Algorithms for Cone Programming (Part I)

Levent Tunçel

Dept. of Combinatorics and Optimization,

Faculty of Mathematics,

University of Waterloo,

Canada.

ltuncel@math.uwaterloo.ca

May 12, 2004

1 Introduction

This part of the course is almost completely focused on

Interior-Point Methods

and a bit on

The Ellipsoid Method

In the modern interior-point theory, all *difficult* constraints are pushed into the *convex set constraints* and/or *convex cone constraints*. Then, each of these *convex inclusion constraints* are treated via a strictly convex *barrier function* with very special properties.

We will present most of our results in the full generality of an arbitrary convex set G in $\mathbb{R}^d$ or an arbitrary convex cone K in $\mathbb{R}^d$. However, special attention will be paid to Semidefinite Programming (SDP).

Recall,

- $\mathcal{S}^n_+$ denotes the convex cone of $n \times n$ symmetric, positive semidefinite matrices over the reals.
- $\mathcal{S}^n_{++}$ is the interior of $\mathcal{S}^n_+$; i.e., the convex cone of $n \times n$ symmetric, positive definite matrices over the reals.

- Second-order cone:

$$SOC^n := \left\{ \begin{pmatrix} x_0 \\ x \end{pmatrix} \in \mathbb{R}^{n+1} : x_0 \geq \|x\|_2 \right\}.$$

SDP stands for *semidefinite programming* where K is made up from direct sums of various $\mathcal{S}_+^{n_i}$ (possibly under some linear isomorphisms).

SOCP stands for *second-order cone programming* where K is made up from direct sums of various SOC^{n_i} (possibly under some linear isomorphisms).

We consider the semidefinite programming (SDP) problems in the following primal (P) and dual (D) forms.

$$
\begin{array}{llrcl}
(P) & \text{inf} & \langle C, X \rangle & & \\
 & & \mathcal{A}(X) & = & b, \\
 & & X & \succeq & 0,
\end{array}
$$

$$
\begin{array}{llrcl}
(D) & \text{sup} & b^T y & & \\
 & & \mathcal{A}^*(y) + Z & = & C, \\
 & & Z & \succeq & 0,
\end{array}
$$

where $\mathcal{A}$ is a linear operator from $\mathcal{S}^n$ to $\mathbb{R}^m$, so that $b \in \mathbb{R}^m$ and $\mathcal{A}^*$ denotes the adjoint of $\mathcal{A}$.

Without loss of generality, we assume that $\mathcal{A}$ is surjective. If not, we can consider the representation $\langle A_i, X\rangle = b_i, i \in \{1, \ldots, m\}$, where $A_i \in \mathcal{S}^n$ for every i. $\mathcal{A}$ being surjective is equivalent to $A_1, A_2, \ldots, A_m$ being linearly independent. The latter can be assumed without loss of generality, since if they are linearly dependent, then either the system $\langle A_i, X\rangle = b_i, i \in \{1, \ldots, m\}$ has no solution, or there are some redundant equations which can be eliminated. In the first case, (P) is infeasible. In the second case, all redundant equations, and corresponding A_i, b_i can be eliminated, to arrive at an equivalent problem satisfying the assumption.

Under this assumption, for any solution, (y, Z), of the equation $\sum_{i=1}^{m} y_i A_i + Z = C$, the Z part of the solution uniquely identifies the corresponding y. Sometimes, in interior-point algorithms, it is convenient to refer only to Z when one mentions a feasible solution of (D). The above setting of the primal-dual SDP pair can be embedded in the following more general setting of conic convex optimization problems:

$$
\begin{array}{lll}
(CP) & \inf & \langle c, x \rangle \\
 & & \begin{array}{rcl} \mathcal{A}(x) & = & b, \\ x & \in & K, \end{array}
\end{array}
$$

where $\mathcal{A}$ is a surjective linear map and K is a pointed, closed, convex cone with non-empty interior.

We define the dual of (CP) as

$$\begin{array}{llrcrcl} (CD) & \text{sup} & \langle b, y\rangle_D & & & & \\ & & \mathcal{A}^*(y) & + & z & = & c, \\ & & & & z & \in & K^+, \end{array}$$

where K^+ is the dual of cone K with respect to $\langle \cdot, \cdot \rangle$, i.e.

$$K^+ \ := \ \left\{ z \in \mathbb{R}^d : \ \langle z, x \rangle \ \geq \ 0, \ \forall\, x \in K \right\}.$$

We will refer to this setting as the conic convex optimization setting.

SDP problem fits into this general setting by letting

- $\mathbb{R}^d := \mathcal{S}^n$ (that is, $d := n(n+1)/2$),
- $K := \mathcal{S}^n_+$,
- $\langle x, z \rangle := \langle X, Z \rangle$.

Under these definitions we have $K^+ = K$. I.e., the cone of symmetric positive semidefinite matrices is *self-dual* under the trace inner product. In addition to being self-dual, the cone $\mathcal{S}^n_+$ enjoys another symmetry property, in that it is *homogeneous*. That is, the set of nonsingular linear transformations keeping $\mathcal{S}^n_+$ the same (the *automorphism group* of $\mathcal{S}^n_+$) is rich enough to contain linear transformations which map any fixed interior point to any other fixed interior point of $\mathcal{S}^n_+$. Convex cones with both properties, i.e. homogeneous self-dual cones, are also called *symmetric*.

- K is self-dual if there exists an inner-product under which $K^* = K$.
- K is homogeneous if Aut(K) acts transitively on int(K).
- K is symmetric if K is homogeneous and self-dual.

So, we have the underlying optimization problems:

- Symmetric Cone Programming (SymCP)
- Homogeneous Cone Programming (HomCP)

- A homogeneous polynomial $p : \mathbb{R}^d \mapsto \mathbb{R}$ is hyperbolic in the direction $h \in \mathbb{R}^d$, if the univariate polynomial (in $t \in \mathbb{R}$)

$$p(x + th)$$

has only real roots for every $x \in \mathbb{R}^d$.

- A convex cone K is hyperbolic if it is

$$\left\{x \in \mathbb{R}^d : p(x + th) \neq 0, \ \forall t \in \mathbb{R}_+\right\}$$

for a polynomial p which is hyperbolic in the direction $h \in \mathbb{R}^d$.

- Homogeneous cones make up a proper subset of hyperbolic cones.

Hyperbolic Cone Programming (HypCP)

Strictly speaking we have,

$$LP \subset SOCP \subset SDP \subset SymCP \subset HomCP \subset HypCP \subset CP.$$

However, in some sense,

$$LP \subset SOCP \subseteq SDP = SymCP = HomCP \subseteq HypCP \subset CP.$$

Yet in an another sense,

$$LP = SOCP \subseteq SDP = SymCP = HomCP \subseteq HypCP \subseteq CP.$$

Recall the *weak duality relation*:

Proposition 1.1 *Let $\bar{X}$ be feasible in (P), and $(\bar{y}, \bar{Z})$ be feasible in (D). Then*

$$\langle C, \bar{X} \rangle - b^T \bar{y} \;=\; \langle \bar{X}, \bar{Z} \rangle \;\geq\; 0.$$

Similarly for the conic convex optimization setting... Therefore, if we start with $X^{(0)}$ and $Z^{(0)}$ both feasible in their respective problems, then decreasing $\langle X, Z \rangle$ will get us closer to optimality!

Since the linear operator $\mathcal{A}$ is surjective, we can always find $\bar{X} \in \mathcal{S}^n$ such that

$$\mathcal{A}(\bar{X}) = b.$$

For the dual, because of the form we chose, we can always find $\bar{y} \in \mathbb{R}^m$, $\bar{Z} \in \mathcal{S}^n$ such that

$$\mathcal{A}^*(\bar{y}) + \bar{Z} = C.$$

Denoting

$$\mathcal{L} \ := \ \{d \in \mathcal{S}^n : \ \mathcal{A}(d) = 0\},$$

we claim that (P) and (D) are *equivalent* to the following pair.

$$
\begin{array}{lll}
(\tilde{P}) & \text{inf} & \langle \bar{Z}, X \rangle \\
 & & X \in (\mathcal{L} + \bar{X}), \\
 & & X \succeq 0,
\end{array}
$$

$$
\begin{array}{lll}
(\tilde{D}) & \text{inf} & \langle \bar{X}, Z \rangle \\
 & & Z \in (\mathcal{L}^{\perp} + \bar{Z}), \\
 & & Z \succeq 0,
\end{array}
$$

here, $\mathcal{L}^{\perp}$ denotes the orthogonal complement of $\mathcal{L}$.

We have

$$(C\tilde{P}) \qquad \begin{array}{rcl} \inf & \langle \bar{z}, x \rangle & \\ x & \in & (\mathcal{L} + \bar{x}) \cap K, \end{array}$$

and

$$(C\tilde{D}) \qquad \begin{array}{rcl} \inf & \langle z, \bar{x} \rangle & \\ z & \in & (\mathcal{L}^{\perp} + \bar{z}) \cap K^{+}. \end{array}$$

in the general conic convex optimization setting. To establish the equivalence, first note that the feasible regions are preserved (in $(C\tilde{D})$ we only refer to z).

Recall the proof of the weak duality relation (Proposition 1.1). For every $x \in \mathbb{R}^d$ satisfying $\mathcal{A}(x) = b$, and for every (y, z) satisfying $\mathcal{A}^*(y) + z = c$ we have

$$\langle c, x \rangle - \langle b, y \rangle_D \; = \; \langle x, z \rangle.$$

If we fix $(\bar{y}, \bar{z})$ such that

$$\mathcal{A}^*(\bar{y}) + \bar{z} = c$$

then for all $x \in \mathbb{R}^d$ satisfying $\mathcal{A}(x) = b$ we have

$$\langle c, x \rangle \; = \; \langle x, \bar{z} \rangle + \mathit{constant},$$

where the constant is $\langle b, \bar{y} \rangle_D$. Therefore, minimizing $\langle c, x \rangle$ subject to any set of constraints, containing the constraint $\mathcal{A}(x) = b$ is equivalent to minimizing $\langle x, \bar{z} \rangle$ subject to the same set of constraints.

Similarly, we can establish the equivalence of the dual problems. We fix $\bar{x}$ such that $\mathcal{A}(\bar{x}) = b$ then for all (y, z) satisfying $\mathcal{A}^*(y) + z = c$, we have

$$-\langle b, y \rangle_D \;=\; \langle \bar{x}, z \rangle + \color{blue}{constant},$$

where the constant is $-\langle c, \bar{x} \rangle$. Therefore, maximizing $\langle b, y \rangle_D$ subject to any set of constraints containing the constraint $\mathcal{A}^*(y) + z = c$ is equivalent to minimizing $\langle \bar{x}, z \rangle$ subject to the same set of constraints.

2 Ellipsoid Method

$E \subset \mathbb{R}^d$ is an *ellipsoid* if there exist $c \in \mathbb{R}^d$ (determining the center) and $A \in \mathcal{S}^d_{++}$ (determining the size and the shape) such that

$$E := E(A,c) := \left\{ x \in \mathbb{R}^d : (x-c)^T A^{-1} (x-c) \le 1 \right\}.$$

We can alternatively express the ellipsoid as the image of the unit ball in $\mathbb{R}^d$ (denoted by $B_d(0,1)$) under an affine mapping as follows:

$$E(A,c) = A^{1/2} B_d(0,1) + c.$$

The volume of the ellipsoid is proportional to the square-root of the determinant of the positive definite matrix determining its shape:

$$\text{vol}(E(A,c)) = \sqrt{\det(A)}\,\text{vol}(B_d(0,1)).$$

The volume of the d-dimensional unit ball is

$$\text{vol}(B_d(0,1)) = \frac{\pi^{d/2}}{\Gamma\left(\frac{d}{2}+1\right)},$$

where $\Gamma(x) := \int_0^\infty \mathbf{e}^{-t} t^{x-1} dt$, for $x > 0$.

We first study the ellipsoid method as an algorithm which computes a point in an implicitly described convex set. In this basic setting, it is easy to see that *the ellipsoid method is a beautiful and theoretically very powerful generalization of the bisection method from $\mathbb{R}$ to $\mathbb{R}^d$, for an arbitrary d*.

2.1 Ingredients: Separation Oracles, Inscribed and Circumscribed Ellipsoids

To appreciate the full theoretical power of the *Ellipsoid Method*, we will move away from the *explicit* description of the convex optimization problem at hand. Instead, we will assume that we are given access to a *separation oracle*. Let $G \subset \mathbb{R}^d$ be the convex set we are interested in optimizing over or less ambitiously just finding a point inside it (the set G). We define δ-relaxation of G as follows:

$$\text{relax}(G, \delta) := \left\{ u \in \mathbb{R}^d : \|u - x\|_2 \leq \delta \text{ for some } x \in G \right\}.$$

A *weak separation oracle for* G takes as input $\bar{x} \in \mathbb{Q}^d$, $\delta \in \mathbb{Q}_{++}$. It either outputs "$\bar{x} \in \mathrm{relax}(G, \delta)$" or $a \in \mathbb{Q}^d$ such that $\|a\|_\infty = 1$ and

$$\langle a, \bar{x} \rangle \geq \langle a, x \rangle - \delta, \quad \forall x \in \mathrm{relax}(G, \delta).$$

Theorem 2.1 *For every compact, convex set in $\mathbb{R}^d$ with nonempty interior, there exists a unique minimal volume ellipsoid containing that set. Moreover, shrinking that ellipsoid (around its center) by a factor of at most d gives an ellipsoid contained in the convex set.*

The unique ellipsoid described in the above theorem is usually called the *Löwner-John ellipsoid*. The factor d in the above theorem is the best possible. (The d-dimensional simplex proves the claim for every d.)

Let

$$E := \left\{ x \in \mathbb{R}^d : (x-c)^T A^{-1} (x-c) \leq 1 \right\},$$

and

$$\tilde{E} := \left\{ x \in E : a^T x \leq a^T c \right\},$$

for some $a \in \mathbb{R}^d \setminus \{0\}$. We will assume $d \geq 2$. We would like to construct the smallest volume ellipsoid E_+ containing the half-ellipsoid $\tilde{E}$. Let $c_+ \in \mathbb{R}^d$ and $A_+ \in \mathcal{S}^d_{++}$ denote the center and the positive definite matrix determining E_+. Then

$$c_+ = c - \frac{1}{(d+1)\sqrt{a^T A a}} A a,$$

$$A_+ = \frac{d^2}{d^2-1} \left[A - \frac{2}{(d+1) a^T A a} A a a^T A \right].$$

We can explicitly compute the volume of E_+ in terms of the volume of E, since A_+ is a rank-1 update of A. When we take the image of $\mathbb{R}^d$ under the mapping $A^{-1/2}$, our ellipsoid E becomes the unit ball. Under this mapping our update formula becomes

$$I_+ = \frac{d^2}{d^2-1}\left[I - \frac{2}{(d+1)\bar{a}^T\bar{a}}\bar{a}\bar{a}^T\right],$$

where $\bar{a} := A^{1/2}a$. The eigenvalues of $\left[I - \frac{2}{(d+1)\bar{a}^T\bar{a}}\bar{a}\bar{a}^T\right]$ are $\frac{d-1}{d+1}$ and 1 (with multiplicity $(d-1)$). Therefore,

$$\det(I_+) = \left(\frac{d}{d+1}\right)^{d+1}\left(\frac{d}{d-1}\right)^{d-1} = \left(\frac{d^2}{d^2-1}\right)^{d-1}\frac{d^2}{(d+1)^2}.$$

Hence the volume of E_+ can be related to the volume of E as follows:

$$\text{vol}(E_+) = \left(\frac{d}{\sqrt{d^2-1}}\right)^{d-1} \frac{d}{d+1}\text{vol}(E).$$

Theorem 2.2 *We have*

$$\hat{E} \subseteq E_+ \textit{ and } \ln\left(\frac{\text{vol}(E_+)}{\text{vol}(E)}\right) \leq -\frac{1}{2d}.$$

2.2 Complexity Analysis for the Ellipsoid Method

Suppose vol$(E_0) = R$ and we want E_k such that vol$(E_k) \leq \epsilon$. Then by the last theorem,

$$\ln\left(\frac{\text{vol}(E_k)}{\text{vol}(E_0)}\right) \leq -\frac{k}{2d},$$

and $O\left(d\ln\left(\frac{R}{\epsilon}\right)\right)$ iterations suffice. Similarly, if R is the radius of the initial ball (whose image under the mapping $x \mapsto A_0^{-1/2}x$ contains the set) and we want the stopping criterion to be that the radius of the current ball (whose image under the mapping $x \mapsto A_k^{-1/2}x$ contains G) is at most ϵ, then using

$$\left(\frac{\text{vol}(E_k)}{\text{vol}(E_0)}\right) = d\ln\left(\frac{R}{\epsilon}\right),$$

we find that $O\left(d^2 \ln\left(\frac{R}{\epsilon}\right)\right)$ iterations suffice.

Theorem 2.3 *Let $G \subset \mathbb{R}^d$ be a convex set such that $G \subseteq B_d(0, R)$. Suppose we are also given $\epsilon \in \mathbb{Q}_{++}$. If we have access to a weak separation oracle for G, then in polynomial time (polynomial in $(d, \ln(R/\epsilon))$), we can compute $\bar{x} \in \text{relax}(G, \epsilon)$ or prove that $\text{vol}(G) \leq \epsilon$.*

These results can be extended to optimizing a convex function over G. We need one more ingredient to deal with the objective function. Let $f : \mathbb{R}^d \mapsto \mathbb{R}$ be a convex function.

Definition 2.1 *A* subgradient oracle for f *takes as input* $\bar{x} \in \mathbb{R}^d$ *and returns in polynomial time (polynomial in* d *and for a proper definition,* size$(\bar{x})$*)* $f(\bar{x})$ *and* $h \in \mathbb{R}^d$ *such that*

$$f(x) \geq f(\bar{x}) + h^T(x - \bar{x}), \quad \forall x \in \mathbb{R}^d.$$

Suppose we are interested in solving the convex optimization problem

$$\inf \{f(x) : x \in G\},$$

where $f : \mathbb{R}^d \mapsto \mathbb{R}$ is a convex function.

Theorem 2.4 *Let $G \subset \mathbb{R}^d$ be a convex set and $r, R \in \mathbb{R}_{++}$ be given such that*

$$B_d(\tilde{x}, r) \subseteq G \subseteq B_d(0, R), \quad \text{for some } \tilde{x} \in \mathbb{R}^d$$

(here, $\tilde{x}$ is **not** *given). Let $\epsilon \in \mathbb{R}_{++}$ be also given. Suppose that a subgradient oracle for f and a weak separation oracle for G are available. Then after*

$$O\left(d^2 \left[\ln\left(\frac{R}{r}\right) + \ln\left(\frac{\mu_0}{\epsilon}\right)\right]\right)$$

iterations, the ellipsoid method returns a feasible solution $\bar{x} \in G$ such that

$$f(\bar{x}) \leq \inf\{f(x) : x \in G\} + \epsilon.$$

In the above, $\mu_0 := \epsilon + \sup_{x \in E_0}\{f(x)\} - \inf_{x \in E_0}\{f(x)\}$.

2.3 Bibliographical Notes

Until 1979 very few mathematicians in the West knew about the ellipsoid method. Khachiyan, in 1979 proved that the Ellipsoid Method can be adapted to solve linear programming problems in polynomial time (hence settling a long outstanding problem). This announcement caused an unprecedented reaction from the media.

The Ellipsoid Method was originally proposed by Iudin and Nemirovski in 1976 also some related work is due to Shor in 1977. This original method was designed to deal with essentially *any* convex optimization problem that can be posed in a finite dimensional space by the potential usage of oracles (an important point which should be emphasized is that the functions involved in defining the feasible solution set, the objective function need **not** be differentiable). For a nice exposure to Ellipsoid Method for linear programming problems, see the survey paper by Bland, Goldfarb and Todd, *Operations Research* 29 (1981) 1039–1091.

Shortly after Khachiyan's result, it was established that this method is a very powerful tool in determining the computational complexity status (hence the degree of difficulty) of various combinatorial optimization problems. A good reference is *Geometric algorithms and combinatorial optimization* by M. Grötschel, L. Lovász and A. Schrijver.

In the late 1980’s, Ellipsoid Method was applied to some problems in System and Control Theory (see the book *Linear Matrix Inequalities in System and Control Theory* by Boyd et al.). These problems were small but relatively difficult convex optimization problems. Since late 1980’s and early 1990’s interior-point methods consistently took over the solution process. These are the methods we discuss and analyze next.

3 Central Path

One of the most important concepts in interior-point methods is *the central path*. We arrive at this concept via another cental concept:

the barrier for the *difficult constraints*.

Let $F : \mathbb{R}^d \mapsto \mathbb{R}$ be a *logarithmically homogeneous self-concordant barrier for* K.

Definition 3.1 *(LHSCB) Let* F : int$(K) \to \mathbb{R}$ *be a* $\mathcal{C}^3$*-smooth convex function such that* F *is a barrier for* K *(i.e.* $F(x) \to \infty$ *as* $x \in$ int(K) *approaches* ∂K*) and there exists* $\vartheta \geq 1$ *such that for each* $t > 0$,

$$F(tx) \;=\; F(x) - \vartheta \ln(t),$$

and

$$D^3 F(x)[h,h,h]| \;\leq\; 2\left(D^2 F(x)[h,h]\right)^{3/2}$$

for all $x \in$ int(K) *and for all* $h \in E$*. Then* F *is called a* ϑ-LHSC barrier *for* K.

$$F_*(z) := -\inf_{x \in \text{int}(K)} \{\langle z, x\rangle + F(x)\}$$ Legendre-Fenchel Conjugate

F_* also has the above mentioned properties for K^+ for the same *barrier parameter* ϑ.

ϑ is a very important parameter of these barriers. Currently, one of the best iteration bounds for interior-point methods for conic convex optimization is

$$O\left(\sqrt{\vartheta}\ln\left(\frac{1}{\epsilon}\right)\right)$$

to compute an ϵ-optimal solution.

Let $\mu > 0$. Consider

$$\begin{array}{rrcl} (CP_\mu) \quad \inf & \langle c, x \rangle & + & \mu F(x) \\ & \mathcal{A}(x) & = & b, \\ & (x & \in & K), \end{array}$$

and

$$\begin{array}{rrcccl} (CD_\mu) \quad \inf & -\langle b, y \rangle_D & + & \mu F_*(z) & & \\ & \mathcal{A}^*(y) & + & z & = & c, \\ & & & (z & \in & K^+). \end{array}$$

It is well-known that

Theorem 3.1 *Suppose* $(x^{(0)}, z^{(0)}) \in \mathrm{int}(K) \oplus \mathrm{int}(K^+)$ *feasible in* (CP_0) *and* (CD_0) *exists. Then* (CP_μ) *and* (CD_μ) *have a unique optimal solution pair* $x(\mu)$, $(y(\mu), z(\mu))$, *for each* $\mu > 0$.

Definition 3.2 $\{(x(\mu), y(\mu), z(\mu)) : \mu > 0\}$ *is called the* primal-dual central path *for the pair* $(CP_0), (CD_0)$.

Sometimes we refer only to $(x(\mu), z(\mu))$.

3.1 Central Path for SDP

Let's focus on SDP first. Here,

$$F(X) := -\ln(\det(X)),$$

$$F_*(Z) = -\ln(\det(Z)) - n = F(Z) - n.$$

The central path is equivalently characterized as the unique solution (via necessary and sufficient conditions for optimality) of the convex optimization problem (P_μ):

$$\begin{aligned} \mathcal{A}(X) = b, \qquad X \succ 0, \\ -\mathcal{A}^*(y) - X^{-1} + \frac{1}{\mu}C = 0. \end{aligned}$$

Let's do the substitutions $y := \mu y$ and $Z := \mu X^{-1}$. We obtain the system

$$\begin{aligned} \mathcal{A}(X) &= b, \quad X \succ 0, \\ \mathcal{A}^*(y) + Z &= C, \\ Z &= \mu X^{-1}. \end{aligned}$$

For each $\mu > 0$, the unique solution of the above system $(X(\mu), y(\mu), Z(\mu))$ defines *the primal-dual central path*.

Path-Following algorithms "closely" or "loosely" follow this path to the set of optimal solutions. (Note that under the assumption of the existence of strictly feasible points for both primal and the dual, both problems do have optimal solutions and there is no duality gap.)

Potential-Reduction algorithms reduce the problem to that of minimizing (or driving to $-\infty$) a combination of the objective function:

- $\langle c, x \rangle$
- $b^T y$
- $\langle X, Z \rangle$

and a barrier (e.g., $-\ln(\det(X))$ or a measure of centrality.

3.2 Neighbourhoods of the Central Path

Again, we first focus on SDP. To follow the central path, or to understand the potential-reduction algorithms in a unifying way, it is worth studying the neighbourhoods of the central path. In the primal-dual setting, this is quite elegant and effective.

Given strictly feasible points X and Z, define

$$\mu := \frac{\langle X, Z \rangle}{n}.$$

Let $\mathcal{F}_0$ denote the set of all strictly feasible solution pairs (X, Z) (X satisfies all the primal constraints and is positive definite, Z satisfies all the dual constraints and is positive definite). Then we can express many neighbourhoods.

3.3 Neighbourhoods Based on the Algebraic Description of the Central Path

Let $\beta \in (0,1)$ be an absolute constant. We define

(i) so-called *wide neighborhoods*

$$\mathcal{N}_\infty^-(\beta) \; := \; \left\{(X,Z) \in \mathcal{F}_0 : \; \lambda_{\min}(X^{1/2} Z X^{1/2}) \geq (1-\beta)\mu \right\},$$

(ii) *infinity-norm neighborhoods*

$$\mathcal{N}_\infty(\beta) \; := \; \left\{(X,Z) \in \mathcal{F}_0 : \; \left\| \lambda(X^{1/2} Z X^{1/2}) - \mu e \right\|_\infty \leq \beta\mu \right\},$$

(ii)$'$ or equivalently,

$$\mathcal{N}_\infty(\beta) \; := \; \left\{(X,Z) \in \mathcal{F}_0 : \; \left\| \frac{X^{1/2} Z X^{1/2}}{\mu} - I \right\|_2 \leq \beta \right\},$$

(iii) so-called *tight (or narrow) neighborhoods*

$$\mathcal{N}_2(\beta) \ := \ \left\{(X,Z) \in \mathcal{F}_0 : \ \left\| \frac{X^{1/2} Z X^{1/2}}{\mu} - I \right\|_F \leq \beta \right\}.$$

Note that (as is well-known),

$$\text{Central Path} \subset \mathcal{N}_2(\beta) \subset \mathcal{N}_\infty(\beta) \subset \mathcal{N}_\infty^-(\beta) \subset \mathcal{F}_0.$$

3.4 Neighbourhoods Based on the Analytic Descriptions of the Central Path

We define a measure of *centrality* based on the barrier values:

$$\psi(X,Z) := n \ln\left(\frac{\langle X, Z\rangle}{n}\right) - \ln(\det(X)) - \ln(\det(Z)).$$

Theorem 3.2 *For every* $(X,Z) \in \mathcal{S}^n_{++} \oplus \mathcal{S}^n_{++}$,

$$\psi(X,Z) \geq 0.$$

Moreover, the equality holds above iff $Z = \mu X^{-1}$ *for some* $\mu > 0$.

This theorem generalizes the *Arithmetic-Geometric Mean Inequality* and the corresponding characterization for equality.

We can also define a proximity measure based on the gradients of the barrier $-\ln(\det(X))$, $-\ln(\det(Z))$: Let

$$\tilde{\mu} := \frac{\langle X^{-1}, Z^{-1} \rangle}{n}.$$

We have

Theorem 3.3 *For every $X \succ 0$, $Z \succ 0$,*

$$\mu\tilde{\mu} \;\geq\; 1.$$

Equality holds above iff $Z = \mu X^{-1}$.

The above theorem generalizes the *Arithmetic-Harmonic Mean Inequality* and the corresponding characterization for equality.

In the general, conic convex optimization setting, the central path equation

$$z = -\mu F'(x)$$

replaces $Z = -\mu X^{-1}$. The first proximity measure is generalized to:

$$\psi(x, z) \; := \; \vartheta \ln \left(\frac{\langle x, z \rangle}{\vartheta} \right) + F(x) + F_*(z) + \vartheta.$$

The next theorem shows that for every pair of interior solutions (x, z), the proximity measure is nonnegative and it is equal to zero if and only if the point (x, z) lies on the central path.

Theorem 3.4 *Let F be a LHSCB for K with parameter ϑ. Then*

$$\psi(x,z) \geq 0, \quad \textit{for all } x \in \mathrm{int}(K),\ z \in \mathrm{int}(K^+).$$

Moreover, the inequality above holds as equality iff

$$z = -tF^{'}(x), \ \textit{for some } t > 0.$$

For convenience, we write $\tilde{x} := -F_*'(z)$ and $\tilde{z} := -F'(x)$. One can think of $\tilde{x}$ and $\tilde{z}$ as the *shadow iterates*, as $\tilde{x} \in \text{int}(K)$ and $\tilde{z} \in \text{int}(K^+)$ and if (x, z) is a feasible pair, then $\mu\tilde{x} = x$ iff $\mu\tilde{z} = z$ iff (x, z) lies on the central path. We also denote $\tilde{\mu} := \langle \tilde{x}, \tilde{z} \rangle / \vartheta$.

Theorem 3.5 *For every* $(x, z) \in \text{int}(K) \oplus \text{int}(K^+)$,

$$\mu\tilde{\mu} \geq 1.$$

Equality holds above iff $x = -\mu F_*'(z)$ *(and hence* $z = -\mu F'(x)$*).*

3.5 Path-Following

While most elegant potential-reduction algorithms might not make any reference to the central path (even including the theoretical analysis), path-following and potential-reduction algorithms are very closely connected.

Currently, most of the practical implementations of IPMs both for LP and SymCP are based on modifications of path-following algorithms.

There are many strategies available to us for following the central path.

Our algorithms generate search directions

$$D_X \text{ and } D_Z$$

and step sizes

$$\alpha_X \text{ and } \alpha_Z$$

and update

$$X \leftarrow X + \alpha_X D_X,$$

and

$$Z \leftarrow Z + \alpha_Z D_Z.$$

In theory, it is very convenient to take

$$\alpha := \alpha_X = \alpha_Z.$$

However, there are possible advantages in practise to allow them to take different values.

What properties do we ask for in the search directions?

- Improve the current duality gap $\langle X, Z \rangle$
- Get closer to the central path without increasing $\langle X, Z \rangle$ (very much)
- A suitable combination of the first two above!

We can also mix these properties “externally” in a predictor-corrector methods.

Let's define

$$X(\alpha) := X + \alpha D_X,$$

and

$$Z(\alpha) := Z + \alpha D_Z.$$

Given $\gamma \in [0, 1]$ (a centrality parameter defining some of the properties of the search direction), there are many search directions achieving

$$\langle X(\alpha), Z(\alpha) \rangle \leq [1 - \alpha(1 - \gamma)] \langle X, Z \rangle,$$

and for α large enough (e.g., $\Omega(1)$, $\Omega(1/n)$, $\Omega(1/\sqrt{n})$, ...), $(X(\alpha), Z(\alpha))$ stays in a suitable neighbourhood of the central path.

More on search directions for SDP at the end of this lecture and at the beginning of the next...

4 Primal-Dual Potential Function

Given $(\hat{X}, \hat{Z})$ and $(\bar{X}, \bar{Z})$ a pair of primal-dual feasible and interior pairs, we would like to have a simple way of comparing them. We have two criteria:

- smaller the duality gap $\langle X, Z \rangle$ is the better,
- smaller the distance to the central path (that is, the value of $\psi(X, Z)$) is the better.

The next function, called *the primal-dual potential function*, serves such a purpose and allows us to design and perform the complexity analysis directly on it.

$$\phi_q(X, Z) := q \ln(\langle X, Z \rangle) + \psi(X, Z),$$

where $q > 0$ (we will take $q := \sqrt{n}$ in our analysis).

Theorem 4.1 *Suppose we have* $\left(X^{(0)}, Z^{(0)}\right)$ *feasible for* (P) *and* (D) *such that*

$$\psi(X^{(0)}, Z^{(0)}) \leq \sqrt{n} \ln\left(\frac{1}{\epsilon}\right), \text{ for some } \epsilon \in (0,1).$$

If we generate $(X^{(k)}, Z^{(k)})$ *feasible in* (P) *and* (D) *such that*

$$\phi_{\sqrt{n}}(X^{(k)}, Z^{(k)}) \leq \phi_{\sqrt{n}}(X^{(k-1)}, Z^{(k-1)}) - \delta, \text{ for every } k \geq 1,$$

for some $\delta > 0$ *an absolute constant, then for some* $\bar{k} = O(\sqrt{n}\ln(1/\epsilon))$, *we have*

$$\langle X^{(k)}, Z^{(k)} \rangle \leq \epsilon \langle X^{(0)}, Z^{(0)} \rangle, \text{ for every } k \geq \bar{k}.$$

5 Algorithm and Computational Complexity Analysis

Based on the last theorem, our problem of producing an approximately optimal solution pair is reduced to decreasing the potential function value by a constant, in every iteration. We want to update $(X^{(k)}, Z^{(k)})$ to $(X^{(k+1)}, Z^{(k+1)})$ such that $\mathcal{A}(X) = b$, $\mathcal{A}^*(y) + Z = C$, $X \succ 0$, $Z \succ 0$ are all maintained; moreover, $\langle X, Z \rangle$ is decreased and $\psi(X, Z)$ is not increased a lot (in comparison to the duality gap). We can express the update from $(X^{(k)}, Z^{(k)})$ to $(X^{(k+1)}, Z^{(k+1)})$ by a pair of *search directions* D_X, $D_Z \in \mathcal{S}^n$ respectively and a step size $\alpha \in \mathbb{R}_{++}$. We drop the iterate numbers for this part of our study and define

$$X(\alpha) := X + \alpha D_X, \text{ and } Z(\alpha) := Z + \alpha D_Z.$$

To maintain the feasibility of the iterates, the search directions must satisfy

$$\mathcal{A}(D_X) = 0 \text{ and } \mathcal{A}^*(d_y) + D_Z = 0,$$

for some $d_y \in \mathbb{R}^m$; i.e., D_X must be in the null space of $\mathcal{A}(\cdot)$ and D_Z must be in the range of $\mathcal{A}^*(\cdot)$.

We will derive an algorithm that is symmetric between the primal and the dual (*primal-dual symmetry*), we would also like to have our algorithm invariant under the symmetries of the cone constrains (*scale-invariance*). To be more specific, instead of formalizing these vague goals, we will describe an approach which attains these goals.

For every $(X, Z) \in \mathcal{S}^n_{++} \oplus \mathcal{S}^n_{++}$ we would like to have a self-adjoint, positive definite linear transformation $T : \mathcal{S}^n \mapsto \mathcal{S}^n$ such that

- $T \in \text{Aut}(\mathcal{S}^n_+)$,
- $T(Z) = T^{-1}(X) =: V$,
- $T(X^{-1}) = T^{-1}(Z^{-1}) = V^{-1}$.

If we can find such transformation T, then we can map our primal-space with the mapping T^{-1} and the dual-space with T. This modification does not change anything significantly, except that our current primal-dual iterate is mapped onto (V, V). Let's elaborate:

Since T (and therefore T^{-1}; because, T is self-adjoint) is an automorphism of the cone of positive semidefinite matrices, the most important part of the problem (for the current interior-point method approach) is unchanged. We define

$$\begin{aligned} \bar{\mathcal{A}}(\cdot) &:= \mathcal{A}(T(\cdot)) \\ \bar{C} &:= T(C) \\ \bar{D}_X &:= T^{-1}(D_X), \\ \bar{D}_Z &:= T(D_Z). \end{aligned}$$

Now, (P) and (D) become

$$\begin{array}{llrcl}
(\bar{P}) & \text{inf} & \langle \bar{C}, X \rangle & & \\
& & \bar{\mathcal{A}}(X) & = & b, \\
& & X & \in & T^{-1}(\mathcal{S}^n_+) = \mathcal{S}^n_+,
\end{array}$$

$$\begin{array}{llrcl}
(\bar{D}) & \text{sup} & b^T y & & \\
& & \bar{\mathcal{A}}^*(y) + Z & = & \bar{C}, \\
& & Z & \in & T(\mathcal{S}^n_+) = \mathcal{S}^n_+.
\end{array}$$

In these *scaled spaces*, the search directions are still orthogonal: $\bar{D}_X$ must lie in the null space of $\bar{\mathcal{A}}(\cdot)$ and $\bar{D}_Z$ must lie in the range of $\bar{\mathcal{A}}^*(\cdot)$.

Let's analyze the duality gap. We have

$$\langle X(\alpha), Z(\alpha)\rangle \;=\; \langle X, Z\rangle + \alpha\langle V, (\bar{D}_Z + \bar{D}_X)\rangle.$$

Therefore, if we take as $\bar{D}_X$ and $\bar{D}_Z$ the orthogonal projection of $-V$ onto the null space of $\bar{\mathcal{A}}(\cdot)$ and range of $\bar{\mathcal{A}}^*(\cdot)$ respectively, then we will have the best search direction to reduce the duality gap in this setting.

Now, let's turn to the *centrality part* of the potential function. What kind of search direction would improve the barrier function values in this setting? We utilize the following technical lemma which summarizes many of the nice properties of the barrier function F.

Lemma 5.1 *Let* $X \in \mathcal{S}^n_{++}$. *Suppose* $D \in \mathcal{S}^n$ *satisfies*

$$\|D\|_X := \langle D, X^{-1} D X^{-1} \rangle^{1/2} \leq 1.$$

Then

$$F(X) + \langle F^{'}(X), D \rangle \leq F(X + D),$$

$$F(X + D) \leq F(X) + \langle F^{'}(X), D \rangle + \frac{\|D\|_X^2}{2(1 - \|D\|_X)^2}.$$

Remark 5.1 *The condition $\|D\|_X \leq 1$ (< 1) implies $(X + D) \succeq 0$ ($\succ 0$). This is clear from the statement of the lemma. But it can also be directly observed as follows:*

$$1 \geq \|D\|_X = \|X^{-1/2} D X^{-1/2}\|_F.$$

Therefore, $\|X^{-1/2} D X^{-1/2}\|_2 \leq 1$. But this is equivalent to $I \pm X^{-1/2} D X^{-1/2} \succeq 0$. If we apply the automorphism $X^{1/2} \cdot X^{1/2}$ of $\mathcal{S}^n_+$ to both sides, we obtain equivalently $X \pm D \succeq 0$.

Let's focus on the first order upper estimate on $F(X(\alpha)) + F(Z(\alpha))$ given by the above lemma. We obtain

$$\left[\langle F^{'}(X), D_X\rangle + \langle F^{'}(Z), D_Z\rangle\right] \;=\; -\langle V^{-1}, \bar{D}_X + \bar{D}_Z\rangle.$$

Thus, as in our analysis of the duality gap, if we take as $\bar{D}_X$ and $\bar{D}_S$ the orthogonal projection of V^{-1} onto the null space of $\bar{\mathcal{A}}(\cdot)$ and range of $\bar{\mathcal{A}}^*(\cdot)$ respectively, then we will have the best search direction to reduce the first order term in the upper bound on the value of the barrier terms $F(X) + F(Z)$, in this setting.

Therefore, to reduce the value of the potential function, it seems desirable to choose a matrix which is a nonnegative linear combination of $-V$ and V^{-1} and then define $\bar{D}_X$ and $\bar{D}_Z$ as the orthogonal projections of this matrix onto the null space of $\bar{\mathcal{A}}(\cdot)$ and the range of $\bar{\mathcal{A}}^*(\cdot)$ respectively.

This is precisely what we do next. Let

$$\tilde{U} := -\frac{(n+\sqrt{n})}{\langle X, Z\rangle} V + V^{-1}.$$

Note that $\|\tilde{U}\|_F = 0$ iff $V^{-1} = \frac{(n+\sqrt{n})}{\langle X,Z\rangle} V$. But the latter leads to a contradiction (that $n = n + \sqrt{n}$) upon taking the inner product with V of both sides. Therefore, $\|\tilde{U}\|_F > 0$ and we define

$$U := \frac{\tilde{U}}{\|\tilde{U}\|_F}.$$

In fact, $\|\tilde{U}\|_F$ is connected to a measure of centrality. Recall

$$\tilde{\mu} := \frac{\langle X^{-1}, Z^{-1}\rangle}{n} = \frac{\langle V^{-1}, V^{-1}\rangle}{n}.$$

Then

$$\begin{aligned} \|U\|_F^2 &= \frac{(n+\sqrt{n})^2}{(n\mu)^2} n\mu - 2\frac{(n+\sqrt{n})}{\mu} + n\tilde{\mu} \\ &= \frac{1}{\mu}\left[n(\mu\tilde{\mu}-1)+1\right]. \end{aligned}$$

Note that $(\mu\tilde{\mu}-1)$ is $\frac{1}{n\mu}$ times the squared norm of the error in the equation $Z = \mu X^{-1}$, where the norm is with respect to the local metric induced by T^2. We have

Corollary 5.1 *For every $X, Z \in \mathcal{S}^n_{++}$, we have*

$$\|\tilde{U}\|_F^2 \geq \frac{1}{\mu}(>0).$$

The equality holds above iff $Z = \mu X^{-1}$.

Then $\bar{D}_X$, d_y, $\bar{D}_Z$ make up the unique solution of the system:

$$\begin{aligned} \bar{\mathcal{A}}(\bar{D}_X) &= 0 \\ \bar{\mathcal{A}}^*(d_y) + \bar{D}_Z &= 0 \\ \bar{D}_X + \bar{D}_Z &= U. \end{aligned}$$

By definition,

$$\|\bar{D}_X\|_F^2 + \|\bar{D}_Z\|_F^2 = \|U\|_F^2 = 1.$$

Therefore, we immediately conclude that

$$\|\bar{D}_X\|_F \le 1 \text{ and } \|\bar{D}_Z\|_F \le 1.$$

Now, we analyze $\|D_X\|_X$ and $\|D_Z\|_Z$. We have

$$\|D_X\|_X^2 = \langle D_X, X^{-1} D_X X^{-1} \rangle$$

$$
\begin{aligned}
&= \langle W\bar{D}_X W, X^{-1}W\bar{D}_X W X^{-1}\rangle \\
&= \langle \bar{D}_X, (WX^{-1}W)\bar{D}_X(WX^{-1}W)\rangle \\
&\leq \left\|V^{-1}\cdot V^{-1}\right\|_2 \|\bar{D}_X\|_F^2 \\
&\leq \left\|V^{-1}\cdot V^{-1}\right\|_2 \\
&\leq \frac{1}{(\lambda_n(V))^2}.
\end{aligned}
$$

Note that in the above derivation, we encountered the linear operator

$$T(F^{''}(X))T(\cdot)$$

which happens to coincide (in this case) with the linear operator

$$\left(T(F^{'}(X))\cdot T(F^{'}(X))\right).$$

Similarly,

$$\begin{aligned} \|D_Z\|_Z^2 &\leq \left\|V^{-1} \cdot V^{-1}\right\|_2 \|\bar{D}_Z\|_F^2 \\ &\leq \left\|V^{-1} \cdot V^{-1}\right\|_2 \\ &\leq \frac{1}{(\lambda_n(V))^2}. \end{aligned}$$

We have,

$$
\begin{aligned}
\phi(\alpha)-\phi(0) &\leq (n+\sqrt{n})\ln\left(\frac{\langle X(\alpha),Z(\alpha)\rangle}{\langle X,Z\rangle}\right)-\alpha\langle U,V^{-1}\rangle \\
&\quad +\alpha^2\left(\frac{\|D_X\|_X^2}{2(1-\alpha\|D_X\|_X)^2}+\frac{\|D_Z\|_Z^2}{2(1-\alpha\|D_Z\|_Z)^2}\right) \\
&\leq \alpha\frac{(n+\sqrt{n})}{\langle X,Z\rangle}\langle U,V\rangle-\alpha\langle U,V^{-1}\rangle \\
&\quad +2\alpha^2\left(\frac{(1/\lambda_n(V))^2}{2(1-\alpha(1/\lambda_n(V)))^2}\right) \\
&= -\alpha\|\tilde{U}\|_F+\alpha^2\left(\frac{(1/\lambda_n(V))^2}{(1-\alpha(1/\lambda_n(V)))^2}\right) \\
&\leq -\frac{\sqrt{3}}{2\lambda_n(V)}\alpha+\alpha^2\left(\frac{(1/\lambda_n(V))^2}{(1-\alpha(1/\lambda_n(V)))^2}\right)
\end{aligned}
$$

$$
\begin{aligned}
&= -\frac{\sqrt{3}}{16} + \frac{1}{64}\frac{1}{(7/8)^2} \\
&= -\frac{\sqrt{3}}{16} + \frac{1}{49} \\
&< -\frac{1}{12}.
\end{aligned}
$$

ALGORITHM

Given $(X^{(0)}, Z^{(0)})$ feasible in (P) and (D) such that $X^{(0)} \succ 0$, $Z^{(0)} \succ 0$. Also given is $\epsilon \in (0,1)$ such that $\psi(X^{(0)}, Z^{(0)}) \leq \sqrt{n}\ln(1/\epsilon)$.

$k := 0$.

While $\langle X^{(k)}, Z^{(k)} \rangle > \epsilon \langle X^{(0)}, Z^{(0)} \rangle$,

$$W^2 := \left(Z^{(k)}\right)^{-1/2} \left[\left(Z^{(k)}\right)^{1/2} X^{(k)} \left(Z^{(k)}\right)^{1/2}\right]^{1/2} \left(Z^{(k)}\right)^{-1/2}$$

$$\bar{\mathcal{A}}(\cdot) := \mathcal{A}(W \cdot W) \left[\bar{A}_i := W A_i W\right]$$

$$V := W Z^{(k)} W$$

$$\tilde{U} := -\frac{n+\sqrt{n}}{\langle X^{(k)}, Z^{(k)} \rangle} V + V^{-1}$$

$$U := \frac{\tilde{U}}{\|\tilde{U}\|_F}$$

Solve the system

$$
\begin{aligned}
\bar{\mathcal{A}}(\bar{D}_X) &= 0 \\
\bar{\mathcal{A}}^*(d_y) + \bar{D}_Z &= 0 \\
\bar{D}_X + \bar{D}_Z &= U.
\end{aligned}
$$

Compute $\bar{\alpha} :=$
$$\min\left\{\phi_{\sqrt{n}}(X^{(k)} + \alpha W \bar{D}_X W, Z^{(k)} + \alpha W^{-1} \bar{D}_Z W^{-1}) : \alpha > 0\right\}.$$

Let $X^{(k+1)} := X^{(k)} + \bar{\alpha} W \bar{D}_X W,$
$Z^{(k+1)} := Z^{(k)} + \bar{\alpha} W^{-1} \bar{D}_Z W^{-1}.$

$k := k + 1.$

end{While}

The above is what is typically called a *potential reduction algorithm*.

We proved the following theorem.

Theorem 5.1 *The above algorithm terminates in at most*

$$24\sqrt{n}\ln(1/\epsilon)$$

iterations with feasible $X^{(k)}, Z^{(k)}$ *such that*

$$\langle X^{(k)}, Z^{(k)}\rangle \leq \epsilon\langle X^{(0)}, Z^{(0)}\rangle.$$

Even though the algorithm requires the iterates to lie in the interior of the underlying cone constraints, we can relax the initial feasibility assumption by using auxiliary optimization problems.

6 Infeasible-Start Algorithms

Another approach is to work in the framework of the algorithm that we discussed (or some other primal-dual interior-point algorithm) but modify the search directions so that the search directions also try eliminate the error in the linear equations defining the primal and dual feasible regions. Instead of having our search directions D_X and D_Z lying in the nullspace of $\mathcal{A}(\cdot)$ and range of $\mathcal{A}^*(\cdot)$, we ask that they satisfy the following system of equations:

$\mathcal{A}(D_X) = b - \mathcal{A}(X^{(k)})$ and $\mathcal{A}^*(d_y) + D_Z = C - \mathcal{A}^*(y^{(k)}) - Z^{(k)}$.

The analysis becomes more complicated, however, this is one of the popular ways to solve SDP problems in practise. The algorithms need to carefully monitor the progress in attaining feasibility, reducing $\langle X, Z \rangle$ as well as the proximity to the "central surface." (Since we allow infeasible iterates, we will be concerned with the distance to the "central surface" rather than the central path.)

For instance, the algorithm should not allow the fast reduction of $\langle X, Z \rangle$ unless the iterates are getting to be near feasible at least as fast.

7 Other Interior-Point Algorithms, General Remarks

The search directions that we discussed are known as the NT direction (for Nesterov-Todd). These algorithms have been generalized to convex optimization problems over arbitrary convex cones.

Other primal-dual algorithms that are useful and popular rely on search directions proposed Helmberg-Rendl-Vanderbei-Wolkowicz/Kojima-Shindoh-Hara/Monteiro (HKM direction) and Alizadeh-Haeberly-Overton (AHO direction). All these directions can be defined and treated in a unified way (due to Y. Zhang, some other related work is due to Monteiro-Y. Zhang):

Let $M \in \mathbb{R}^{n\times n}$. Define $H_P : \mathbb{R}^{n\times n} \mapsto \mathcal{S}^n$ as follows

$$H_P(M) := PMP^{-1} + P^{-T}M^TP^T.$$

This $H_P(\cdot)$ is called the *symmetrized similarity transformation*. To compute the search direction, we solve the system

$$\begin{aligned} \mathcal{A}(D_X) &= b - \mathcal{A}(X^{(k)}), \\ \mathcal{A}^*(d_y) + D_Z &= C - \mathcal{A}^*(y^{(k)}) - Z^{(k)}, \\ H_P\left(X^{(k)}D_Z + D_X Z^{(k)}\right) &= 2\gamma\mu I - H_P\left(X^{(k)}Z^{(k)}\right), \end{aligned}$$

where $\gamma \in [0,1]$ a parameter fixed by the user/algorithm and $\mu = \langle X^{(k)}, Z^{(k)}\rangle / n$ as before.

Choosing $P := I$ gives the AHO direction, $P := (Z^{(k)})^{1/2}$ yields the HKM direction, choosing any $P \in \mathbb{R}^{n \times n}$ such that

$$P^T P = (X^{(k)})^{-1/2} \left((X^{(k)})^{1/2} Z^{(k)} (X^{(k)})^{1/2} \right)^{1/2} (X^{(k)})^{-1/2}$$

(for instance, $P := \left((X^{(k)})^{1/2} Z^{(k)} (X^{(k)})^{1/2} \right)^{1/4} (X^{(k)})^{-1/2}$) gives the NT direction.

The next lecture starts with a discussion of the computational issues related to the search directions for SDP which ties in nicely with the bundle methods.

We can also design a wide range of primal-dual algorithms without the conic structure or logarithmic homogeneity:

Polynomial time IPMs

WITH OR WITHOUT

the Conic Structure and Logarithmically Homogeneous Barriers!

(From a recent paper by Nemirovski and T.)

We are given

- a ϑ-SCB Φ with a domain $\tilde{G}$ and the Legendre-Fenchel conjugate Φ_* (with a slight difference from the previous defn.) of Φ; the domain of Φ_* is denoted $\tilde{G}^+$. $\tilde{G}^+$ is a cone:

$$y \in \tilde{G}^+ \Rightarrow \tau y \in \tilde{G}^+ \quad \forall \tau > 0$$

- a linear embedding $x \mapsto \mathcal{A}(x)$ with the null space $\mathcal{A} = \{0\}$ and the image intersecting $\tilde{G}$;
- a vector $c \neq 0$.

- the optimization problem

$$\inf_x \{\langle c, x\rangle : x \in G\}, \quad G = \{x : \mathcal{A}(x) \in \tilde{G}\},$$

  we are interested in solving;

- the function $F(x) = \Phi(\mathcal{A}(x))$ which is a ϑ-SCB for cl(G).

A shifted central path

Lemma 7.1 *For $t > 0$, the "primal-dual pair" $(x, y) = (x_*(t), y_*(t))$ is uniquely defined by the relations*

$$
\begin{array}{ll}
(a) & y \in \tilde{G}^+, \quad x \in G \\
(b) & \mathcal{A}^*(y) = -tc \\
(c) & \Phi_*'(y) = \mathcal{A}(x) \quad [\Leftrightarrow y = \Phi'(\mathcal{A}(x))].
\end{array}
$$

Moreover,

$$
y_*(t) = \text{argmin}_y \{\Phi_*(y) : \mathcal{A}^*(y) = -tc\}.
$$

7.1 Proximity measure

Let us define the *proximity measure* as the function

$$\Psi(x,y) = \Phi(\mathcal{A}(x)) + \Phi_*(y) - \langle y, \mathcal{A}(x)\rangle : G \oplus \tilde{G}^+ \to \mathbb{R}$$

(Legendre-Fenchel gap between Φ and Φ_*). Notice that for every $x \in G$ and every $y \in \tilde{G}^+$, we have $\Psi(x,y) \geq 0$ and for such a pair (x,y) we have $\Psi(x,y) = 0$ iff $y = \Phi'(\mathcal{A}(x))$.

Using this set up many path-following and potential-reduction algorithms can be derived and analyzed.

References

[1] F. Alizadeh, “Interior point methods in semidefinite programming with applications to combinatorial optimization,” *SIAM Journal on Optimization* **5** (1995) 13–51.

[2] F. Alizadeh, J-P. A. Haeberly, and M. L. Overton, Primal-dual interior-point methods for semidefinite programming: Convergence rates, stability and numerical results, *SIAM J. Optim.*, 8:746–768, 1998.

[3] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization. Analysis, Algorithms, and Engineering Applications*, MPS/SIAM Series on Optimization, Philadelphia, PA, USA, 2001.

[4] C. B. Chua, Relating homogeneous cones and positive definite cones via T-algebras, *SIAM J. Opt.* 14 (2004) 500-506.

[5] J. Faraut and A. Korányi, *Analysis on Symmetric Cones*, Oxford University Press, NY, USA, 1994.

[6] M. Grötschel, L. Lovász and A. Schrijver, *Geometric algorithms and combinatorial optimization* (Springer, New York, 1988).

[7] O. Güler, Barrier functions in interior point methods, *Math. of Oper. Res.* 21 (1996), pp. 860–885.

[8] O. Güler, Hyperbolic Polynomials and Interior Point Methods for Convex Programming, *Math. Oper. Res.* 22 (1997) 350–377.

[9] C. Helmberg, F. Rendl, R. Vanderbei and H. Wolkowicz, An interior-point method for semidefinite programming. *SIAM J. Optim.* **6** (1996) 342–361.

[10] R.A. Horn and C.R. Johnson, *Topics in matrix analysis*, Cambridge University Press, Cambridge, 1994. Corrected reprint of the 1991 original.

[11] Jarre, F. *Interior-Point Methods via Self-Concordance or Relative Lipschitz Condition*, Habilitationsschrift, University of Würzburg, July 1994.

[12] N. Karmarkar, A new polynomial time algorithm for linear programming. *Combinatorica* 4 (1984) 373–395.

[13] M. Kojima, M. Shida and S. Shindoh, Search directions in the SDP and the monotone SDLCP: generalization and inexact computation. *Math. Programming*, 85:51–80, 1999.

[14] M. Kojima, S. Shindoh and S. Hara, Interior-point methods for the monotone linear complementarity problem in symmetric matrices. *SIAM J. Optim.* **7** (1997) 86–125.

[15] R. D. C. Monteiro, First and second order methods for semidefinite programming, *Math. Prog.B* 97 (2003) 209–244.

[16] R. D. C. Monteiro, Primal-dual path following algorithms for semidefinite programming. *SIAM J. Optim.* 7 (1997) 663–678.

[17] R. D. C. Monteiro and T. Tsuchiya, Polynomial convergence of a new family of primal-dual algorithms for semidefinite programming, *SIAM J. Optim.* 9 (1999) 551–577.

[18] A. Nemirovski and L. Tunçel, "Cone-free" primal-dual path-following and potential reduction polynomial time interior-point methods, *Research Report* CORR 2002-32, Dept. of Combinatorics and Optimization, Faculty of Mathematics, University of Waterloo, Canada, October 2002.

[19] Yu. E. Nesterov and A. S. Nemirovskii, *Interior-Point Polynomial Algorithms in Convex Programming*, SIAM, Philadelphia, PA, USA, 1994.

[20] Nesterov, Yu. E. and M. J. Todd (1997). Self-scaled barriers and interior-point methods for convex programming, *Math. Oper. Res.* **22** 1–46.

[21] Yu. E. Nesterov and M. J. Todd, Primal-dual interior-point methods for self-scaled cones, *SIAM J. Optimization* **8** (1998) 324–364.

[22] J. Renegar, *A Mathematical View of Interior-Point Methods in Convex Optimization*, MPS/SIAM Series on Optimization, Philadelphia, PA, USA, 2001.

[23] Rockafellar, R. T. (1970). *Convex Analysis*, Princeton University Press, Princeton, New Jersey.

[24] A. Schrijver, *Theory of Linear and Integer Programming.* John Wiley and Sons, 1986.

[25] A. Shapiro, First and second order analysis of nonlinear semidefinite programs, *Math. Programming B* 77 (1997) 301–320.

[26] M. J. Todd, On search directions in interior-point methods for semidefinite programming, *Optimization Methods and Software* **11** (1999) 1–46.

[27] M. J. Todd, Potential-reduction methods in mathematical programming, *Mathematical Programming* 76 (1997) 3–45.

[28] L. Tunçel, Generalization of primal-dual interior-point methods to convex optimization problems in conic form, *Foundations of Computational Mathematics* **1** (2001) 229–254.

[29] L. Tunçel, Potential reduction and primal-dual methods, *Handbook of Semidefinite Programming: Theory, Algorithms and Applications*, H. Wolkowicz, R. Saigal and L. Vandenberghe (eds.), Kluwer Academic Publishers, Boston, MA, USA, 2000, pp. 235–265.

[30] L. Tunçel and H. Wolkowicz, Strengthened existence and uniqueness conditions for search directions in semidefinite programming, *Research Report* CORR 2003-20, Dept. of Combinatorics and Optimization, Faculty of Mathematics, University of Waterloo, Ontario, Canada, July 2003.

[31] L. Vandenberghe and S. Boyd, A primal-dual potential reduction method for problems involving matrix inequalities, *Mathematical Programming* 69 (1995) 205–236.

[32] L. Vandenberghe and S. Boyd, “Semidefinite Programming,” *SIAM Review* **38** (1996) 49–95.

[33] H. Wolkowicz, Some applications of optimization in matrix theory. *Linear Algebra and its Applications*, 40:101–118, 1981.

[34] 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.

[35] Y. Ye, Approximating quadratic programming with bound and quadratic constraints, *Math. Programming* 84 (1999) 219–226.

[36] Y. Zhang, On extending some primal-dual interior-point algorithms from linear programming to semidefinite programming. *SIAM J. Optim.*, 8:365–386, 1998.