otimes S) - (T \otimes I) \succeq 0 \\ & & S = S^T,\ T = T^T.\end{array}$$

However, since Λ and Σ are diagonal matrices, (233) is equivalent to the ordinary linear program:

$$
\begin{array}{lll}
\text{LD} & \max & e^T s + e^T t \\
& \text{s.t.} & \lambda_i \sigma_j - s_j - t_i \geq 0, \quad i,j = 1,\ldots,n.
\end{array}
$$

But LD is the dual of the linear assignment problem:

$$
\begin{array}{lll}
\text{LP} & \min & \sum_{i,j} \lambda_i \sigma_j x_{ij} \\
& \text{s.t.} & \sum_{j=1}^{n} x_{ij} = 1, \quad i = 1,\ldots,n \\
& & \sum_{i=1}^{n} x_{ij} = 1, \quad j = 1,\ldots,n \\
& & x_{ij} \geq 0, \quad i,j = 1,\ldots,n.
\end{array}
$$

Assume without loss of generality that $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$, and $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n$. Then LP can be interpreted as the problem of finding a permutation $\pi(\cdot)$ of $\{1, \ldots, n\}$ so that $\sum_{i=1}^n \lambda_i \sigma_{\pi(i)}$ is minimized. But the minimizing permutation is then $\pi(i) = i$, $i = 1, \ldots, n$, and from the minimal inner-product expression (3.4), the solution value μ^D is exactly μ^O. $\blacksquare$

Other applications:

Weighted Sums of Eigenvalues;

Graph Partitioning Problem;

TRS like constraints $\{X : XX^T \preceq I\}$ ([5]) (extension of Hoffman-Wielandt inequality)

3.3.2 Quadratic Assignment Problem, QAP

$$(QAP) \quad \mu^* := \min_{X \in \Pi} \operatorname{trace} AXBX^T - 2CX^T,$$

A, B real symmetric $n \times n$ matrices, C real $n \times n$ matrix, Π set of permutation matrices.

model for e.g. allocating set of n facilities to set of n locations while minimizing quadratic objective (distance-flow)

QAP is NP-hard and, in practice, problems of moderate sizes, such as $n = 16$, are still considered very hard. (Recent advances to $n = 30$, [4].)

Questions for SDP type bounds:

1. interesting numerical and theoretical difficulties, e.g. loss of constraint qualification and loss of sparsity in the optimality conditions.
2. Can the new bound compete with other bounding techniques in speed and quality?
3. comparison on the Nugent test problems?
4. add new facet inequalities?

Notation:

$\mathcal{E} := \{X : Xe = X^T e = e\}$
set of matrices satisfying *assignment constraints*

$\mathcal{Z} := \{X : X_{ij} \in \{0,1\}\}$
set of (0,1)-matrices

$\mathcal{N} := \{X : X_{ij} \geq 0\}$
set of *nonnegative matrices*

$\mathcal{O} := \{X : XX^T = X^T X = I\}$
set of orthogonal matrices

$\Pi = \mathcal{E} \cap \mathcal{Z} = \mathcal{O} \cap \mathcal{Z}$, Diag $\leftrightarrow$ diag is vector $\leftrightarrow$ matrix

rewrite QAP as (add redundant constraints)

$$
\begin{array}{ll}
\min & \operatorname{trace} AXBX^T - 2CX^T \\
\text{s.t.} & XX^T = X^TX = I \\
& \|Xe - e\|^2 = 0 \\
& \|X^Te - e\|^2 = 0 \\
& X_{ij}^2 - X_{ij} = 0, \quad \forall i, j. \\
& \operatorname{diag}(X_{:i}X_{:j}^T) = 0, \quad \text{if } i \neq j \\
& X_{:i}X_{:j}^T - \operatorname{Diag}\left(\operatorname{diag}(X_{:i}X_{:j}^T)\right) = 0, \quad \text{if } i = j
\end{array}
$$

Lagrangian Relaxation

$$
\begin{aligned}
\mu_{\mathcal{O}} \geq \mu_{\mathcal{L}} & \\
:= \max_{W,u_0,v_0\ldots} \min_{XX^T=X^TX=I} & \{\operatorname{trace} AXBX^T - 2CX^T \\
& + \textstyle\sum_{ij} W_{ij}(X_{ij}^2 - X_{ij}) \\
& + u_0\|Xe - e\|^2 \\
& + v_0\|X^Te - e\|^2 \\
& + \ldots\}.
\end{aligned}
$$

homogenize the Lagrangian using scalar x_0 and constraint $x_0^2 = 1$.

We get the lower bound
(quadratic, linear constant in X)

$$\begin{aligned}
\max_{W,S_b,S_o,u_0 v_0,,w_0} \ \min_{X,\ x_0} \quad & \{\mathrm{trace}\,[AXBX^T \\
& + u_0\|Xe\|^2 + v_0\|X^T e\|^2 \\
& + W(X \circ X)^T + w_0 x_0^2 \\
& + S_b X X^T + S_o X^T X] \\
& - \mathrm{trace}\, x_0(2C + W)X^T \\
& - 2x_0 u_0 e^T (X + X^T)e \\
& + \ldots \\
& - w_0 - \mathrm{trace}\, S_b - \mathrm{trace}\, S_o + 2n u_0 x_0^2\}.
\end{aligned}$$

The hidden semidefinite constraint yields the equivalent SDP:

$$
(D_{\mathcal{O}}) \quad \begin{array}{ll} \max & -w_0 - \operatorname{trace} S_b - \operatorname{trace} S_o \\ \text{s.t.} & L_Q + \operatorname{Arrow}(w) + \\ & \mathrm{B}^0\operatorname{Diag}(S_b) + \mathrm{O}^0\operatorname{Diag}(S_o) + w_0 D \succeq 0, \end{array}
$$

The dual of this dual yields the semidefinite relaxation.

$Y \succeq 0$ is $(n^2+1) \times (n^2+1)$
the dual matrix variable

$$
(SDP_{\mathcal{O}}) \quad \begin{array}{llll} \min & \operatorname{trace} L_Q Y & & \\ \text{s.t.} & \mathrm{b}^0\operatorname{diag}(Y) = I & & \mathrm{o}^0\operatorname{diag}(Y) = I \\ & \operatorname{arrow}(Y) = e_0 & & \operatorname{trace} DY = 0 \\ & Y \succeq 0 & & \end{array}
$$

adjoint operators are:

$$\operatorname{arrow}(Y) := \operatorname{diag}(Y) - (0, (Y_{0,1:n^2})^t.$$

$$\mathrm{b}^0\operatorname{diag}(Y) := \sum_{k=1}^{n} Y_{(k-1)n+1:kn,(k-1)n+1:kn}$$

$$[\mathrm{o}^0\operatorname{diag}(Y)]_{ij} := \operatorname{trace} Y_{(i-1)n+1:in,(j-1)n+1:jn}$$

check Slater's condition in SDP relaxation

$$
\begin{array}{ll}
\min & \operatorname{trace} L_Q Y \\
\text{s.t.} & \mathrm{b}^0\mathrm{diag}\,(Y) = I, \quad \mathrm{o}^0\mathrm{diag}\,(Y) = I \\
& \operatorname{arrow}(Y) = e_0, \quad \operatorname{trace} DY = 0 \\
& \ldots \\
& Y \succeq 0,
\end{array}
$$

$D \succeq 0$ so it fails!!
But we can project onto the minimal face!!

Now remove redundant constraints to get simplified SDP relaxation simple projected relaxation with $n^3 - 2n^2 + 1$ constraints.

$$\begin{array}{rll} \mu_{R2} := & \min & \operatorname{trace}(\hat{V}^T L_Q \hat{V})R \\ & \text{s.t.} & \mathcal{G}_{\bar{J}}(\hat{V} R \hat{V}^T) = E_{00} \\ & & R \succeq 0. \end{array}$$

The dual problem is

$$\begin{array}{rll} \mu_{R2} = & \max & -Y_{00} \\ & \text{s.t.} & \hat{V}^T (L_Q + \mathcal{G}_{\bar{J}}^*(Y))\hat{V} \succeq 0. \end{array}$$

Direct Approach to SDP Relaxation

Let

$X \in \Pi_n$ be a permutation matrix

$x = \mathrm{vec}\,(X),\ c = \mathrm{vec}\,(C)$.

$$\begin{array}{rcl} q(X) & = & \mathrm{trace}\, AXBX^t - 2CX^t \\ & = & x^t(B \otimes A)x - 2c^t x \\ & = & \mathrm{trace}\, xx^t(B \otimes A) - 2c^t x \\ & = & \mathrm{trace}\, L_Q Y_X, \end{array}$$

where L_Q is as above and

$$Y_X := \begin{bmatrix} 1 & x^t \\ x & xx^t \end{bmatrix}.$$

Adding Generic Inequality Constraints

From the relaxation for the (0,1)-constraints of the original problem:

$$y_{ij} \geq 0 \quad \text{since} \quad x_i x_j \geq 0.$$

We also get so called triangle inequalities

$$y_{ij} + y_{ik} + y_{jk} - (y_{ii} + y_{jj} + y_{kk}) + 1 \geq 0$$

and

$$y_{ij} - y_{ik} - y_{jk} + y_{kk} \geq 0$$

THEOREM Let $x = \operatorname{vec}(X)$ and F_S be the feasible set of (SDP). Define the centroid point

$$\hat{Y} := \frac{1}{n!} \sum_{X \in \Pi_n} \begin{bmatrix} 1 & x^t \\ x & xx^t \end{bmatrix}.$$

Then:

1. $\hat{Y}$ has a 1 in the (1,1) position and n diagonal $n \times n$ blocks with diagonal elements $1/n$. The first row and column equal the diagonal. The rest of the matrix is made up of $n \times n$ blocks with all elements equal to $1/(n(n-1))$ except for the diagonal elements which are 0:

$\hat{Y} =$

$$= \left[\begin{array}{c|cccc} 1 & \multicolumn{4}{c}{\frac{1}{n}e^t} \\ \hline & \mathrm{diag}\,(\frac{1}{n}e) & (\frac{1}{n(n-1)})(E-I) & \cdots & (\frac{1}{n(n-1)})(E-I) \\ & \cdots & \cdots & \cdots & \cdots \\ \frac{1}{n}e & \cdots & \cdots & \cdots & \cdots \\ & \cdots & \cdots & \cdots & \cdots \\ & (\frac{1}{n(n-1)})(E-I) & \cdots & (\frac{1}{n(n-1)})(E-I) & \mathrm{diag}\,(\frac{1}{n}e) \end{array}\right]$$

$$= \left[\begin{array}{c|c} 1 & \frac{1}{n}e^t \\ \hline \frac{1}{n}e & E \otimes \left(\frac{1}{n(n-1)}(E-I)\right) - I \otimes \left(\frac{1}{n(n-1)}E - \frac{1}{n-1}I\right) \end{array}\right]$$

2.

$$\mathrm{rank}\,(\hat{Y}) = (n-1)^2 + 1$$

3.

$$\operatorname{diag} \hat{Y} = \begin{pmatrix} 1 \\ \frac{1}{n} e \end{pmatrix}$$

4.

$$\operatorname{trace}(\hat{Y}) = 1 + n$$

5.

The $n^2 + 1$ eigenvalues of $\hat{Y}$ are given in the vector

$$(2, \frac{1}{n-1} e_{(n-1)^2}, 0 e_{2n-1})$$

6.
$$\mathcal{N}(\hat{Y}) = \left\{ \begin{pmatrix} -\frac{1}{n}e^t u \\ u \end{pmatrix} : u \in \mathcal{R}(T^t) \right\},$$
where T is the assignment constraint matrix.

——————————————-

We can use the matrix T to project onto the minimal face.

We get a simplified SDP with a positive definite feasible point.

Recent results of Anstreicher uses a weakened form of the SDP relaxation for QAP in combination with the strong duality result for the eigenvalue bound.

3.3.3 Graph Partitioning, GP

(special case of QAP)

k and m given integers
G an edgeweighted undirected graph on $n := km$ nodes, given by its adjacency matrix A.
(a_{ij} weight of edge $i \leftrightarrow j$)

a k-partition of $V(G)$ is a partitioning of $V(G)$ into k subsets $(S_1, \ldots, S_k)$ of equal cardinality.

denoted (k-GP)

the columns of $Y \in \Re^{n \times k}$ are the characteristic vectors for the sets S_k

$$\mathcal{F}_k := \{Y : Yu_k = u_n,\ Y^t u_n = mu_k, y_{ij} \in \{0,1\}\}$$

$L := \operatorname{diag}(Au_n) - A$ *Laplacian matrix associated to G.*

weight of the edges of G, cut by some k-partition $Y \in \mathcal{F}_k$,

$$\frac{1}{2} \operatorname{trace} Y^t L Y$$

THE PROBLEM:

$$(k-GP) \quad \begin{array}{cc} \min & \frac{1}{2} \operatorname{trace} Y^t L Y \quad (= \operatorname{trace} L Y Y^t) \\ \text{subject to} & Y \in \mathcal{F}_k \end{array}$$

$$\mathcal{T}_k := \{X : X = YY^t \text{ for some } Y \in \mathcal{F}_k\}$$

$$\mathcal{E}_m := \{X : X = X^t, \operatorname{diag}(X) = u_n, Xu_n = mu_n\}$$

note $X \in \mathcal{T}_k$ implies that the only eigenvalues of X are 0 and m
Therefore $mI - X \succeq 0$

THE SDP RELAXATION:

$$\min\{\frac{1}{2}\operatorname{trace} LX : X \in \mathcal{E}_m,\ X \succeq 0,\ mI - X \succeq 0\}.$$

We can add further constraints, e.g. $X \geq 0$ and other 'polyhedral constraints'

Numerical tests by:
Stefan Karisch and Franz Rendl

n	$\lvert E\rvert$	$\lvert E_{cut}\rvert$	$(k-GP_{R1})$	$(k-GP_{R4})$
36a	297	117	111.76	116.22
36b	303	118	111.63	117.18
36c	316	124	119.48	123.34
60a	885	367	354.46	364.84
60b	863	357	343.24	354.57
60c	841	343	329.43	341.61
84a	1742	742	716.38	734.70
84b	1793	772	744.67	762.38
84c	1753	753	728.01	744.25
108a	2873	1247	1205.80	1231.12
108b	2933	1283	1243.12	1264.06
108c	2871	1240	1201.78	1227.50
132a	4294	1885	1832.76	1861.55
132b	4301	1883	1842.16	1870.04
132c	4257	1854	1808.86	1837.78

Partitioning unweighted random graphs into $k = 2$ components.

n	$\vert E_{cut}\vert$	$(k-GP_{R1})$	$(k-GP_{R2})$		$(k-GP_{R3})$
36a	160	149.02	154.32	(102)	157.36
36b	163	148.84	155.84	(98)	159.11
36c	173	159.31	166.22	(103)	169.40
60a	502	472.61	484.14	(174)	489.88
60b	485	457.65	468.21	(185)	474.03
60c	470	439.24	451.49	(200)	457.79
84a	1011	955.18	972.89	(300)	981.17
84b	1044	992.89	1007.73	(290)	1016.83
84c	1024	970.68	986.38	(276)	993.57
108a	1686	1607.74	1630.80	(399)	1642.86
108b	1738	1657.50	1675.22	(376)	1685.67
108c	1688	1602.38	1628.73	(404)	1639.40
132a	2556	2443.69	2469.33	(498)	2483.07
132b	2569	2456.21	2486.40	(510)	2498.01
132c	2526	2411.82	2442.92	(527)	2456.18

Partitioning unweighted random graphs into $k = 3$ components.

n	$\lvert E_{cut} \rvert$	$(k - GP_{R1})$	$(k - GP_{R2})$	$(k - GP_{R3})$
36a	186	167.65	179.73	180.74
36b	189	167.45	182.63	183.65
36c	201	179.22	193.99	195.27
60a	577	531.69	556.13	557.92
60b	558	514.86	538.32	540.09
60c	540	494.14	521.45	523.07
84a	1155	1074.57	1112.74	1114.69
84b	1195	1117.00	1152.30	1154.58
84c	1164	1092.01	1126.55	1128.25
108a	1918	1808.71	1860.94	1863.37
108b	1981	1864.68	1907.70	1909.83
108c	1923	1802.68	1858.23	1860.17
132a	2910	2749.15	2810.40	2812.43
132b	2921	2763.24	2827.65	2829.61
132c	2874	2713.30	2781.83	2784.19

Partitioning unweighted random graphs into $k = 4$ components.

n	$\|E_{cut}\|$	$(k - GP_{R1})$	$(k - GP_{R2})$
36a	216	186.27	210.22
36b	221	186.06	215.53
36c	234	199.13	227.62
60a	660	590.77	638.21
60b	640	572.06	618.24
60c	622	549.05	600.66
84a	1313	1193.97	1265.56
84b	1355	1241.11	1310.61
84c	1325	1213.35	1279.39
108a	2184	2009.68	2106.69
108b	2237	2071.87	2158.07
108c	2185	2002.97	2104.26
132a	3290	3054.61	3171.61
132b	3292	3070.26	3187.10
132c	3251	3014.78	3141.52

Partitioning unweighted random graphs into $k = 6$ components.

n	$w(E)$	$w(E_{cut})$	$(k-GP_{R1})$	$(k-GP_{R4})$
36d	3055	1426	1402.76	1425.27
36e	3168	1482	1462.12	1481.12
36f	3134	1454	1435.05	1453.24
60d	8831	4151	4118.59	4150.11
60e	8833	4154	4102.88	4149.89
60f	8777	4132	4078.85	4125.43
84d	17241	8152	8057.77	8132.57
84e	17547	8327	8233.21	8296.74
84f	17473	8264	8166.24	8237.93
108d	29153	13895	13728.41	13825.87
108e	28877	13699	13550.40	13649.46
108f	28892	13709	13588.80	13678.63
132d	43208	20581	20380.59	20514.25
132e	43220	20618	20415.68	20536.67
132f	43391	20707	20488.58	20617.09

Partitioning weighted random graphs into $k = 2$ components.

n	$w(E_{cut})$	$(k-GP_{R1})$	$(k-GP_{R2})$		$(k-GP_{R3})$
36d	1924	1870.35	1891.39	(96)	1904.92
36e	2006	1949.49	1973.55	(102)	1987.21
36f	1974	1913.40	1944.87	(111)	1956.33
60d	5609	5491.45	5533.67	(182)	5558.27
60e	5592	5470.51	5520.86	(195)	5543.31
60f	5555	5438.46	5481.36	(177)	5507.30
84d	10960	10743.70	10828.31	(286)	10860.45
84e	11181	10977.62	11034.14	(322)	11068.14
84f	11105	10888.33	10966.49	(315)	11001.26
108d	18624	18304.55	18398.93	(433)	18438.50
108e	18420	18067.20	18180.35	(429)	18223.56
108f	18442	18118.40	18219.78	(413)	18258.13
132d	27659	27174.12	27325.69	(564)	27374.23
132e	27657	27220.90	27332.27	(506)	27392.74
132f	27768	27318.11	27444.35	(543)	27500.82

Partitioning weighted random graphs into $k = 3$ components.

n	$w(E_{cut})$	$(k - GP_{R1})$	$(k - GP_{R2})$	$(k - GP_{R3})$
36d	2182	2104.14	2152.81	2159.53
36e	2268	2193.18	2243.20	2250.25
36f	2241	2152.57	2212.78	2218.76
60d	6341	6177.88	6288.66	6278.02
60e	6344	6154.33	6259.22	6266.31
60f	6282	6118.27	6214.29	6222.40
84d	12421	12086.66	12260.11	12268.31
84e	12645	12349.82	12482.87	12492.12
84f	12553	12249.37	12411.77	12421.67
108d	21069	20592.62	20796.31	20804.05
108e	20836	20325.60	20562.29	20571.90
108f	20859	20383.20	20599.70	20606.26
132d	31277	30570.89	30888.41	30894.53
132e	31273	30623.52	30885.18	30894.99
132f	31392	30732.87	31016.60	31025.67

Partitioning weighted random graphs into $k = 4$ components.

n	$w(E_{cut})$	$(k - GP_{R1})$	$(k - GP_{R2})$
36d	2454	2337.93	2434.02
36e	2552	2436.87	2533.87
36f	2525	2391.75	2502.79
60d	7116	6864.32	7039.53
60e	7115	6838.14	7033.36
60f	7048	6798.08	6985.54
84d	13923	13429.63	13740.85
84e	14171	13722.03	13983.97
84f	14093	13610.41	13909.54
108d	23575	22880.69	23258.89
108e	23303	22584.00	23007.83
108f	23345	22648.00	23043.71
132d	34985	33967.66	34518.13
132e	34965	34026.13	34522.69
132f	35109	34147.64	34663.42

Partitioning weighted random graphs into $k = 6$ components.

k	gap_{R1}	gap_{R2}	gap_{R3}	gap_{R4}
2	3.34	–	–	0.73
3	5.54	3.50	2.46	–
4	7.34	3.41	3.17	–
6	9.88	3.29	–	–

Average gaps in percent for unweighted graphs.

k	gap_{R1}	gap_{R2}	gap_{R3}	gap_{R4}
2	1.14	–	–	0.23
3	2.07	1.31	0.95	–
4	2.66	1.22	1.12	–
6	3.54	1.15	–	–

Average gaps in percent for weighted graphs.

n	$\|E\|$	$\|E_{cut}\|$		$(k-GP_{R1})$	$(k-GP_{R4})$
124a	149	13	(13)	7.34	12.00
124b	318	63	(63)	46.86	61.01
124c	620	178	(179)	153.01	170.91
124d	1271	449	(449)	418.98	440.06
250a	331	29	(29)	15.44	24.87
250b	612	114	(116)	81.86	100.45
250c	1283	357	(361)	303.53	327.88
250d	2421	828	(833)	747.32	779.55

Partitioning graphs of Johnson et al. into $k = 2$ components.

n	k	$\|E_{cut}\|$	$(k-GP_{R1})$	$(k-GP_{R2})$		$(k-GP_{R3})$
124a	4	23	11.01	13.65	(638)	17.16
124b	4	107	70.29	81.36	(604)	92.06
124c	4	294	229.52	251.51	(704)	260.65
124d	4	726	628.48	662.64	(749)	669.14
250a	5	63	24.71	31.08		31.08
250b	5	218	130.99	150.32		152.82
250c	5	648	485.66	527.12		530.82
250d	5	1447	1195.72	1268.17		1269.03

Partitioning graphs of Johnson et al. into $k = 4$ or $k = 5$ components.

n	k	$\lvert E_{cut} \rvert$	$(k-GP_{R1})$	$(k-GP_{R2})$	$(k-GP_{R3})$	$(k-GP_{R4})$
120	2	8	1.67	–	–	4.83
120	3	10	2.22	4.64	7.67	–
120	4	22	2.50	9.07	12.07	–
120	5	20	2.67	14.26	16.50	–
120	6	36	2.78	22.38	25.90	–

Partitioning a geometric graph of size $n = 120$ with $|E| = 413$.

n	$\lvert E\rvert$	cut	$\mathcal{E}_m \cap \mathcal{P}$
36	305	119	112
60	903	367	355
84	1762	747	717
108	2897	1252	1206
132	4325	1901	1833

Partitioning random graphs into 2 components of equal size. The first two columns describe the size of the graphs, column 3 contains the best equipartition found, column 4 contains the lower bound.

n	cut	$\mathcal{E}_m \cap \mathcal{P}$	$\mathcal{E}_m \cap \mathcal{P} \cap \mathcal{N}$	sign constr.
36	160	149.1	154.3	120
60	506	472.6	484.1	223
84	1014	955.1	972.8	349
108	1693	1607.7	1630.8	469
132	2563	2443.7	2469.3	554

Partitioning random graphs into $k = 3$ components of equal size. The first column identifies the graph, column 2 contains the best 3-partition found, columns 3 and 4 contain lower bounds from semidefinite relaxations, the last column indicates the number of nonnegativity constraints $x_{ij} \geq 0$ used, to insure $X \geq 0$.

n	cut	$\mathcal{E}_m \cap \mathcal{P}$	$\mathcal{E}_m \cap \mathcal{P} \cap \mathcal{N}$	sign constr.
36	186	167.6	179.7	211
60	585	531.7	556.1	405
84	1165	1074.6	1112.7	596
108	1935	1808.7	1860.9	859
132	2915	2749.2	2810.4	993

Partitioning random graphs into $k = 4$ components of equal size. The first column identifies the graph, column 2 contains the best 4-partition found, columns 3 and 4 contain lower bounds from semidefinite relaxations, the last column indicates the number of nonnegativity constraints $x_{ij} \geq 0$ used, to insure $X \geq 0$.

3.4 Trust Region Subproblems for NLP

3.4.1 Background

Let

$$q(x) := x^T A x - 2a^T x,$$

$$(TRS) \qquad \mu^* := \begin{array}{ll} \min & q(x) \\ \text{s.t.} & x^T x = s^2 \ \ (\leq s^2). \end{array}$$

where

$A = A^T$, not necessarily semidefinite
$a \in \mathbb{R}^n$, $s > 0$.

Motivation

Newton's method for minimization - global convergence difficulties

get around these difficulties by minimizing at each iteration the same quadratic model as in Newton's method, but restricted to a ball, the *trust region.*

Main Result: an algorithm that solves TRS using only matrix-vector multiplication; therefore exploits sparsity; SDP framework

Semidefinite Programming Framework

$$
(SDP) \quad \begin{array}{ll} \max & \operatorname{trace} CZ \\ \text{s.t.} & L(Z) = B \\ & Z \succeq 0 \end{array}
$$

$Z \succeq 0$ denotes positive semidefinite
L is a linear operator

Back to TRS:

$$
\begin{array}{llll}
(TRS) & \mu^* := & \min & q(x) \\
 & & \text{s.t.} & x^T x = s^2 \ \ (\leq s^2).
\end{array}
$$

nonlinear least squares, Levenberg 1944, Marquadt 1963

Applications to general minimization, Goldfeld, Quandt and Trotter 1966

Early theoretical results, Forsythe and Golub 1965

Hebden 1973, Reinsch 1967

efficient numerics, Gay 1981

state of the art algorithm, More-Sorenson 1983

Currently:

strong Lagrangian Duality, Stern and Wolkowicz 1993

polynomial time, Ye 1992

at most one local-nonglobal, Martinez 1993

two trust regions, CDT 1984

multiple trust regions: Shor, Ramana, Wolkowicz

Lanczos-eigenvalue approach:
Sorensen 1994,
(with SDP) Rendl and Wolkowicz 1994,
Santos and Sorensen 1995
Charles Fortin, (thesis) 2000

$$L(x,\lambda) = q(x) - \lambda(x^T x - s^2)$$

Theorem(Stern and Wolkowicz, 1993)
(i) Strong duality holds for TRS, i.e.

$$\mu^* = \min_x \max_\lambda L(x,\lambda) = \max_\lambda \min_x L(x,\lambda).$$

Moreover, attainment holds for x and uniquely for λ.
(ii) A dual problem for TRS, without a duality gap, is

$$(D) \quad \max_{A-\lambda I \succeq 0} h(\lambda),$$

where

$$h(\lambda) = \lambda s^2 - a^T (A - \lambda I)^\dagger a, \quad \text{(concave)}$$

and $\cdot^\dagger$ denotes the Moore-Penrose generalized inverse.

Proof

$$\begin{aligned}\mu^* &= \min_{||x||=s} q(x) \\ &\geq \max_{\lambda} \min_{x} L(x,\lambda) \\ &= \max_{A-\lambda I \succeq 0} \min_{x} L(x,\lambda),\end{aligned}$$

hidden constraints - Hessian $\succeq 0$ and stationarity
global minimum of the Lagrangian is

$$x_\lambda := (A - \lambda I)^\dagger a,$$

Therefore, the right hand side of the above is equal to

$$\begin{aligned}&= \max_{A-\lambda I \succeq 0} \lambda s^2 - a^T (A - \lambda I)^\dagger a \\ &= \max_{A-\lambda I \succeq 0} h(\lambda),\end{aligned}$$

This functional is strictly concave and coercive.
$h' = 0$ feasibility

Example (hard case) Let

$$A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \; a = (1\ 0)^T, \; s = 1.$$

$A - \lambda I \succeq 0$, implies $\lambda \leq -1$

$\lambda^* = -1$, $\mu^* = -1.5 = h(\lambda^*)$

$x_\lambda = (A - \lambda I)^\dagger a = (.5\ 0)^T$

$q(x_\lambda) = -.75 > \mu^*$

$||x_\lambda|| < 1$

complementary slackness fails

zero duality gap but no primal attainment

Nonlinear Primal-Dual Pair

The dual problem is

$$\max_{A-\lambda I \succeq 0} h(\lambda) \qquad (D)$$

where

$$h(\lambda) = \lambda s^2 - a^T (A - \lambda I)^\dagger a$$

dual of the dual:

$$\begin{array}{rc} \min & h(\lambda) + \operatorname{trace} Z(A - \lambda I) \\ \text{s.t.} & s^2 - ||(A - \lambda I)^\dagger a||^2 - \operatorname{trace} Z = 0 \qquad (DD) \\ & Z \succeq 0. \end{array}$$

For feasible pair Z, λ, the duality gap is:

$$\operatorname{trace} Z(A - \lambda I)$$

complementary slackness

Gale and More-Sorenson Algorithm explained by SDP Framework

maintain 2nd order cond. $(A - \lambda I) \succ 0$
maximize the dual function $h(\lambda)$
(Newton's method)

solve $0 = h'(\lambda) = s^2 - ||x_\lambda||^2$
better solve $0 = \frac{1}{s} - \frac{1}{||x_\lambda||}$
(almost linear, convex)

Cholesky factoriz: $R^T R = A - \lambda I$
(safeguard positive def.)

$$\begin{array}{ll} \text{Solve} & : \; R^T R p = a \\ \text{Solve} & : \; R^T q = p \\ \text{iterate} & : \; \lambda \leftarrow \lambda + \left(\frac{||p||}{||q||}\right)^2 \left(\frac{||p|| - s}{s}\right) \end{array}$$

Hard Case

indicated by $0 < s^2 - ||x_\lambda||^2$ inside TR

DD feasible

$$s^2 - ||(A - \lambda I)^\dagger a||^2 - \operatorname{trace} Z = 0$$

so take a primal step to boundary and try to minimize the duality gap with

$$Z = zz^T, \ \operatorname{trace} Z = ||z||^2$$

$$\begin{aligned} q(x+z) &= (\lambda s^2 - x_\lambda^T (A - \lambda I) x_\lambda) + z^T (A - \lambda I) z \\ &= h(\lambda) + z^T (A - \lambda I) z. \end{aligned}$$

$$\min \operatorname{trace} Z(A - \lambda I) = z^T (A - \lambda I) z$$

$$A - \lambda I = RRc^T \text{ Cholesky decomp}$$

(singular value/vector of R)

Another approach:

Homogenization of TRS

$$\begin{aligned}
\mu^* &= \min_{||x||=s,\ y_0^2=1} x^T A x - 2y_0 a^T x \\
&= \max_t \min_{||x||=s} x^T A x - 2y_0 a^T x + t y_0^2 - t \\
&= \max_t \min_{||x||=s,\ y_0^2=1} x^T A x - 2y_0 a^T x + t y_0^2 - t \\
&= \max_t \min_{||x||^2+y_0^2=s^2+1} x^T A x - 2y_0 a^T x + t y_0^2 - t
\end{aligned}$$

$$= \max_t (s^2+1)\lambda_1(D(t)) - t$$

$$D(t) = \begin{pmatrix} t & -a^T \\ -a & A \end{pmatrix}.$$

unconstrained dual problem to TRS

$$D(t) = \begin{pmatrix} t & -a^T \\ -a & A \end{pmatrix}$$

$y = \begin{pmatrix} y_0 \\ x \end{pmatrix}$ normalized eigenvector for $\lambda_{\min} D(t)$

$$k(t) = (s^2 + 1)\lambda_{\min}(D(t)) - t,$$

$$*** \qquad \max_t k(t)$$

Note

$$k'(t) = (s^2 + 1)y_0^2 - 1 = 0$$

is feasibility again for x

SDP Primal-Dual Pair

$$\max_t k(t) = (s^2 + 1)\lambda_{\min}(D(t)) - t,$$

add the variable λ

$$\begin{array}{ll} \max & (s^2+1)\lambda - t \\ \text{s.t.} & D(t) \succeq \lambda I \end{array} \qquad (DSDP)$$

Lagrangian dual of this dual is:

$$\begin{array}{ll} \min & \operatorname{trace} D(0)X \\ \text{s.t.} & \operatorname{trace} X = s^2 + 1 \\ & X_{11} = 1 \\ & X \succeq 0 \end{array} \qquad (PSDP)$$

primal-dual interior point method:

approx. solve perturbed optimality conditions using Newton's method:

$$\operatorname{trace} X = s^2 + 1$$

$$X_{11} = 1$$

$$D(t) - \lambda I - Z = 0$$

$$\mu Z^{-1} - X = 0$$

$$X \succ 0, Z \succ 0$$

Dual Simplex Method

basic feasible dual solution at given t:

$$D(t) - \lambda_{\min}(D(t))I \succeq 0 \text{ and singular}$$

eigenvector $y = \begin{pmatrix} y_0 \\ x \end{pmatrix}$

normalize $y_0 = 1$

test primal feasibility of $X = yy^T$ to get better value for t (interpolation problem)

Alternatively, normalize $||y|| = 1$ with $y_0 \geq 0$, and exploit almost linear structure of

$$\frac{1}{y_0(t)} = \sqrt{s^2 + 1}$$

$$Z = D(t) - \lambda I$$

complementary slackness

$$\operatorname{trace} ZX = 0$$

$$0 = ZX = \begin{bmatrix} t - \lambda & -a^T \\ -a & A - \lambda I \end{bmatrix} \begin{bmatrix} X_{11} & y^T \\ y & \bar{X} \end{bmatrix}.$$

$$t = a^T y + \lambda;$$

$$\bar{X} a = (t - \lambda) y;$$

$$(A - \lambda I) y = a;$$

$$(A - \lambda I) \bar{X} = a y^T;$$

$$A - \lambda I \succeq 0.$$

$$\lambda_i(A) = \alpha_i$$

$J := \{i : a_i \neq 0\}$

$\det(D(t) - \lambda I) =$

$$\begin{aligned} &= (t-\lambda)\prod_{k=1}^{n}(\alpha_k - \lambda) - \sum_{k=1}^{n} a_k^2 \prod_{j\neq k}^{n}(\alpha_j - \lambda) \\ &= [t - \lambda - \sum_{j\in J}\frac{a_j^2}{\alpha_j - \lambda}]\prod_{k=1}^{n}(\alpha_k - \lambda) \\ &= d(\lambda)\prod_{k=1}^{n}(\alpha_k - \lambda) \end{aligned}$$

Easy case:

$$\lambda_1(D(t)) \text{ is simple and } \lambda_1(D(t)) < \alpha_1$$

if $\lambda < \alpha_1$ given. Then

$$D(d(\lambda)) - \lambda I \succeq 0 \text{ and singular.}$$

Let $y(t)$ be a normalized eigenvector corresponding to $\lambda_1(D(t))$ and denote its first component by $y_0(t)$. Then $y_0(t) \neq 0$. (wlog > 0)

$y_0(t) : \mathcal{R} \mapsto (0, 1)$ is strictly monotonically decreasing.

Theorem

$y(t) = \begin{pmatrix} y_0(t) \\ z(t) \end{pmatrix}$ normal eigenvector for

$$\lambda_1(D(t))$$

Then

$y_0(t) \neq 0$ and $v := \frac{1}{y_0(t)} z(t)$ is unique opt of

$$\min\{v^T A v - 2a^T v : v^T v = \frac{1 - y_0(t)^2}{y_0(t)^2}\}.$$

Conversely, if $v \in \mathbb{R}^n$ *and* $\lambda \in \mathbb{R}$ satisfy

$$(A - \lambda I)v = a; \quad A - \lambda I \succeq 0; \; v^T v = \frac{1 - y_0^2}{y_0^2}$$

thereby defining $y_0 > 0$,

then

$$y := y_0 \begin{pmatrix} 1 \\ v \end{pmatrix}$$

is eigenvector for $D(t)$ for $t := a^T v + \lambda$

and

$$\lambda_1(D(t)) = \lambda$$

Dual-Simplex Method

Initialization:

Find bounds for intervals of uncertainty.

$t_0^l \leq t^* \leq t_0^u$

$\mu^l \leq \mu^* \leq \mu^u$

Begin the main iterations:

while (feasibility or duality gap tols too large)

Part 1: Find a new estimate t_+.

$t_+ = \frac{t_c^l + t_c^u}{2}$ (default value)

Find intersection point of 2 tangent lines

to the graph of $k(t)$

update the upper bound μ^u

if last iterate, on infeasible side, good side

use inverse interpolation

elseif on feasible side, bad side

  use inverse interpolation

endif

Part 2: Update info for new estimate t_+.

if $\lambda > 0$ (wrong sign for Lagrange multiplier)

  if $t_+ <= t^*$ then **STOP**

else (correct sign for Lagrange multiplier)

  Update lower bound using dual function

endif

if $t_+ < t_*$ (bad side)

  Perform a PSDP (negative curvature)

    step, update bounds

else

  steepest descent step - update bounds

endif

Update the tolerances

endwhile

Tests on SUN SPARC station 1 using MATLAB

1. dimensions 1540 to 1565
   30 problems for each dimension
   density of nonzeros .01
   average iterations 4.4453
   average cpu time 54.1876 sec
   average work 1.73 times one Lanczos step

2. dimensions 1540 to 1558
   30 problems for each dimension
   density of nonzeros .013
   average iterations 4.1148
   average cpu time 56.1322 sec
   average work 1.75 times one Lanczos step
   comment: a multiple of the identity was added to get a positive definite Hessian

3. dimensions 1500 to 1565
   2 problems for each dimension
   density of nonzeros .01
   average iterations 4.2951
   average cpu time 57.0429 sec
   average work 4.8589 times one Lanczos step
   comment: hard case; multiplicity of smallest eigenvalue varied from 1 to 6

4. dimensions 1800 to 1865
   12 problems for each dimension
   density of nonzeros .005
   average iterations 4.2453
   average cpu time 28.1692 sec
   average work 4.2589 times one Lanczos step
   comment: hard case; multiplicity of smallest eigenvalue varied from 1 to 6

3.5 Sequential Quadratic Programming for NLP

•SQP uses a quadratic program to find the search direction, i.e. quadratic approximation of objective and linear approximations of constraints.

•Extend the classical SQP method to include a quadratic model of the constraints. Solve the Lagrangian relaxation of this QQP subproblem using SDP.

Until now we have used classical tools of nonlinear programming to develop and analyse a modern problem in linear optimization. In a reversal of roles, we now attempt to use semidefinite programming as the subproblem solver of a nonlinear optimization toolbox. This should be viewed as an application of semidefinite programming. The results presented here have been published before.

A proven approach for the unconstrained minimization of a function, $f(x), x \in \mathbb{R}^n$, is to build and solve a quadratic model at a local estimate $x^{(k)}$, that is, apply the Newton's method. In this chapter we propose a direct extension of this modeling approach to constrained minimization: A local quadratic model of both the objective function and the constraints is built; since this model is too hard to solve, it is relaxed using the Lagrangean dual, which is then solved by semidefinite programming techniques. The key idea in this approach is to use the latest technique of cone linear programming to obtain a better model than is usual in SQP methods and the key ingredient is the equivalence between the Lagrangean and semidefinite relaxations.

As illustration of how semidefinite programs is used to good effect, recall the well-known Rayleigh-Ritz quotient to obtain the smallest eigenvalue of a symmetric matrix A. An equivalent formulation yields (for example, see [33])

$$\lambda_1(A) = \min \left\{ x^T A x \mid x^T x = 1 \right\}. \tag{3.4}$$

One approach to prove this result involves Lagrange multipliers: the optimal x must be a stationary point of the Lagrangean $L(x, \lambda) = x^T Ax - \lambda(x^T x - 1)$. This shows that the optimal x is an eigenvector; and substitution into the objective function shows that the corresponding eigenvalue is the smallest. But now, consider instead, $x^T Ax = \operatorname{trace}(x^T Ax) = \operatorname{trace}(Axx^T)$ and let $X := xx^T$. We write the program (3.4) as

$$\min \left\{ \langle A, X \rangle \mid \langle I, X \rangle = 1, X \succeq 0, X = xx^T \right\},$$

where $\langle A, B \rangle = \operatorname{trace}(AB)$, the trace inner product; $A \succeq 0$ (resp. $A \succ 0$) denotes positive semidefiniteness (resp. positive definiteness); and $A \succeq B$ denotes $A - B \succeq 0$, i.e. the symmetric matrix space $\mathcal{S}^n$ is equipped with the Löwner partial order.

Note that the rank one constraint ($X = xx^T$) is redundant because we have only one constraint [66]. We therefore drop it and construct the dual to obtain

$$\max\left\{\lambda \mid \lambda I \preceq A, \lambda \in \mathbb{R}\right\},$$

which obviously has $\lambda_1(A)$ as optimal value. Since the dual has a strictly interior point, the primal attains the same value and we get the Rayleigh-Ritz result. The reader should concur that this has to be one of the simplest proof of (3.4). In this manner we use semidefinite programming to solve Lagrangean relaxations.

In this chapter, we wish to illustrate some of the strengths, both theoretical and practical, of considering semidefinite relaxations of quadratic programs as the tool of choice for solving Lagrangean relaxations that arise from quadratic models of general nonlinear programs.

3.5.1 The Simplest Case

Moving up in complexity we consider the unconstrained problem

$$\min\left\{f(x) \mid x \in \mathbb{R}^n\right\}.$$

When possible, the method of choice for this problem is Newton's method, which solves a quadratic model of the objective function. To ensure a solution (or convexity) of the model, Newton's method is often implemented within a Trust-Region, or Restricted-Step approach. This very efficient variation proceeds from an initial estimate of the solution; develops a second-order model of the objective function deemed valid in a region around the estimate; and finally solves the model (the trust-region subproblem)

$$\min\left\{q_0(d) := d^TQd + 2b^Td \mid q_1(d) := d^Td \leq \delta^2, d \in \mathbb{R}^n\right\}. \quad (3.4)$$

The model is constructed from $Q = \nabla^2 f(x^{(k)})$ (or an approximation of the Hessian), $b = \nabla f(x^{(k)})$ and the parameter δ represents the radius where the model is deemed valid. The trust-region may be scaled or even arise from a non-convex quadratic. A solution d is then used as the step to the next estimate $x^{(k+1)} = x^{(k)} + d$.

One of the interesting properties of (3.4), first shown in Stern and Wolkowicz [80] using semidefinite programming, is that even though generally non-convex, the problem exhibits no duality gaps. The Lagrangean dual of (3.4) is written as

$$\max\left\{-b^T(Q+\lambda I)^\dagger b - \lambda\delta^2 \;\middle|\; Q+\lambda I \succ 0, \lambda \geq 0\right\}, \tag{3.4}$$

a nonlinear, concave semidefinite program, where $(\cdot)^\dagger$ is the Moore-Penrose generalized inverse. In addition, the Lagrangean dual has been shown [71] equivalent to the following linear semidefinite program,

$$\max\left\{(\delta^2+1)\lambda - t \;\middle|\; \begin{bmatrix} t & b^T \\ b & Q \end{bmatrix} \succeq \lambda I, t \in \mathbb{R}, \lambda \geq 0\right\}. \qquad (3.4)$$

We take the dual of the above linear semidefinite program (3.4) to get a semidefinite program equivalent to (3.4).

$$\min\left\{\langle P_0, Y\rangle \;\middle|\; \langle E_0, Y\rangle = 1, \langle P_I, Y\rangle \leq \delta^2, Y \succeq 0\right\}. \qquad (3.4)$$

The variable in this program, Y, belongs to the cone of symmetric positive semidefinite matrices of dimension $(n+1)\times(n+1)$. Also,

$$P_0 = \begin{bmatrix} 0 & b^T \\ b & Q \end{bmatrix}, P_I = \begin{bmatrix} 0 & 0 \\ 0 & I \end{bmatrix}, E_0 = \begin{bmatrix} 1 & 0 \\ 0 & \mathbf{0} \end{bmatrix}.$$

The reader will note that (3.4) may be obtained as we did for the for the Rayleigh-Ritz program, by homogenization of (3.4), transformation to matrix space and then by dropping the rank one constraint. (We abuse the term homogenization to mean a quadratic function without a linear term.) We will do this in detail for a more general program later on.

This pair of linear primal-dual semidefinite programs (3.4, 3.4) have strict interior points. Therefore the optimal values are equal; moreover, they are both attained. Finally, part of the first column of the primal semidefinite solution, the matrix Y, is feasible for (3.4). And, possibly with an additional displacement, chosen in the nullspace of the Lagrangean, this first column yields the same objective value for (3.4) as its dual optimal. By this procedure, usually known as *lifting*, of (3.4) to the cone of semidefinite matrices, and projecting back (by the first column), we see that there are no duality gaps for (3.4). This was first shown in [80].

Theorem 3.5.1.1 *The optimal solution to (3.4) and to its Lagrangean dual problem (3.4) are attained and the corresponding* *objective values are equal.*

The interesting aspect of this theorem is that the Lagrangean dual is shown equivalent to a semidefinite program and its optimal value is deduced from this latter program. Therefore, interior-point algorithms, as developed in the previous chapters, may be used to solve (3.4), even if the objective function and the feasible set are non-convex. The result has been extended to upper and lower bounded trust-region subproblems but, interestingly, not to a finite number of constraints. With as few as two constraints, a duality gap may appear [87, 88].

3.5.2 Multiple Trust-Regions

Consider now a quadratic objective function constrained by multiple quadratics,

$$\min\left\{x^TQ_0x+2b_0^Tx-a_0 \mid x^TQ_kx+2b_k^Tx-a_k \leq 0, 1 \leq k \leq m, x \in \mathbb{R}^n\right\}. \tag{3.4}$$

As soon as two or more trust-regions are considered, the necessary and sufficient conditions that hold for one trust region may no longer be necessary for (3.4). This is reflected in the duality gap exhibited by some instances of multiple trust-region programs.

To directly derive the relaxations, we introduce the vector $y = (x_0\ x)^T$. We then require $x_0^2 = 1$ or, in terms of the new variable, $y^TE_0y = 1$, to get an homogeneous program equivalent to (3.4),

$$\min\left\{y^TP_0y \mid y^TE_0y = 1, y^TP_ky \leq 0, 1 \leq k \leq m, y \in \mathbb{R}^{n+1}\right\}, \tag{3.4}$$

where

$$E_0 = \begin{bmatrix} 1 & 0 \\ 0 & \mathbf{0} \end{bmatrix} \text{ and } P_k = \begin{bmatrix} -a_k & b_k^T \\ b_k & Q_k \end{bmatrix}, 0 \leq k \leq m.$$

The homogenization simplifies the notation and opens the way to the semidefinite relaxation: We rewrite (3.4) using matrix variables.

$$\min \left\{ \langle Y, P_0 \rangle \;\middle|\; \langle Y, E_0 \rangle = 1, \langle Y, P_k \rangle \leq 0, 1 \leq k \leq m, Y = yy^T \right\}.$$

The rank-one constraint is relaxed to a semidefinite constraint; a procedure we justify by showing its equivalence with the Lagrangean relaxation. After some rearrangement of terms, the Lagrangean dual of (3.4) reads

$$\max \left\{ \min \left\{ y^T (P_0 + \sum_{k=1}^{m} \lambda_k P_k + \lambda_0 E_0) y - \lambda_0 \mid y \in \mathbb{R}^{n+1} \right\} \;\middle|\; \lambda \in \mathbb{R} \times \mathbb{R}_+^m \right\}.$$

For the inner minimization to be bounded we must now have

$$P_0 + \sum_{k=1}^{m} \lambda_k P_k + \lambda_0 E_0 \succeq 0, \quad \text{which implies} \quad Q + \sum_{k=1}^{m} \lambda_k Q_k \succeq 0.$$

This, by the way, is where the duality gap arises. The standard necessary optimality conditions for (3.4) do not require the Hessian of the Lagrangean to be semidefinite. But the Lagrangean dual program we are deriving here requires the same Hessian to be semidefinite. We therefore cannot expect the primal variables corresponding to an optimal dual solution to be, in general, optimal for (3.4).

To complete the derivation, we note that the minimum over y will be attained at $y = 0$ from which we get the dual program

$$\max \left\{ -\lambda_0 \;\middle|\; P_0 + \lambda_0 E_0 + \sum_{k=1}^{m} \lambda_k P_k \succeq 0, \lambda \in \mathbb{R} \times \mathbb{R}^m_+ \right\}. \tag{3.4}$$

We have now justified the claim of equivalence of the Lagrangean and semidefinite relaxations since dropping the rank-one condition on the homogenized primal (3.5.2) or taking the semidefinite dual of (3.4) will result in the following, which we will therefore simply refer to as the relaxation of (3.4),

$$\min \left\{ \langle P_0, Y \rangle \mid \langle E_0, Y \rangle = 1, \langle P_k, Y \rangle \leq 0, 1 \leq k \leq m, Y \succeq 0 \right\}. \quad (3.4)$$

This resulting semidefinite relaxation of (3.4) is equivalent to the one considered in the literature ([78, 12, 69]).

The optimal value of the relaxation provides a lower bound for the (3.4). We now need an approximation for the optimum x. Feasibility properties of the first column of the semidefinite relaxation were first shown by Fujie and Kojima [21] for an equivalent problem with linear objective function. For an alternate view of this result, see old geometry KW paper??? from which we extract the next results. Consider the feasible set of the nonlinear program (3.4),

$$\hat{F} := \{x \in \mathbb{R}^n \mid x^T Q_k x + 2b_k^T x - a_k \le 0, 1 \le k \le m\};$$

the feasible set of the semidefinite program (3.4),

$$\tilde{F} := \{Y \succeq 0 \mid \langle E_0, Y \rangle = 1, \langle P_k, Y \rangle \le 0, 1 \le k \le m\};$$

and the projector map,

$$P_R\colon \mathcal{S}^n \to \mathbb{R}^n, \quad P_R(Y) = P_R\left(\begin{bmatrix} a & x^T \\ x & X \end{bmatrix}\right) = x.$$

Theorem 3.5.2.1 *Suppose that Y is a feasible solution of (3.4). The projected vector, $x = P_R(Y)$, is then feasible for all convex constraints of (3.4).*

Proof. Since we are concerned only with convex constraints, we may consider only those where $Q_k \succeq 0$ and compute

$$\begin{aligned} x^T Q_k x + 2b_k^T x - a_k - \langle P_k, Y \rangle &= x^T Q_k x - \langle Q_k, X \rangle \\ &= -\langle Q, X - xx^T \rangle. \end{aligned}$$

Since $Y \succeq 0$ implies $X - xx^T \succeq 0$, we obtain

$$\begin{aligned} x^T Q_k x + 2b_k^T x - a_k &= \langle P_k, Y \rangle - \langle Q_k, X - xx^T \rangle \\ &\leq \langle P_k, Y \rangle \\ &\leq 0. \end{aligned}$$

And therefore x is feasible for all convex constraints of (3.4). ■

This feasibility of the first column is interesting to consider in more detail. First, in the case of a problem where the quadratic constraints are convex (but maybe the objective is not) there is an obvious way to improve this first column solution when it is not optimal.

A optimal pair Y,λ to the semidefinite relaxation, if Y is not rank one, will in general map to a vector x for which complementarity fails but improving the objective value while remaining feasible is then easy.

Lemma 3.5.2.1 *Consider a (3.4) with convex constraints. If the semidefinite primal optimal solution* Y *is not rank one, let* $\tilde{x} = P_R(Y)$, *(part of the first column of* Y*). Then there is a* $\bar{x}$ *chosen in* $\mathcal{N}(Q_0 + \sum \lambda_k Q_k)$, *the nullspace of the Lagrangean, such that* $x = \tilde{x} + \bar{x}$, *is feasible and will improve the primal objective value of (3.4).*

The idea is to choose a displacement along the nullspace of the Lagrangean until one or more slack constraints is satisfied with equality. The value of the objective function is lowered since

$$0 = \bar{x}^T (Q_0 + \sum \lambda_k Q_k) \bar{x} \geq \bar{x}^T Q_0 \bar{x}$$

and therefore $(\tilde{x}+\bar{x})^T Q_0 (\tilde{x}+\bar{x}) \leq \tilde{x}^T Q_0 \tilde{x}$.

Consider now a more general case where the constraints may not be convex. Note that Theorem 3.5.2.1 implies that the projected first column x is feasible for any nonnegative combination of constraints,

$$\sum_{k=1}^{m} \lambda_k (x^T Q_k x + 2b_k^T x - a_k) \leq 0, \lambda \geq 0, \tag{3.4}$$

which results in a convex function. Thus we obtain feasible points for convex combinations of constraints of (3.4) as in (3.4) from feasible points Y of the relaxation (3.4), even when these are not rank one. Therefore the relaxation provides a convex approximation to the feasible set $\hat{F}$. However, it actually provides a better approximation than this would initially lead us to believe.

Let us define a *valid inequality* for (3.4) as

$$\sum_{k=1}^{m} \lambda_k (x^T Q_k x + 2b_k^T x - a_k) \le 0, \quad \text{where} \quad Q_0 + \sum_{k=1}^{m} \lambda_k Q_k \succeq 0, \lambda \ge 0. \tag{3.4}$$

These inequalities, an infinite number of them, are not, in general, convex. (Simply consider (3.4) where the objective is strictly convex while the constraint is not.) However, they provide geometric insight into the **SDP** relaxation. The set of vectors satisfying all valid inequalities,

$$\left\{ x \;\middle|\; \sum_{k=1}^{m} \lambda_k (x^T Q_k x + 2b_k^T x - a_k) \le 0, \quad Q_0 + \sum_{k=1}^{m} \lambda_k Q_k \succeq 0, \lambda \ge 0 \right\}$$

establishes a relation between the set of projected columns of **SDP** solutions and some intersection of the original constraints.

We now use the above geometric descriptions to provide an approximate solution to (3.4) from the optimum of **SDP**. We use the first column of the optimum Y but then we use the properties of the valid inequalities (3.4) to improve this column by moving onto a boundary of a valid inequality.

In the general case of a non-convex feasible region, we obtain a step, which, unlike Lemma 3.5.2.1, *attains* complementary slackness, though not necessarily feasibility. Again, the value of the objective function is improved. This additional step is a generalization of an idea introduced by Moré and Sorensen to solve (3.4) and there is an explicit expression for the step as there is for (3.4), given here in Lemma (3.5.2.2). We give the technical construction of the step in the following lemma and its value in Corollary 3.5.2.1.

Lemma 3.5.2.2 *Suppose that* λ *and*

$$Y = \begin{bmatrix} 1 & x^T \\ x & X \end{bmatrix}$$

are feasible for the primal-dual pair DSDP and PSDP, respectively. Let

$$y := \begin{bmatrix} 1 \\ x \end{bmatrix}, \qquad Z := P_0 + \lambda_0 E_0 + \sum_{k \in \mathcal{I}} \lambda_k P_k, \quad \mathcal{I} = \{1, \ldots m\}$$

and suppose that they satisfy $ZY = 0$.

Let the matrix Y *be factored as*

$$Y = TT^T,$$

where T *is* $(n+1) \times r$ *and full column rank* $r \geq 2$. *Let the matrix* S *be* $r \times (r-1)$ *and full column rank with* $\mathcal{R}(S) = \mathcal{N}(T_{1,:})$, *i.e. with range space given by the orthogonal complement to the first row of* T. *Define*

$$R := TS, \;\; \bar{P} := \sum_{k \in \mathcal{I}} \lambda_k P_k, \;\; c := y^T \bar{P} y,$$

$$K := R^T \left(\bar{P} y y^T \bar{P} - c\bar{P} \right) R.$$

Choose v such that

$$Rv \neq 0 \text{ and } v^T K v \geq 0, \tag{3.4}$$

and define

$$a := v^T R^T \bar{P} R v, \;\; b := 2 v^T R^T \bar{P} y.$$

Then, for z defined as follows, we have

$$TSv = \begin{bmatrix} 0 \\ z \end{bmatrix} \neq 0, \tag{3.4}$$

and

$$b^2 - 4ac \geq 0. \tag{3.4}$$

Moreover, if we define

$$\alpha := \begin{cases} \left(-b \pm \sqrt{b^2 - 4ac}\right) / (2a) & \text{if } a \neq 0 \\ \frac{-c}{b} & \text{if } a = 0, \end{cases}$$

and

$$w := y + \alpha \begin{bmatrix} 0 \\ z \end{bmatrix},$$

then

$$w^T \bar{P} w = 0, \;\; \textit{and } Zw = 0. \tag{3.4}$$

Proof. That (3.4) holds and $Zw = 0$ follows directly from construction of R and the assumption of complementary slackness, $ZY = 0$. Note that $ZY = ZTT^T = 0$ implies $ZT = 0$. We still need to show the equality of the quadratic form in (3.4). Now

$$w^T \bar{P} w = y^T \bar{P} y + \alpha 2 v^T R^T \bar{P} y + \alpha^2 v^T R^T \bar{P} R v.$$

(We assume that a w exists to make this quadratic 0. This may be seen from using the ordinary TRS with the given λ defining the single constraint.) The discriminant for this quadratic in α is defined in (3.4), where

$$b^2 - 4ac = 4v^T K v.$$

Therefore, the discriminant is nonnegative, and the quadratic has a real solution α as given by the standard formula. ∎

We now make explicit the value of the above lemma in finding an approximate solution to (3.4).

Corollary 3.5.2.1 *Suppose that Y,Z,λ,w are defined as in Lemma 3.5.2.2 above. Then the Lagrangean dual bound is attained by w as well as complementary slackness.*

Proof. That complementarity is attained is seen directly from the second equation of (3.4). And from both equations we obtain

$$0 = w^T Z w - w^T \bar{P} w = w^T (P_0 - \lambda_0 E_0) w.$$

Therefore, $w^T P_0 w = q_0(x + \alpha z) = -\lambda_0$, the dual Lagrangean bound. ■

3.5.3 Approximations of Nonlinear Programs

We assume the reader is familiar with *Sequential Quadratic Programming*, denoted **SQP**. We recall only the main features and refer the reader to ???? for details. The usual justifications for the application of **SQP** to the nonlinear program

$$\textbf{NEP} \qquad \min\left\{ f_0(x) \mid f_i(x) = 0, 1 \le i \le m, x \in \mathbb{R}^n \right\},$$

stem from applying Newton's method to obtain stationarity of its Lagrangean $\mathcal{L}(x, \lambda) := f_0(x) + \sum \lambda_i f_i(x)$,

$$\nabla f_0(x^*) + \sum \nabla f_i(x^*) \lambda_i^* = 0,$$

$$f(x^*) = 0.$$

We will sometimes use the notation $f(x) := [f_1(x) \dots f_m(x)]^T$ and $f'(x)$ for the first derivative. An iterative attempt at the non-linear system above by Newton's method with some simplification involving $d = x^{(k+1)} - x^{(k)}$ and $\delta_\lambda = \lambda^{(k+1)} - \lambda^{(k)}$, will produce the First-Order Newton Step,

$$\begin{bmatrix} \nabla^2 \mathcal{L}(x^{(k)}, \lambda^{(k)}) & f'(x^{(k)}) \\ f'(x^{(k)})^T & 0 \end{bmatrix} \begin{bmatrix} d \\ \lambda^{(k+1)} \end{bmatrix} = \begin{bmatrix} -\nabla f_0(x^{(k)}) \\ -f(x^{(k)}) \end{bmatrix}.$$

This system produces a direction d and a new vector of Lagrangean multiplier estimates $\lambda^{(k+1)}$. The key justification for **SQP** is that the system of equations (3.5.3) may be derived as the first-order necessary conditions of the quadratic program

$$\begin{aligned} \min \quad & f_0(x^{(k)}) + \nabla f_0(x^{(k)})^T d + \tfrac{1}{2} d^T \nabla^2 \mathcal{L}(x^{(k)}, \lambda^{(k)}) d \\ \text{s.t.} \quad & f_i(x^{(k)}) + \nabla f_i(x^{(k)})^T d = 0, \qquad 1 \leq i \leq m. \end{aligned} \tag{3.4}$$

Stationarity of the Lagrangean of (3.4) yields the first line of (3.5.3), and feasibility yields the second line. This is why **SQP** is viewed as an extension of Newton's method to constrained optimization.

It is now standard procedure to extend the above derivation to the inequality constrained program.

$$\min \left\{ f_0(x) \;\middle|\; f_i(x) \leq 0, 1 \leq i \leq m, x \in \mathbb{R}^n \right\}, \tag{3.4}$$

and obtain the subproblem,

$$\begin{aligned} \min \quad & f_0(x^{(k)}) + \nabla f_0(x^{(k)})^T d + \tfrac{1}{2} d^T \nabla^2 \mathcal{L}(x^{(k)}, \lambda^{(k)}) d \\ \text{s.t.} \quad & f_i(x^{(k)}) + \nabla f_i(x^{(k)})^T d \leq 0, \qquad 1 \leq i \leq m, \end{aligned} \tag{3.4}$$

In summary, the above derivation, from the Taylor first-order expansion of $\mathcal{L}(x, \lambda)$, obtained the standard **SQP** subproblem, which approximates the objective function to second order yet approximates the constraints only to first order. Consider now a second-order Taylor expansion of $\mathcal{L}(x, \lambda)$,

$$\begin{bmatrix} \sum \lambda_i \nabla f_i(x^{(k)}) + \nabla^2 \mathcal{L}(x^{(k)}, \lambda^{(k)})d + H_3(\delta_x, \delta_\lambda) \\ f'(x^{(k)})d + \frac{1}{2} d^T f''(x^{(k)})d \end{bmatrix} = \begin{bmatrix} -\nabla f_0(x^{(k)}) \\ -f(x^{(k)}) \end{bmatrix},$$

where we have grouped the third-order derivatives under the name H_3 because we intend to neglect them. Consider also replacing $\nabla^2 \mathcal{L}(x^{(k)}, \lambda^{(k)})$ by $\nabla^2 \mathcal{L}(x^{(k)}, \lambda^{(k+1)})$. We then obtain an approximation of the necessary optimality conditions which sits between a first and a second-order expansion and is obtained by solving

$$
\begin{aligned}
\min \quad & f_0(x^{(k)}) + \nabla f_0(x^{(k)})^T d + \tfrac{1}{2} d^T \nabla^2 f_0(x^{(k)}) d \\
\text{s.t.} \quad & f_i(x^{(k)}) + \nabla f_i(x^{(k)})^T d + \tfrac{1}{2} d^T \nabla^2 f_i(x^{(k)}) d \le 0, 1 \le i \le m \\
& d^T d \le \delta^2,
\end{aligned}
\tag{3.4}
$$

without the additional trust-region, which is added to ensure a bounded solution.

Such a straightforward subproblem has often been considered, but has, just as often, been discarded as unsolvable. One notable exception is an algorithm by Maany [47] developed, interestingly enough, because the standard **SQP** approach failed on the highly nonlinear orbital trajectory problems they were studying. (See [17].) Because (3.4) is a closer approximation to the original problem (3.4) than the quadratic program, we expect it to be a better subproblem to solve in a sequential programming approach and, in fact we have the following,

Lemma 3.5.3.1 *Assume that $x^{(k)}$ is feasible for (3.4). If the (3.4) subproblem is solved by $d = 0$ with multipliers λ, then the pair of vectors $x^{(k)}$ and λ satisfies the first-order conditions and second-order conditions of (3.4). Conversely, if $x^{(k)}$ and λ satisfy the first and second-order necessary conditions of (3.4), then the pair of vectors $d = 0$, λ satisfy the first and second-order conditions of (3.4).*

This implies that the (3.4) subproblem does better than the (3.4) subproblem since they both solve the first-order conditions but only the former guarantees second-order optimality conditions. This is expected of a trust-region approach.

Is also does better by providing second-order multiplier estimates in the sense that the multipliers $\lambda^{(k+1)}$ obtained from (3.4) satisfy

$$\min\left\{\|\nabla f_0(x^{(k)}) + \nabla^2\mathcal{L}(x^{(k)},\eta)d + \sum \eta_i \nabla f_i(x^{(k)})\|_2^2 \mid \eta \in \mathbb{R}^m\right\}.$$

If we are close to the solution we therefore obtain, directly from the solution of the subproblem, not only a good search direction in primal space, but better multiplier estimates than provided by the standard (3.4) subproblem. (For more details on second-order multiplier estimates, see

3.5.4 Quadratically Constrained Programming

Note that, for simplicity, we assume that our constraints are nonlinear. Linear constraints have to be treated differently, essentially squared, see [69]. Equivalently, linear constraints may be eliminated or mapped to a linear constraint in matrix space.

Homogenization of (3.4), obtained by adding a component d_0 to the vector d, together with the constraint $d_0^2 = 1$, yield the semidefinite relaxation,

$$\min \Big\{ \langle P_0, Y \rangle \;\Big|\; \langle E_0, Y \rangle = 1, \langle P_i, Y \rangle \le 0, 1 \le i \le m, \langle P_I, Y \rangle \le \delta^2 \Big\}, \tag{3.4}$$

where

$$P_i = \begin{bmatrix} -a_i & \nabla f_i(x^{(k)})^T & 0 \\ \nabla f_i(x^{(k)}) & \nabla^2 f_i(x^{(k)}) & 0 \\ 0 & 0 & 0 \end{bmatrix}, \;\; a_i = -2 f_i(x^{(k)}), \quad 0 \le i \le m,$$

and where E_0 and P_I have their usual definitions,

$$E_0 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, P_I = \begin{bmatrix} 0 & 0 \\ 0 & I \end{bmatrix},$$

and $Y \succeq 0$.

But this relaxation may be infeasible if the current estimate is too far from the feasible region. To overcome this difficulty in **SQP**, Vardi suggested a heuristic shift of the linear constraints. We do a related shift of our second-order constraints by allowing the additional component d_0 to take values between zero and one. That is, we change $d_0^2 = 1$ to $d_0^2 \leq 1$. This additional relaxation allows for a feasible subproblem. Of course we would want d_0 to be as close to 1 as possible and examination of the subproblem shows that it automatically tries to make d_0 "large". We need no heuristic to choose a Vardi-type parameter.

The dual program to (3.4) is then

$$\max \left\{ -\lambda_0 \,\middle|\, P_0 + \lambda_0 E_0 + \sum_{i=1}^{m} \lambda_i P_i + \lambda_I P_I \succeq 0, \lambda \in \mathbb{R} \times \mathbb{R}_+^m \right\}. \quad (3.4)$$

Solving the above primal-dual pair (3.4),(3.4), in the case of gap-free (3.4), is enough since, as we have seen, the first column is optimal for the quadratic approximation. But, in general, we need an appropriate merit function to ensure sufficient decrease at each step and guarantee global convergence of the algorithm, whether we use a line search or a trust-region strategy.

After solving the (3.4) subproblem for a direction $d \neq 0$, the next iterate is obtained by $x^{(k+1)} = x^{(k)} + d$. This new point serves for the expansion of a new problem by second-order polynomials and we iterate until the subproblem yields $d = 0$. As with any trust-region based algorithm, we adjust the trust-region radius according to the ratio of predicted improvement to actual improvement. At the end, we have a solution satisfying both first and second-order conditions of (3.4). Somewhat more formally, Algorithm sqqp??? describe the approach.

Sequential Quadratically Constrained Programming

1. - Given $f_i, \nabla f_i, \nabla^2 f_i, x^{(0)}$ *Functions and derivatives*
   - Given ϵ *Steplength tolerance*
   - $k := 0$ *Iteration count*

   REPEAT

   - $Y \in \operatorname{argmin}\{\langle P_0, Y\rangle : \langle P_i, Y\rangle \leq 0, \langle E_0, Y\rangle = 1, Y \succeq 0\}$; ——- *Solve semidefinite pair*
   - $\lambda^{(k+1)} \in \operatorname{argmax}\{-\lambda_0 : P_0 + \sum \lambda P_i + \lambda_0 E_0 \succeq 0, \lambda \in \mathbb{R} \times \mathbb{R}^m_+\}$;
   - $d := P_R(Y)$; *Project down by first column*
   - $x^{(k+1)} := x^{(k)} + d$; *New point*
   - $r^k := \frac{\varphi(x^{(k)}) - \varphi(x^{(k+1)})}{q_0(x^{(k)}) - q_0(x^{(k+1)})}$ *Decrease ratio*

   IF $(r^k < \frac{1}{4})$

   - $\delta = \delta/4$ *Bad model, shrink trust-region*

ELSE IF $(r^k > \frac{3}{4})$ and $\|x^{(k+1)} - x^{(k)}\| = \delta$

- $\delta = 2\delta$ *Good model, expand trust-region*

ENDIF ENDIF

- $k := k + 1$ *Bump iteration*

UNTIL $(\|d\| \leq \epsilon)$ *Attained optimality*

- Find maximal $d \in \mathcal{N}(\nabla^2 \mathcal{L})$ such that $f(x^{(k)} + d) \leq 0$
- $x^{(k)} := x^{(k)} + d;$ *Nullspace move*
- **return**$(x^{(k)}, \lambda^{(k)})$ *Primal iterate and multiplier*

If the (3.4) subproblem is convex, or more generally, if it is an instance without duality gaps, then solving the semidefinite relaxation, which may be done efficiently, will be enough since the primal semidefinite solution will be rank one. We will have a pair of primal-dual vectors satisfying the sufficient conditions for optimality of (3.4).

This takes care of the convex case and of many non-convex cases.

In other cases, we move along the nullspace of the Lagrangean until we hit one of the constraints. This nullspace-restricted step improves the objective value even if it does not lead to an optimal solution.

3.5.5 Conclusion

Efficient approaches to unconstrained optimization based on Newton's method all involve local quadratic models of the objective function. Yet for constrained optimization, the extension of Newton's method, **SQP**, uses linear approximations. Some second-order information is included in the model, but in an aggregate form.

In this chapter we have outlined an approach that deals more closely with the true quadratic model of the problem at hand. One of the key features is the relationship between the Lagrangean and

Semidefinite relaxations which leads to what we have called the **SQQP** algorithm for general nonlinear programs.

This algorithm builds second-order approximations of both the objective function and the constraints and then solves the Lagrangean relaxation of this quadratic model via semidefinite programming. The approach provides a stronger relaxation than the standard quadratic program used in **SQP** methods; at every step it provides better multiplier estimates; it handles potential infeasibility of the subproblem in a straight-forward manner; finally, it aims at solutions satisfying both first and second-order optimality conditions. Many implementation issues still need to be resolved but the recent advances in numerical solutions of large semidefinite programs encourage further study.

As a final note, we should make clear that it may turn out that the semidefinite relaxation is not exactly the right one to use for efficiency reasons. But the point remains that we are nowadays in a position to do better than the linear relaxations so popular during the seventies and eighties. Because we had at our disposals good linear programming solvers, the world seemed linear or, at least, mathematical models tried to make it so. We now have good solvers for quadratic programs, either via semidefinite relaxations or, possibly via some second-order cone relaxation, and we should make full use of these new tools.

3.6 Matrix Completions

3.6.1 Outline

- background
- Positive Semidefinite matrix completions
- Euclidean Distance matrix completions
- exploiting sparsity
- A New Characterization of EDMs using Gale Matrices

3.6.2 Positive Definite Completions of Partial Hermitian Matrices

- $\mathcal{G}(V, E)$ *finite undirected graph*, (node set V, edge set E)

- $A(\mathcal{G})$ is a $\mathcal{G}$*-partial matrix*, (a_{ij} defined *iff* $\{i, j\} \in E$)

- $A(\mathcal{G})$ is a $\mathcal{G}$*-partial positive matrix* if $a_{ij} = \overline{a_{ji}}, \forall \{i, j\} \in E$ and all existing principal minors are positive.

- with $\mathcal{J} = (V, \bar{E}), E \subset \bar{E}$ *a* $\mathcal{J}$*-partial matrix* $B(\mathcal{J})$ *extends* the $\mathcal{G}$-partial matrix $A(\mathcal{G})$ if $b_{ij} = a_{ij}, \forall \{i, j\} \in E$

- $\mathcal{G}$ is *positive completable* if every $\mathcal{G}$-partial positive matrix can be extended to a positive definite matrix.

$\mathcal{G}$ is **chordal** if there are no minimal cycles of length ≥ 4.
(i.e. every cycle of length ≥ 4 has a chord)

Theorem 3.6.2.1 *(Grone, Johnson, Sa, Wolkowicz)*
$\mathcal{G}$ is positive completable **iff** *$\mathcal{G}$ is chordal.* ∎

Equivalently, positive completability can be expressed as strict feasibility in SDP:

$$\begin{aligned} \operatorname{trace} E_{ij} P &= a_{ij}, \quad \forall \{i,j\} \in E \\ P &\succ 0 \end{aligned}$$

where $E_{ij} = e_i e_j^T + e_j e_k^T$

3.6.3 Approximate Positive Semidefinite Completions

GIVEN:

- $H = H^T \geq 0$ a real, nonnegative (elementwise) **symmetric matrix of weights**
  (with positive diagonal elements $H_{ii} > 0, \ \forall i$)
- $A = A^*$, the **given partial Hermitian matrix**
  i.e. some elements approximately fixed, others free (for notational purposes, assume free elements set to 0 if not specified.)

$||A||_F = \sqrt{\mathrm{trace}\, A^*A}$ *Frobenius norm,* $\circ$ denotes *Hadamard product.*

$$f(P) := ||H \circ (A - P)||_F^2$$

the weighted, best approximate, completion problem

$$\text{(AC)} \qquad \begin{array}{rcl} \mu^* := & \min & f(P) \\ & \text{subject to} & \mathcal{K}P = b \\ & & P \succeq 0, \end{array}$$

where $\mathcal{K} : \mathcal{H}^n \to \mathcal{C}^m$ linear operator

Lagrangian : $\quad L(P,y,\Lambda)=f(P)+\langle y,b-\mathcal{K}P\rangle-\operatorname{trace}\Lambda P$

Dual problem:

$$\text{(DAC)}\qquad \begin{array}{ll} \max & f(P)+\langle y,b-\mathcal{K}P\rangle-\operatorname{trace}\Lambda P \\ \text{subject to} & \nabla f(P)-\mathcal{K}^*y-\Lambda=0 \\ & \Lambda\succeq 0. \end{array}$$

Theorem 3.6.3.1 *The matrix* $\bar{P}\succeq 0$ *and vector-matrix* $\bar{y},\bar{\Lambda}\succeq 0$ *solve AC and DAC resp.* **iff**

$$\begin{array}{rcl} \mathcal{K}\bar{P}=b & & \textit{primal feas.} \\ 2H^{(2)}\circ(\bar{P}-A)-\mathcal{K}^*\bar{y}-\bar{\Lambda}=0 & & \textit{dual feas.} \\ \operatorname{trace}\bar{\Lambda}\bar{P}=0 & & \textit{compl. slack.} \end{array}$$

■

For simplicity, replace $\mathcal{K}$ with appropriate weights in H.
Use *square* perturbed optimality conditions.
(For $P, \Lambda \succeq 0$: $\operatorname{trace} P\Lambda = 0 \iff P\Lambda = 0$.)

$$\begin{array}{ll} 2H^{(2)} \circ (P - A) - \Lambda = 0 & \text{dual feasibility} \\ -P + \mu\Lambda^{-1} = 0 & \text{perturbed C.S.} \end{array}$$

Linearize second equation and solve for h and l

$$h = \mu\Lambda^{-1} - \mu\Lambda^{-1} l \Lambda^{-1} - P$$

$$l = \frac{1}{\mu}\left\{-\Lambda(P + h)\Lambda\right\} + \Lambda$$

Dual Step First: (if many elements of P are free; similar approach for primal step first if many elements are fixed)

We can eliminate the primal step h and solve for the dual step l.

$$\begin{aligned} l &= 2H^{(2)} \circ h + (2H^{(2)} \circ (P - A) - \Lambda) \\ &= 2H^{(2)} \circ (\mu\Lambda^{-1} - \mu\Lambda^{-1} \\ &\qquad\qquad l\Lambda^{-1} - P) + (2H^{(2)} \circ (P - A) - \Lambda). \end{aligned}$$

Equivalently, we get the Newton equation

$$2H^{(2)} \circ (\mu\Lambda^{-1} l \Lambda^{-1}) + l = 2H^{(2)} \circ (\mu\Lambda^{-1} - A) - \Lambda.$$

This shows that l, Λ have the same sparsity pattern as H (order is number of nonzeros in $H/2$.)

Each test appears on one line and includes 20 test problems.

dim	toler	Hdens./infty	Apsd	cond(A)	Hpd	min/max	iters
15	10^{-5}	.0751/.02	yes	222.5	no	8/17	10.3
15	10^{-6}	.1/.95	yes	19.6	no	10/23	15.6
15	10^{-6}	.01/.95	yes	21.	no	10/20	13.2
19	10^{-6}	.005/.1	yes	14.8	no	10/18	12.9
21	10^{-6}	.005/.1	yes	20.3	no	8/24	15
38	10^{-6}	1/.99	yes	49.6	yes	14/24	16.
45	10^{-6}	1/.99	yes	46.8	yes	15/22	17.
55	10^{-6}	1/.99	yes	37.2	yes	15/30	17.5
85	10^{-5}	.0219/.02	yes	1374.5	no	16/23	18.9
95	10^{-5}	.0206/.02	yes	2.7	no	8/14	11.1
95	10^{-6}	1/.999	yes	196.	yes	14/18	16.8
145	10^{-6}	.01/.997	yes	658.5	yes	13/17	14.9

data for primal-step-first (20 problems per test):
dimension; tolerance for duality gap;
density of nonzeros in H/ density of infinite values in H;
positive semidefiniteness of A; condition number of A; positive definiteness of H;
min and max number of iterations; average number of iterations.

dim	toler	Hdens./infty	Apsd	cond(A)	Hpd	min/max	iters
60	10^{-6}	.01/.001	yes	79.7	no	15/23	16.8
65	10^{-6}	.015/.001	yes	49.9	yes	18/24	21.3
83	10^{-6}	.007/.001	no	235.1	no	24/29	25.5
85	10^{-5}	.008/.001	yes	94.7	no	11/17	13.1
85	10^{-6}	.0075/.001	no	299.9	no	23/27	25.2
87	10^{-6}	.006/.001	yes	74.2	yes	14/19	16.9
89	10^{-6}	.006/.001	no	179.3	no	23/28	15.2
110	10^{-6}	.007/.001	yes	172.3	yes	15/20	17.8
155	10^{-6}	.01/0	yes	643.9	yes	14/18	15.3
655	10^{-6}	.017/0	yes	1.4	no	14/14	14.
755	10^{-6}	.002/0	yes	1.5	no	15/15	15.

data for dual-step-first (20 problems per test): dimension; tolerance for duality gap; density of nonzeros in H/ density of infinite values in H; positive semidefiniteness of A; condition number of A; positive definiteness of H;

(only one test for: 655,755; *bottleneck* is generating large random probs)

3.6.4 Euclidean Distance Matrix Completion Problem

What are EDMs?

- A **pre-distance matrix** (or dissimilarity matrix) is: an $n \times n$ symmetric matrix $D = (d_{ij})$ with nonnegative elements and zero diagonal
- A (squared) **Euclidean distance matrix** (EDM) is: a pre-distance matrix such that there exists points $x^1, x^2, \ldots, x^n$ in $\mathbb{R}^r$ such that

$$d_{ij} = \|x^i - x^j\|^2, \quad i, j = 1, 2, \ldots, n.$$

- The smallest value of r is called **the embedding dimension** of D. (r is always $\leq n-1$, e.g. translate x^1 to origin)

EDM problem:
Given a partial symmetric matrix A with certain elements specified, the Euclidean distance matrix completion problem (EDMCP) consists in finding the unspecified elements of A that make A a EDM.

WHY?

e.g.:

• The shape of an enzyme determines its chemical function. Once the shape is known, then the proper drug can be designed.

• distance geometry on molecules: Atoms are points in space with pairwise distances; find a set of points which yield those distances.

• **approximate EDMCP**, let: A be a pre-distance matrix, H be an $n \times n$ symmetric matrix with nonnegative elements, $\|A\|_F = \sqrt{\operatorname{trace} A^T A}$ denote the *Frobenius norm* of A.

• **objective function**

$$f(D) := \|H \circ (A - D)\|_F^2,$$

where $\circ$ denotes *Hadamard product.*

• **weighted, closest Euclidean distance matrix problem** is

$$(CDM_0) \quad \begin{array}{lll} \mu^* := & \min & f(D) \\ & \text{subject to} & D \in \mathcal{D}, \end{array}$$

where $\mathcal{D}$ denotes the (convex) cone of EDMs.

DISTANCE GEOMETRY

A pre-distance matrix D is a EDM if and only if D is negative semidefinite on

$$M := \left\{x \in \mathbb{R}^n : x^T e = 0\right\},$$

where e is the vector of all ones.

Define V $n \times (n-1)$, full column rank such that $V^T e = 0$. Then

$$J := VV^\dagger = I - \frac{ee^T}{n}$$

is the orthogonal projection onto M, where $V^\dagger$ denotes *Moore-Penrose generalized inverse.*

Define the **centered** and **hollow** subspaces

$$\begin{aligned} \mathcal{S}_C &:= \{B \in \mathcal{S}^n : Be = 0\}, \\ \mathcal{S}_H &:= \{D \in \mathcal{S}^n : \mathrm{diag}\,(D) = 0\}. \end{aligned}$$

Define the two linear operators

$$\mathcal{K}(B) \quad := \quad \mathrm{diag}(B)\, e^T + e\,\mathrm{diag}(B)^T - 2B,$$

$$\mathcal{T}(D) \quad := \quad -\tfrac{1}{2} JDJ.$$

The operator $-2\mathcal{T}$ is an orthogonal projection onto $\mathcal{S}_C$.

Theorem 3.6.4.1

$$\begin{aligned} \mathcal{K}(\mathcal{S}_C) &= \mathcal{S}_H, \\ \mathcal{T}(\mathcal{S}_H) &= \mathcal{S}_C, \end{aligned}$$

and $\mathcal{K}_{|\mathcal{S}_C}$ and $\mathcal{T}_{|\mathcal{S}_H}$ are inverses of each other. ∎

D (hollow matrix) is EDM $\iff B = \mathcal{T}(D) \succeq 0$

D is EDM $\iff D = \mathcal{K}(B)$, for some B with $Be = 0$ and $B \succeq 0$.

In this case the embedding dimension r is given by the rank of B. Moreover if $B = XX^T$, then the coordinates of the points $x^1, x^2, \ldots, x^n$ that generate D are given by the rows of X and, since $Be = 0$, it follows that the origin coincides with the centroid of these points.

The cone of EDMs, $\mathcal{D}$, has empty interior.

This can cause problems for interior-point methods.

$$V \cdot V : \mathcal{S}_{n-1} \rightarrow \mathcal{S}_n$$

$$V \cdot V : \mathcal{P}_{n-1} \rightarrow \mathcal{P}_n$$

Define the composite operators

$$\mathcal{K}_V(X) \quad := \quad \mathcal{K}(VXV^T),$$

and

$$\mathcal{T}_V(D) \quad := \quad V^\dagger \mathcal{T}(D)(V^\dagger)^T = -\frac{1}{2} V^\dagger D (V^\dagger)^T.$$

LEMMA

$$\mathcal{K}_V(\mathcal{S}_{n-1}) = \mathcal{S}_H,$$
$$\mathcal{T}_V(\mathcal{S}_H) = \mathcal{S}_{n-1},$$

and $\mathcal{K}_V$ and $\mathcal{T}_V$ are inverses of each other on these two spaces. ■

COROLLARY

$$\begin{aligned} \mathcal{K}_V(\mathcal{P}) &= \mathcal{D}, \\ \mathcal{T}_V(\mathcal{E}) &= \mathcal{P}. \end{aligned}$$

■

Summary

(Re)Define the closest EDM problem:

$$\begin{aligned} f(X) &:= \|H \circ (A - \mathcal{K}_V(X))\|_F^2 \\ &= \|H \circ \mathcal{K}_V(B - X)\|_F^2, \end{aligned}$$

where $B = \mathcal{T}_V(A)$.
($\mathcal{K}_V$ and $\mathcal{T}_V$ are both linear operators)

$$\begin{array}{llll} & \mu^* := & \min & f(X) \\ (CDM) & & \textbf{subject to} & \mathcal{A}X = b \\ & & & X \succeq 0. \end{array}$$

The additional constraint using $\mathcal{A} : \mathcal{S}_{n-1} \longrightarrow \mathbb{R}^m$, could represent some of the fixed elements in the given matrix A.

Primal-Dual Interior-Point Framework:

1. derive a dual program
2. state optimality conditions for log-barrier problem (perturbed primal-dual optimality conditions)
3. find a search direction for solving the perturbed optimality conditions
4. take a step and backtrack to stay strictly feasible (positive definite)
5. Update and go to Step 3 (adaptive update of log-barrier parameter)

Step 1. derive a dual program:

$\Lambda \in \mathcal{S}_{n-1}, \Lambda \succeq 0$ and $y \in R^m$,

Lagrangian is

$$L(X, y, \Lambda) = f(X) + \langle y, b - \mathcal{A}(X) \rangle - \langle \Lambda, X \rangle$$

primal program (CDM) is

$$= \min_{X} \max_{\substack{y \\ \Lambda \succeq 0}} L(X, y, \Lambda).$$

dual program is:

$$= \max_{\substack{y \\ \Lambda \succeq 0}} \min_{X} L(X, y, \Lambda),$$

The inner minimization of the convex, in X, Lagrangian is unconstrained so we add the hidden constraint which makes the minimization redundant.

dual program (DCDM)

$$\max_{\substack{\nabla f(X)-\mathcal{A}^* y=\Lambda \\ \Lambda \succeq 0}} f(X)+\langle y, b-\mathcal{A}(X)\rangle-\operatorname{trace}\Lambda X.$$

or

$$\begin{array}{cc} \max & f(X)+\langle y, b-\mathcal{A}(X)\rangle-\operatorname{trace}\Lambda X \\ \text{subject to} & \nabla f(X)-\mathcal{A}^* y-\Lambda=0 \\ & \Lambda \succeq 0,(X \succeq 0). \end{array}$$

the duality gap,

$f(X) - (f(X) + \langle y, b - \mathcal{A}(X) \rangle - \operatorname{trace} \Lambda X)$,

in the case of primal and dual feasibility, is given by the complementary slackness condition:

$$\operatorname{trace} X(\mathcal{K}_V^*(H^{(2)} \circ \mathcal{K}_V(X - B)) - \mathcal{A}^* y) = 0,$$

or equivalently

$$X(\mathcal{K}_V^*(H^{(2)} \circ \mathcal{K}_V(X - B)) - \mathcal{A}^* y) = 0,$$

where $H^{(2)} = H \circ H$.

Theorem 3.6.4.2 *Suppose that Slater's condition holds for CDM. Then $\bar{X} \succeq 0$, and $\bar{y}$, $\bar{\Lambda} \succeq 0$ solve (CDM) and (DCDM), respectively, if and only if the following three equations hold.*

$$\begin{array}{rc} \mathcal{A}(\bar{X}) = b & \textit{prim. feas.} \\ 2\mathcal{K}_V^*(H^{(2)} \circ \mathcal{K}_V(\bar{X} - B)) - \mathcal{A}^*\bar{y} - \bar{\Lambda} = 0 & \textit{dual feas.} \\ \operatorname{trace} \bar{\Lambda}\bar{X} = 0 & \textit{C.S.} \end{array}$$

■

Lemma 3.6.4.1 *Let H be an $n \times n$ symmetric matrix with nonnegative elements and 0 diagonal such that the graph of H is connected. Then*

$$\mathcal{K}_V^*(H^{(2)} \circ \mathcal{K}_V(I)) \succ 0,$$

where $I \in \mathcal{S}_{n-1}$ is the identity matrix. ∎

i.e. when $\mathcal{A} = 0$, we have a Slater point for the dual (and primal)

Step 2. state optimality conditions for log-barrier problem (perturbed primal-dual optimality conditions):

The log-barrier problem for (CDM) is

$$\min_{X \succ 0} B_\mu(X) := f(X) - \mu \log \det(X),$$

where $\mu \downarrow 0$.

For each $\mu > 0$ we take one Newton step for solving the stationarity condition

$$\nabla B_\mu(X) = 2\mathcal{K}_V^*(H^{(2)} \circ \mathcal{K}_V(X - B)) - \mu X^{-1} = 0.$$

Let

$$C := 2\mathcal{K}_V^*(H^{(2)} \circ \mathcal{K}_V(B)) = 2\mathcal{K}_V^*(H^{(2)} \circ A).$$

Then the stationarity condition is equivalent to

$$\nabla B_\mu(X) = 2\mathcal{K}_V^* \left(H^{(2)} \circ \mathcal{K}_V(X)\right) - C - \mu X^{-1} = 0.$$

equating $\Lambda = \mu X^{-1}$ and multiplying through by X

optimality conditions, $F := \begin{pmatrix} F_d \\ F_c \end{pmatrix} = 0,$

$$\begin{array}{rcll} 2\mathcal{K}_V^*\left(H^{(2)} \circ \mathcal{K}_V(X)\right) - C - \Lambda & = & 0 & \text{dual feas.} \\ \Lambda X - \mu I & = & 0 & \text{pert. C.S.,} \end{array}$$

(an OVERDETERMINED nonlinear system since ΛX not symmetric)

estimate of the barrier parameter

$$\mu = \frac{1}{n-1} \operatorname{trace} \Lambda X$$

σ_k centering parameter

$\mathcal{F}^0$ set of strictly feasible primal-dual points

F' derivative of F

Algorithm 3.6.4.1 *(p-d i-p framework:)*

Given $(X^0, \Lambda^0) \in \mathcal{F}^0$

for $k = 0, 1, 2 \ldots$

solve *for the search direction*

$$F'(X^k, \Lambda^k) \begin{pmatrix} \delta X^k \\ \delta \Lambda^k \end{pmatrix} = \begin{pmatrix} -F_d \\ -\Lambda^k X^k + \sigma_k \mu_k I \end{pmatrix}$$

where σ_k *centering,* $\mu_k = \frac{\operatorname{trace} X^k \Lambda^k}{(n-1)}$

$$(X^{k+1}, \Lambda^{k+1}) = (X^k, \Lambda^k) + \alpha_k (\delta X^k, \delta \Lambda^k)$$

so that $(X^{k+1}, \Lambda^{k+1}) \succ 0$

end (for).

search direction (Gauss-Newton direction, cannot use *square* system because of $\mathcal{K}_V$) - the Frobenius norm lss of $F's = -F$, i.e.

$$\begin{aligned} 2\mathcal{K}_V^* \left(H^{(2)} \circ \mathcal{K}_V(h)\right) - l &= -F_d \\ \Lambda h + lX &= -F_c. \end{aligned}$$

$t(n) = \frac{(n+1)n}{2}$ be dimension of $\mathcal{S}^n$.

$$F's = \begin{bmatrix} F'_{u1} & F'_{u2} \\ F'_{l1} & F'_{l2} \end{bmatrix} \begin{pmatrix} h \\ l \end{pmatrix} = rhs = \begin{pmatrix} rhs_1 \\ rhs_2 \end{pmatrix}.$$

Operator F' maps $\mathbb{R}^{2(t(n-1))}$ to $\mathbb{R}^{t(n-1)+(n-1)^2}$.

3.6.5 Large Sparse Problems

Instead of projecting and reducing the dimension to get Slater's condition, add a variable and increase the dimension.

Lemma 3.6.5.1 *Let*

$$\begin{array}{lcl}
\mathcal{D} & := & \{X \in \mathcal{S}_H : v^T e = 0 \quad \Rightarrow \quad v^T X v \le 0\}, \\
\mathcal{D}_{-1} & := & \left\{X \in \mathcal{S}_H : 0 = \max\{v^T X v : v^T E v = 0\}\right\}, \\
\mathcal{D}_0 & := & \{X \in \mathcal{S}_H : X - \alpha e e^T \preceq 0, \quad \textit{for some } \alpha\}, \\
\mathcal{D}_1 & := & \{X \in \mathcal{S}_H : X - \alpha e e^T \preceq 0, \quad \forall\ \alpha \ge \bar{\alpha},\ \textit{for some } \bar{\alpha}\}.
\end{array}$$

Then

$$\operatorname{ri}(\mathcal{D}) \subset \mathcal{D}_0 = \mathcal{D}_1 \subset \mathcal{D} \subset \mathcal{D}_{-1} \subset \overline{\mathcal{D}_0}. \tag{3.5}$$

Proof. That $\mathcal{D} = \mathcal{D}_{-1}$ is clear and provides motivation for the other inclusions based on Lagrange multipliers. Suppose $\bar{X} \in \mathrm{ri}\,(\mathcal{D})$ (i.e. $v^T e = 0, v \neq 0 \Rightarrow v^T \bar{X} v < 0$) but $\bar{X} \notin \mathcal{D}_0$. Then, for each $\alpha \geq 0$, there exists w_α with $||w_\alpha|| = 1$, such that $w_\alpha \to \bar{w}$, as $\alpha \to \infty$ and

$$w_\alpha^T (\bar{X} - \alpha e e^T) w_\alpha > 0, \quad \forall\, \alpha \geq 0, \qquad \text{i.e.}$$

$$w_\alpha^T \bar{X} w_\alpha > \alpha w_\alpha^T e e^T w_\alpha, \quad \forall\, \alpha \geq 0.$$

Since w_α converges and the left-hand-side of the above inequality must be finite, this implies that $e^T \bar{w} = \bar{w}^T \bar{X} \bar{w} = 0$, a contradiction. Therefore, $\mathrm{ri}\,(\mathcal{D}) \subset \mathcal{D}_0$. That $\mathcal{D}_0 = \mathcal{D}_1$ is clear.

Now suppose that $\bar{X} - \alpha e e^T \preceq 0,\ \alpha \geq 0$. Let $v^T e = 0$. Then $0 \geq v^T (\bar{X} - \alpha e e^T) v = v^T \bar{X} v$, i.e. $\mathcal{D}_0 \subset \mathcal{D}$. The final inclusion comes from the first and the fact that $\mathcal{D}$ is closed. ∎

let: $E = ee^T$; $f(P) := \|H \circ (A - P)\|_F^2$;
$\mathcal{K}$ lin. operator with constraint $\operatorname{diag}(P) = 0$.
primal problem is:

$$\begin{array}{rccc} & \mu^* := & \min & f(P) \\ \text{(CDM)} & & \text{subject to} & \mathcal{K}P = b \\ & & & \alpha E - P \succeq 0 \end{array}$$

and **dual problem (DCDM)** is

$$\begin{array}{rc} \mu^* = \nu^* := \max & f(P) + \langle y, b - KP \rangle - \operatorname{trace} \Lambda(\alpha E - P) \\ \text{s.t.} & \nabla_P f(P) - \mathcal{K}^* y + \Lambda = 0 \\ & -\operatorname{trace} \Lambda E = 0 \\ & \Lambda \succeq 0. \end{array}$$

(Slater's holds for primal but fails for dual.)

(perturbed) Optimality Conditions are:

$$
\begin{aligned}
\operatorname{diag}(P) &= 0 && \text{primal feas.} \\
2H^{(2)} \circ (P - A) - \operatorname{Diag}(y) + \Lambda &= 0, && \\
- \operatorname{trace} \Lambda E &= 0 && \text{dual feas.} \\
-(\alpha E - P) + \mu \Lambda^{-1} &= 0, && \text{pert. C.S.}
\end{aligned}
$$

$$\begin{array}{rlr} h & \text{denotes the step for} & P \\ w & \text{denotes the step for} & \alpha \\ l & \text{denotes the step for} & \Lambda \\ s & \text{denotes the step for} & y. \end{array}$$

maintain

$$\operatorname{diag}(h) = \operatorname{diag}(P) = 0$$

linearization of complementary slackness

$$-(\alpha + w)E + (P + h) + \mu\Lambda^{-1} - \mu\Lambda^{-1} l \Lambda^{-1} \quad = \quad 0,$$

solve for h

$$h \quad = \quad -\mu\Lambda^{-1} + \mu\Lambda^{-1} l \Lambda^{-1} - P + (\alpha + w)E.$$

(or solve for l)

linearization dual feasibility

$$\begin{aligned} 2H^{(2)} \circ h - \mathrm{Diag}\,(s) + l \quad &= \quad -(2H^{(2)} \circ (P - A) \\ &\qquad - \mathrm{Diag}\,(y) + \Lambda) \\ -\mathrm{trace}\, lE \quad &= \quad \mathrm{trace}\, \Lambda E \end{aligned}$$

substitute for h, s

Newton equation is

$$
\begin{aligned}
2H^{(2)} \circ \left(wE + \mu\Lambda^{-1} l \Lambda^{-1}\right) - \operatorname{Diag} \operatorname{diag}(l) + l \\
= 2H^{(2)} \circ \left\{\mu\Lambda^{-1} + A - \alpha E\right\} + \operatorname{Diag}(y) - \Lambda \\
\operatorname{diag}\left(\mu\Lambda^{-1} l \Lambda^{-1}\right) + we \\
= \operatorname{diag}\left(\mu\Lambda^{-1}\right) - \alpha e \\
\operatorname{trace}(lE) = -\operatorname{trace}(\Lambda E).
\end{aligned}
$$

square system, order $1 + nnz$ where nnz are the number of nonzeros in the upper triangular part of H, ($\operatorname{diag}(H) = 0$).

F denotes $(nnz+n)\times 2$ matrix
row p contains indices of the p-th nonzero, upper triangular, element of $H+I$ ordered by columns,

$$\{(F_{p1},F_{p2})_{p=1,\dots nnz+n}\} \\ = \{ij : H_{ij}\neq 0, i\leq j, \text{ ordered by columns}\}.$$

δ_{ij} is *Kronecker delta function*
$\delta_{(ij)(kl)}$ is 1 if $(ij)=(kl)$, 0 otherwise.
$E_{ij}=\left(e_ie_j^T+e_j^Te_i\right)/\sqrt{2}$, ij unit matrix in $\mathcal{S}^n$, where
$E_{ij}=\left(e_ie_j^T+e_j^Te_i\right)/2$ if $i=j$.
(orthonormal basis of $\mathcal{S}^n$)

operator equation:

$$
\begin{aligned}
&k \neq l, i \neq j \text{ LHS} = \\
&= \operatorname{trace} E_{kl} \left\{ 2H^{(2)} \circ \left(\mu \Lambda^{-1} E_{ij} \Lambda^{-1} \right) - \operatorname{Diag} \operatorname{diag} (E_{ij}) \right. \\
&\qquad\qquad\qquad \left. + E_{ij} \right\} \\
&= \mu \operatorname{trace} (e_k e_l^T + e_l e_k^T) \left(H^{(2)} \circ \Lambda^{-1} (e_i e_j^T + e_j e_i^T) \Lambda^{-1} \right) \\
&\qquad\qquad\qquad + \delta_{(ij)(kl)} \\
&= \mu \left[2e_l^T \left(H^{(2)} \circ \Lambda_{:,i}^{-1} \Lambda_{j:}^{-1} \right) e_k + 2e_k^T \left(H^{(2)} \circ \Lambda_{:,i}^{-1} \Lambda_{j:}^{-1} \right) e_l \right] \\
&\qquad\qquad + \delta_{(ij)(kl)};
\end{aligned}
$$

$k \neq l, i \neq j$ LHS $=$

$$= 2\mu H_{kl}^{(2)} \left(\Lambda_{li}^{-1}\Lambda_{jk}^{-1} + \Lambda_{ki}^{-1}\Lambda_{jl}^{-1}\right) + \delta_{(ij)(kl)};$$

$k \neq l, i = j$ LHS $=$

$$= \operatorname{trace} E_{kl} \left\{2\mu H^{(2)} \circ \left[\Lambda^{-1} E_{jj} \Lambda^{-1}\right] - \operatorname{Diag} \operatorname{diag}\left(E_{jj}\right) + E_{jj}\right\}$$

$$= 2\sqrt{2}\mu \operatorname{trace} e_k e_l^T \left(H^{(2)} \circ \Lambda^{-1} e_j e_j^T \Lambda^{-1}\right)$$

$$= 2\sqrt{2}\mu H_{kl}^{(2)} \left(\Lambda_{lj}^{-1}\Lambda_{jk}^{-1}\right);$$

$k = l, i \neq j$ LHS $=$

$$= \sqrt{2}\mu \Lambda_{ki}^{-1}\Lambda_{jk}^{-1}, \quad k = 1, \ldots n;$$

$k = l, i = j$ LHS $=$

$$= \mu \Lambda_{ki}^{-1}\Lambda_{ik}^{-1}, \quad k = 1, \ldots n.$$

last column of LHS, matrix $l = 0$ and $w = 1$:

$$
\begin{aligned}
w = 1, k \neq l \text{ LHS} &= \operatorname{trace}\left(E_{kl}(2H^{(2)} \circ E)\right); \\
w = 1, k = l \text{ LHS} &= 1.
\end{aligned}
$$

last row of LHS:

$$
\begin{aligned}
i \neq j \text{ LHS} &= \operatorname{trace}\left(E_{ij}E)\right) = \sqrt{2}; \\
i = j \text{ LHS} &= 1.
\end{aligned}
$$

Newton system is:

$$\operatorname{sMat}\left[L(\operatorname{svec}(l))\right] = \operatorname{sMat}\left[\operatorname{svec}(RHS)\right],$$

$\operatorname{svec}(S)$ vector formed from nonzero elements of columns of upper triangular part, where strict upper triangular part is multiplied by $\sqrt{2}$. (trace $XY = \operatorname{svec}(X)^T \operatorname{svec}(Y)$, i.e. isometry) sMat is inverse

Solve for $\operatorname{svec}(l)$:

$$L(\operatorname{svec}(l)) = \operatorname{svec}(RHS).$$

$$L_{pq} =$$

$$\begin{cases} 2\mu H^{(2)}_{F_{p_2},F_{p_1}} \left(\Lambda^{-1}_{F_{p_2},F_{q_1}} \Lambda^{-1}_{F_{q_2},F_{p_1}} + \Lambda^{-1}_{F_{p_1},F_{q_1}} \Lambda^{-1}_{F_{q_2},F_{p_2}} \right) \\ \qquad \text{if } p \neq q,\ k \neq l,\ i \neq j; \\ 2\sqrt{2}\mu H^{(2)}_{F_{p_2},F_{p_1}} \left(\Lambda^{-1}_{F_{p_2},F_{q_2}} \Lambda^{-1}_{F_{q_2},F_{p_1}} \right) \\ \qquad \text{if } p \neq q,\ k \neq l,\ i = j; \\ 2\sqrt{2}\mu H^{(2)}_{F_{p_2},F_{p_1}} \left(\Lambda^{-1}_{F_{p_2},F_{q_2}} \Lambda^{-1}_{F_{q_2},F_{p_1}} \right) \\ \qquad \text{if } p = q,\ k \neq l,\ i = j; \\ 2\mu H^{(2)}_{F_{p_2},F_{p_1}} \left(\Lambda^{-1}_{F_{p_2},F_{q_1}} \Lambda^{-1}_{F_{q_2},F_{p_1}} + \Lambda^{-1}_{F_{p_1},F_{q_1}} \Lambda^{-1}_{F_{q_2},F_{p_2}} \right) + 1 \\ \qquad \text{if } p = q, k \neq l,\ i \neq j; \end{cases}$$

$$
\begin{cases}
\sqrt{2}\mu\Lambda^{-1}_{F_{p_1},F_{q_1}}\Lambda^{-1}_{F_{q_2},F_{p_1}} & \\
\qquad\qquad\qquad \text{if } k = l,\ i \neq j; & \\
\mu\Lambda^{-1}_{F_{p_1},F_{q_1}}\Lambda^{-1}_{F_{q_1},F_{p_1}} & \text{if } k = l,\ i = j \\
2\sqrt(2)H^{(2)}_{F_{p_2},F_{p_1}} & \text{if } w = 1,\ k \neq l \\
1 & \text{if } w = 1,\ k = l.
\end{cases}
$$

The p-th row calculated using Hadamard product of pairs of columns of Λ^{-1},

$$
\Lambda^{-1}_{F_{p_2},F_{:,1}} \circ \Lambda^{-1}_{F_{p_1},F_{:,2}}.
$$

complete vectorization

$p = kl, k \leq l$, and last row, component of the right-hand-side of the system is

$RHS_p =$

$$\begin{cases} \sqrt{2}\left(2H_p^{(2)} \circ \{\mu\Lambda_p^{-1} + A_p - \alpha\} - \Lambda_p\right), & \text{if } k \neq l \\ \mu\Lambda_{kk}^{-1} - \alpha & \text{if } k = l \\ -\operatorname{trace}(\Lambda E) & \text{last row} \end{cases}$$

3.6.6 On a New Characterization of EDMs

The above application for EDM uses

$$D = \lambda E - P, \; P \succeq 0, \; \operatorname{diag} P = \lambda e, \; \lambda \geq 0;$$

this raises the **question**:

> **which** *Euclidean distance matrices D can be expressed as $D = \lambda(E - C)$ for some nonnegative scalar λ and some correlation matrix C, where E is the matrix of all ones.*

We **show (and characterize)** that the cones

$$\operatorname{cone}(E - \mathcal{E}_n) \subsetneq \overline{\operatorname{cone}(E - \mathcal{E}_n)} = \mathcal{D}_n,$$

where $\mathcal{E}_n$ is the elliptope (set of correlation matrices) and $\mathcal{D}_n$ is the (closed convex) cone of Euclidean distance matrices.

The characterization is given using the Gale transform of the points generating D. We also show that given points $p^1, p^2, \ldots, p^n \in \Re^r$, for any scalars $\lambda_1, \lambda_2, \ldots, \lambda_n$ such that

$$\sum_{j=1}^{n} \lambda_j \, p^j = 0, \qquad \sum_{j=1}^{n} \lambda_j = 0,$$

we have

$$\sum_{j=1}^{n} \lambda_j \, \|p^i - p^j\|^2 = \alpha \text{ for all } i = 1, \ldots, n,$$

for some scalar α independent of i.

Recall:

- $n \times n$ matrix $D = (d_{ij})$ is *Euclidean distance matrix (EDM)* if

$$\exists \quad p^1, p^2, \ldots, p^n \in \mathbb{R}^r, \text{ such that } \|p^i - p^j\|^2 = d_{ij}, \quad \forall i, j$$

- dimension of smallest Euclidean space containing $p^1, p^2, \ldots, p^n$ is the *embedding dimension* of D.
- D (hollow) EDM $\iff$ D negative semidefinite on $e^\perp$
- the set of $n \times n$ EDM matrices is a closed convex cone, $\mathcal{D}_n$

Since $\mathcal{E}_n$ denotes set of $n \times n$ *correlation matrices*, i.e., set of positive semidefinite symmetric matrices whose diagonal is equal to e, we are using the well known fact [16] that $\mathcal{D}_n$ is the tangent cone of $\mathcal{E}_n$ at E, the matrix of all ones, i.e.,

$$\mathcal{D}_n = \overline{\text{cone }(E - \mathcal{E}_n)} = \overline{\{\lambda(E - C) : \lambda \geq 0, C \in \mathcal{E}_n\}}, \tag{3.6}$$

where $\bar{\cdot}$ denotes closure.

In general, it is difficult to determine whether the generated cone of a set, $cone(C)$ is closed, though there are sufficient conditions, e.g. C is convex and compact and 0 is not in C.

- Let V be $n \times (n-1)$ such that

$$V^T e = 0 \,, \quad V^T V = I_{n-1} \,. \tag{3.7}$$

- orthogonal projection on $M = e^\perp$, denoted by J, is $J := VV^T = I - ee^T/n$.
- D hollow is EDM $\iff$ $B := -\frac{1}{2}\, JDJ \succeq 0$
- embedding dimension of D is equal, r, rank of B.
- the points $p^1, p^2, \ldots, p^n$ that generate D are given by the rows of the $n \times r$ matrix P where $B := PP^T$.
- since $Be = 0$, centroid of the points p^i, $i = 1, \ldots, n$ coincides with the origin

Let $p^1, p^2, \ldots, p^n$ be points in $\mathbb{R}^r$ whose centroid coincides with the origin, and are not contained in a proper hyperplane. Then

$$P := \begin{bmatrix} {p^1}^T \\ {p^2}^T \\ \vdots \\ {p^n}^T \end{bmatrix}$$

is of rank r. Let $B = PP^T$. Then it follows that the EDM matrix D generated by p^i, $i = 1, \ldots, n$ is given by

$$D = \operatorname{diag} Be^T + e\left(\operatorname{diag} B\right)^T - 2B. \tag{3.8}$$

- Let $\bar{r} = n - 1 - r$ and

$$Z \text{ be } n \times \bar{r}, \text{ full column rank with, } P^T Z = 0, e^T Z = 0$$

- ${z^i}^T$ denotes the i-th row of Z. i.e.,

$$Z^T := \left[z^1 \mid z^2 \mid \ldots \mid z^n \right].$$

- z^i is called Gale transform of p^i; Z is called a *Gale matrix* corresponding to D.
- The columns of Z represent the *affine dependence relations* among the points $p^1, p^2, \ldots, p^n$, i.e., among the rows of P.

Main Results

Theorem 3.6.6.1 *Let D be a Euclidean distance matrix and let Z be a Gale matrix corresponding to D. Then, the columns of DZ are proportional to e.*

Equivalently

Theorem 3.6.6.2 *Let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be coefficients, not all zero, of the affine dependence equation of the points $p^1, p^2, \ldots, p^n$, in $\mathbb{R}^r$, i.e.,*

$$\sum_{j=1}^{n} \lambda_j p^j = 0, \quad \sum_{j=1}^{n} \lambda_j = 0.$$

Then

$$\sum_{j=1}^{n} \lambda_j \|p^i - p^j\|^2 = \alpha \text{ for all } i = 1, \ldots, n,$$

for some scalar α independent of i.

Theorem 3.6.6.3 *Let D be a Euclidean distance matrix and let Z be a Gale matrix corresponding to D. Then* **the following are equivalent***:*

1.

$$D = \lambda(E - C), \tag{3.9}$$

for some nonnegative scalar λ and some correlation matrix C;

2.

$$DZ = 0. \tag{3.10}$$

Proof of the Main Results

Lemma 3.6.6.1 *Let D be a Euclidean distance matrix and let $B = -\frac{1}{2}JDJ$. Then:*

1.
$$-\frac{1}{2}V^T DV = V^T BV;$$

2.
$$\mathcal{N}(V^T DV) = \mathcal{N}(P^T V).$$

Proof. The first part follows directly from (3.8) and the definition of V. This yields the second part since $B = PP^T$ and $\mathcal{N}(V^T BV) = \mathcal{N}(V^T PP^T V) = \mathcal{N}(P^T V)$. ■

Lemma 3.6.6.2 *Let D be a Euclidean distance matrix and let U be the matrix whose columns form an orthonormal basis of the null space of V^TDV. Then VU is a Gale matrix corresponding to D.*

Proof. It follows from Lemma 3.6.6.1 that $P^TVU = V^TDVU = 0$ and from the definition of V in (3.7) that $e^TVU = 0$. Hence, the columns of VU form an orthonormal basis for the null space of $\begin{bmatrix} P^T \\ e^T \end{bmatrix}$. ■

Proof of Theorem 3.6.6.1. Let Z be a Gale matrix corresponding to D. Then It follows from Lemma 3.6.6.2 that $VU = ZQ$ for some nonsingular $\bar{r} \times \bar{r}$ matrix Q. Thus $V^T DZ = V^T DVUQ^{-1} = 0$. Hence, the columns of DZ are proportional to e. ■

Proof of Theorem 3.6.6.3. $D = \lambda(E - C)$ for some nonnegative scalar λ and some correlation matrix C if and only if $E - \frac{1}{\lambda}D$ is positive semidefinite. Let $Q = [\frac{e}{\sqrt{n}} \; V]$. Then, $E - D/\lambda \succeq 0$ if and only if $Q^T \, (E - D/\lambda) \, Q \succeq 0$. But

$$Q^T \, (E - D/\lambda) \, Q = \begin{bmatrix} n - \frac{1}{\lambda \, n} \, e^T De & -\frac{1}{\lambda \sqrt{n}} \, e^T DV \\ -\frac{1}{\lambda \sqrt{n}} \, V^T De & -\frac{1}{\lambda} \, V^T DV \end{bmatrix}.$$

Recall that $V^T(-D)V \succeq 0$ follows from Lemma 3.6.6.1. Let W and U be the matrices whose columns form an orthonormal basis for the range space and null space of $V^T(-D)V$, respectively.

Hence, $V^T(-D)V = W\Lambda W^T$, where Λ is the diagonal matrix of the positive eigenvalues of $V^T(-D)V$. Let $Q' = \begin{bmatrix} 1 & 0 & 0 \\ 0 & W & U \end{bmatrix}$. Then, $E - D/\lambda$ is positive semidefinite if and only if

$$\begin{aligned} R &= Q'^T Q^T (E - D/\lambda) Q Q' \\ &= \begin{bmatrix} n - \frac{1}{\lambda n} e^T D e & -\frac{1}{\lambda\sqrt{n}} e^T D V W & -\frac{1}{\lambda\sqrt{n}} e^T D V U \\ -\frac{1}{\lambda\sqrt{n}} W^T V^T D e & \frac{1}{\lambda}\Lambda & 0 \\ -\frac{1}{\lambda\sqrt{n}} U^T V^T D e & 0 & 0 \end{bmatrix} \succeq 0. \end{aligned} \tag{3.11}$$

Now for sufficiently large λ the submatrix

$$\begin{bmatrix} n - \frac{1}{\lambda\, n}\, e^T De & -\frac{1}{\lambda\, \sqrt{n}}\, e^T DVW \\ -\frac{1}{\lambda\, \sqrt{n}}\, W^T V^T De & \frac{1}{\lambda}\, \Lambda \end{bmatrix}$$

is positive definite.

Thus $E - D/\lambda$ is positive semidefinite if and only if $e^T DVU = e^T DZ = 0$. But it follows from Theorem 3.6.6.1 that $e^T DZ = 0$ if and only if $DZ = 0$ and the result follows. ∎

Given the two Euclidean distance matrices:

$$D_1 = \begin{bmatrix} 0 & 1 & 4 \\ 1 & 0 & 1 \\ 4 & 1 & 0 \end{bmatrix}, \qquad D_2 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix},$$

Gale matrices Z_1, Z_2 corresponding to D_1, D_2, resp. are:

$$Z_1^T = [1 \ -2 \ 1], \quad Z_2^T = [1 \ 0 \ -1],$$

Now $D_1 Z_1 = 2e$ and $D_2 Z_2 = 0$. Then: $D_2 = E - C_2$, where

$$C_2 = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \succeq 0.$$

However, there exists no $\lambda \geq 0$ such that $D_1 = \lambda(E - C_1)$ for some correlation matrix C_1.

A Related Papers

A.1 Outline

Following are summaries of several papers closely related to the above notes.

1. Alizadeh has notes that get updated each time he gives the course [1].
2. Nemirovski has course notes that he gave at Delft [9].
3. Todd, December, 2000 has a survey paper on SDP [81].

References

[1] F. ALIZADEH. Lecture notes on semidefinite programming. Technical report, Rutgers University, Piscataway, NJ, 2000. URL: http://karush.rutgers.edu/~alizadeh/.

[2] F. ALIZADEH, J-P.A. HAEBERLY, and M.L. OVERTON. A new primal-dual interior-point method for semidefinite programming. In J.G. Lewis, editor, *Proceedings of the Fifth SIAM Conference on Applied Linear Algebra*, pages 113–117. SIAM, 1994.

[3] M.F. ANJOS and H. WOLKOWICZ. A strengthened SDP relaxation via a second lifting for the Max-Cut problem. *Special Issue of Discrete Applied Mathematics devoted to Foundations of Heuristics in Combinatorial Optimization*, to appear.

[4] K.M. ANSTREICHER and N.W. BRIXIUS. Solving quadratic assignment problems using convex quadratic programming relaxations. Technical report, University of Iowa, Iowa City, IA, 2000.

[5] K.M. ANSTREICHER, X. CHEN, H. WOLKOWICZ, and Y. YUAN. Strong duality for a trust-region type relaxation of the quadratic assignment problem. *Linear Algebra Appl.*, 301(1-3):121–136, 1999.

[6] K.M. ANSTREICHER and H. WOLKOWICZ. On Lagrangian relaxation of quadratic matrix constraints. *SIAM J. Matrix Anal. Appl.*, 22(1):41–55, 2000.

[7] R. BELLMAN. *Introduction to Matrix Analysis.* McGraw Hill, New York, 1957.

[8] R. BELLMAN and K. FAN. On systems of linear inequalities in Hermitian matrix variables. In *Proceedings of Symposia in*

Pure Mathematics, Vol 7, AMS, 1963.

[9] A. BEN-TAL and A.S. NEMIROVSKI. Convex Optimization in Engineering: Modeling, Analysis, Algorithms. Technical Report on notes from course given by A. Nemirovski, Delft University of Technology, Delft, The Netherlands, 1998.

[10] F. BOHNENBLUST. Joint positiveness of matrices. Manuscript, 1948.

[11] 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(4):390–411, 1980/81.

[12] S. BOYD, V. BALAKRISHNAN, E. FERON, and L. El GHAOUI. Control system analysis and synthesis via linear matrix inequalities. *Proc. ACC*, pages 2147–2154, 1993.

[13] S. BOYD and C. H. BARRATT. *Linear Controller Design.*

Limits of Performance. Prentice-Hall, 1991.

[14] J. CULLUM, W.E. DONATH, and P. WOLFE. The minimization of certain nondifferentiable sums of eigenvalues of symmetric matrices. *Mathematical Programming Study*, 3:35–55, 1975.

[15] Jr. DENNIS, J. E. and ROBERT B. SCHNABEL. *Numerical methods for unconstrained optimization and nonlinear equations*, volume 16 of *Classics in Applied Mathematics.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1996. Corrected reprint of the 1983 original.

[16] M.M. DEZA and M. LAURENT. *Geometry of cuts and metrics.* Springer-Verlag, Berlin, 1997.

[17] L.C.W. DIXON, S.E. HERSOM, and Z.A. MAANY. Initial experience obtained solving the low thrust satellite trajectory optimisation problem. Technical Report T.R. 152, The

Hatfield Polytechnic Numerical Optimization Centre, 1984.

[18] H. DYM and I. GOHBERG. Extensions of band matrices with band inverses. *Linear Algebra Appl.*, 36:1–24, 1981.

[19] A.V. FIACCO and G.P. McCORMICK. *Nonlinear programming*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, second (first 1968) edition, 1990. Sequential unconstrained minimization techniques.

[20] A. FISCHER. A Newton-type method for positive-semidefinite linear complementarity problems. *J. Optim. Theory Appl.*, 86(3):585–608, 1995.

[21] T. FUJIE and M. KOJIMA. Semidefinite programming relaxation for nonconvex quadratic programs. *J. Global Optim.*, 10(4):367–380, 1997.

[22] D.M. GAY. Computing optimal locally constrained steps. *SIAM J. Sci. Statist. Comput.*, 2:186–197, 1981.

[23] Y. GENIN, Y. HACHEZ, Y.E. NESTEROV, and P. VAN DOOREN. Optimization problems over positive pseudo-polynomial matrices. Technical report, CORE, Universite Catholique de Louvain, Belgium, 2000.

[24] M.X. GOEMANS. Semidefinite programming in combinatorial optimization. *Math. Programming*, 79:143–162, 1997.

[25] M.X. GOEMANS. Semidefinite programming and combinatorial optimization. *Documenta Mathematica*, Extra Volume ICM 1998:657–666, 1998. Invited talk at the International Congress of Mathematicians, Berlin, 1998.

[26] M.X. GOEMANS and D.P. WILLIAMSON. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. Assoc. Comput. Mach.*, 42(6):1115–1145, 1995.

[27] B. GRONE, C.R. JOHNSON, E. MARQUES de SA, and

H. WOLKOWICZ. Positive definite completions of partial Hermitian matrices. *Linear Algebra Appl.*, 58:109–124, 1984.

[28] Ming Gu. Primal-dual interior-point methods for semidefinite programming in finite precision. *SIAM J. Optim.*, 10(2):462–502 (electronic), 2000.

[29] O. GÜLER and L. TUNÇEL. Characterization of the barrier parameter of homogeneous convex cones. *Math. Programming*, 81:55–76, 1998.

[30] R. HAUSER. Self-scaled barrier functions: Decomposition and classification. Technical Report DAMTP 1999/NA13, Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, England, 1998.

[31] C. HELMBERG, F. RENDL, R. J. VANDERBEI, and H. WOLKOWICZ. An interior-point method for semidefinite programming. *SIAM J. Optim.*, 6(2):342–361, 1996.

[32] R.B. HOLMES. *Geometric Functional Analysis and its Applications.* Springer-Verlag, Berlin, 1975.

[33] R.A. HORN and C.R. JOHNSON. *Matrix analysis.* Cambridge University Press, Cambridge, 1990. Corrected reprint of the 1985 original.

[34] R.A. HORN and C.R. JOHNSON. *Topics in matrix analysis.* Cambridge University Press, Cambridge, 1994. Corrected reprint of the 1991 original.

[35] J. JAHN. *Mathematical Vector Optimization in Partially Ordered Linear Spaces.* Peter Lang, Frankfurt am Main, 1986.

[36] G. JAMESON. *Ordered linear spaces.* Springer-Verlag, New York, 1970.

[37] B. JANSEN, C. ROOS, and T. TERLAKY. A family of polynomial affine scaling algorithms for positive semidefinite linear complementarity problems. *SIAM J. Optim.*,

7(1):126–140, 1997.

[38] C.R. JOHNSON. Matrix completion problems: a survey. *Proceedings of Symposium in Applied Mathematics*, 40:171–198, 1990.

[39] M. KOJIMA, M. SHIDA, and S. SHINDOH. Reduction of monotone linear complementarity problems over cones to linear programs over cones. Technical report, Dept. of Information Sciences, Tokyo Institute of Technology, Tokyo, Japan, 1995.

[40] M. KOJIMA, S. SHINDOH, and S. HARA. Interior-point methods for the monotone semidefinite linear complementarity problem in symmetric matrices. *SIAM J. Optim.*, 7(1):86–125, 1997.

[41] M. LAURENT. A tour d'horizon on positive semidefinite and Euclidean distance matrix completion problems. In *Topics in Semidefinite and Interior-Point Methods*, volume 18 of *The*

Fields Institute for Research in Mathematical Sciences, Communications Series, pages 51–76, Providence, Rhode Island, 1998. American Mathematical Society.

[42] M. LAURENT and S. POLJAK. On a positive semidefinite relaxation of the cut polytope. *Linear Algebra Appl.*, 223/224:439–461, 1995.

[43] A.S. LEWIS. *Take-home final exam, Course CO663 in Convex Analysis.* University of Waterloo, Ontario, Canada, 1994.

[44] A.S. LEWIS. Derivatives of spectral functions. *Math. Oper. Res.*, 21(3):576–588, 1996.

[45] L. LOVÁSZ. On the Shannon capacity of a graph. *IEEE Transactions on Information Theory*, 25:1–7, 1979.

[46] D.G. LUENBERGER. *Optimization by Vector Space Methods.* John Wiley, 1969.

[47] Z.A. MAANY. A new algorithm for highly curved constrained optimization. Technical Report T.R. 161, The Hatfield Polytechnic Numerical Optimization Centre, 1985.

[48] A.W. MARSHALL and I. OLKIN. *Inequalities: Theory of Majorization and its Applications*. Academic Press, 1979.

[49] R.D.C. MONTEIRO. Primal-dual path-following algorithms for semidefinite programming. *SIAM J. Optim.*, 7(3):663–678, 1997.

[50] R.D.C. MONTEIRO and T. TSUCHIYA. Polynomiality of primal-dual algorithms for semidefinite linear complementarity problems based on the Kojima-Shindoh-Hara family of directions. *Math. Programming*, 84:39–53, 1999.

[51] R.D.C. MONTEIRO and Y. ZHANG. A unified analysis for a class of long-step primal-dual path-following interior-point algorithms for semidefinite programming. *Math. Programming*,

81(3, Ser. A):281–299, 1998.

[52] J.J. MORÉ and D.C. SORENSEN. Computing a trust region step. *SIAM J. Sci. Statist. Comput.*, 4:553–572, 1983.

[53] Y.E. NESTEROV. Structure of non-negative polynomial and optimization problems. Technical report, Louvan-la-Neuve, Belgium, 1997. Preprint DP 9749.

[54] Y.E. NESTEROV and A.S. NEMIROVSKI. A general approach to polynomial-time algorithms design for convex programming. Technical report, Centr. Econ. & Math. Inst., USSR Acad. Sci., Moscow, USSR, 1988.

[55] Y.E. NESTEROV and A.S. NEMIROVSKI. Polynomial barrier methods in convex programming. *Èkonom. i Mat. Metody*, 24(6):1084–1091, 1988.

[56] Y.E. NESTEROV and A.S. NEMIROVSKI. Self–concordant functions and polynomial–time methods in convex

programming. Book–Preprint, Central Economic and Mathematical Institute, USSR Academy of Science, Moscow, USSR, 1989. Published in Nesterov and Nemirovsky [60].

[57] Y.E. NESTEROV and A.S. NEMIROVSKI. *Optimization over positive semidefinite matrices: Mathematical background and user's manual.* USSR Acad. Sci. Centr. Econ. & Math. Inst., 32 Krasikova St., Moscow 117418 USSR, 1990.

[58] Y.E. NESTEROV and A.S. NEMIROVSKI. Self-concordant functions and polynomial time methods in convex programming. Technical report, Centr. Econ. & Math. Inst., USSR Acad. Sci., Moscow, USSR, April 1990.

[59] Y.E. NESTEROV and A.S. NEMIROVSKI. Conic formulation of a convex programming problem and duality. Technical report, Centr. Econ. & Math. Inst., USSR Academy of Sciences, Moscow USSR, 1991.

[60] Y.E. NESTEROV and A.S. NEMIROVSKI. *Interior Point Polynomial Algorithms in Convex Programming.* SIAM Publications. SIAM, Philadelphia, USA, 1994.

[61] Y.E. NESTEROV and M.J. TODD. Self-scaled barriers and interior-point methods for convex programming. *Math. Oper. Res.*, 22(1):1–42, 1997.

[62] L.W. NEUSTADT. *Optimization.* Princeton University Press, Princeton, N. J., 1976. A Theory of Necessary Conditions, With a chapter by H. T. Banks.

[63] J.M. ORTEGA and W.C. RHEINBOLDT. *Iterative Solution of Nonlinear Equations in Several Variables.* Academic Press, New York, NY, 1970.

[64] M.L. OVERTON. On minimizing the maximum eigenvalue of a symmetric matrix. *SIAM J. Matrix Anal. Appl.*, 9:256–268, 1988.

[65] M.L. OVERTON. Large-scale optimization of eigenvalues. *SIAM J. Optim.*, 2:88–120, 1992.

[66] G. PATAKI. On the rank of extreme matrices in semidefinite programs and the multiplicity of optimal eigenvalues. *Math. Oper. Res.*, 23(2):339–358, 1998.

[67] G. PATAKI and L. TUNÇEL. On the generic properties of convex optimization problems in conic form. *Math. Programming*, to appear.

[68] A.L. PERESSINI. *Ordered Topological Vector Spaces*. Harper & Row Publishers, New York, 1967.

[69] S. POLJAK, F. RENDL, and H. WOLKOWICZ. A recipe for semidefinite relaxation for $(0,1)$-quadratic programming. *J. Global Optim.*, 7(1):51–73, 1995.

[70] F. RENDL. Semidefinite programming and combinatorial optimization. *Appl. Numer. Math.*, 29:255–281, 1999.

[71] F. RENDL and H. WOLKOWICZ. A semidefinite framework for trust region subproblems with applications to large scale minimization. *Math. Programming*, 77(2, Ser. B):273–299, 1997.

[72] J. RENEGAR. On the computational complexity and geometry of the first-order theory of the reals. III. Quantifier elimination. *J. Symbolic Comput.*, 13(3):329–352, 1992.

[73] J. RENEGAR. Is it possible to know a problem instance is ill-posed? Some foundations for a general theory of condition numbers. *J. Complexity*, 10(1):1–56, 1994.

[74] J. RENEGAR. Incorporating condition measures into the complexity theory of linear programming. *SIAM J. Optim.*, 5(3):506–524, 1995.

[75] J. RENEGAR. Linear programming, complexity theory and elementary functional analysis. *Math. Programming*, 70(3, Ser.

A):279–351, 1995.

[76] J. RENEGAR. Condition numbers, the barrier method, and the conjugate-gradient method. *SIAM J. Optim.*, 6(4):879–912, 1996.

[77] S. SCHAIBLE and W.T. ZIEMBA, editors. *Generalized Concavity in Optimization and Economics*, New York, 1981. Academic Press Inc. [Harcourt Brace Jovanovich Publishers].

[78] N.Z. SHOR. Quadratic optimization problems. *Izv. Akad. Nauk SSSR Tekhn. Kibernet.*, 222(1):128–139, 222, 1987.

[79] R. STERN and H. WOLKOWICZ. Trust region problems and nonsymmetric eigenvalue perturbations. *SIAM J. Matrix Anal. Appl.*, 15(3):755–778, 1994.

[80] R. STERN and H. WOLKOWICZ. Indefinite trust region subproblems and nonsymmetric eigenvalue perturbations. *SIAM J. Optim.*, 5(2):286–313, 1995.

[81] M.J. TODD. Semidefinite programming. Technical report, School of OR and IE, Cornell University, Ithaca, NY, 2000.

[82] L. TUNÇEL and S. XU. On homogeneous convex cones, Caratheodory number, and duality mapping semidefinite liftings. Technical Report CORR 99-21, Department of Combinatorics and Optimization, Waterloo, Ont, 1999.

[83] L. VANDENBERGHE and S. BOYD. Positive definite programming. In *Mathematical Programming: State of the Art, 1994*, pages 276–308. The University of Michigan, 1994.

[84] L. VANDENBERGHE and S. BOYD. Semidefinite programming. *SIAM Rev.*, 38(1):49–95, 1996.

[85] H. WOLKOWICZ. Some applications of optimization in matrix theory. *Linear Algebra Appl.*, 40:101–118, 1981.

[86] H. WOLKOWICZ, R. SAIGAL, and L. VANDENBERGHE, editors. *HANDBOOK OF SEMIDEFINITE*

PROGRAMMING: Theory, Algorithms, and Applications. Kluwer Academic Publishers, Boston, MA, 2000. xxvi+654 pages.

[87] Y. Yuan. On a subproblem of trust region algorithms for constrained optimization. *Math. Programming*, 47(1 (Ser. A)):53–63, 1990.

[88] Y. YUAN. A dual algorithm for minimizing a quadratic function with two quadratic constraints. *Journal of Computational Mathematics*, 9:348–359, 1991.

[89] Q. ZHAO, S.E. KARISCH, F. RENDL, and H. WOLKOWICZ. Semidefinite programming relaxations for the quadratic assignment problem. *J. Comb. Optim.*, 2(1):71–109, 1998. Semidefinite programming and interior-point approaches for combinatorial optimization problems (Toronto, ON, 1996).