Second Order Optimization Algorithms II: Interior-Point Algorithms

Yinyu Ye
Department of Management Science and Engineering
Stanford University
Stanford, CA 94305, U.S.A.

http://www.stanford.edu/~yyye

Chapter 5.4-7, 6.6

Linear Programming Methodological Philosophy

Optimality Conditions: (1) Primal Feasibility, (2) Dual Feasibility, (3) Zero-Duality Gap/Prima-Dual Complementarity.

Recall that the (primal) Simplex Algorithm maintains the primal feasibility and complementarity while working toward dual feasibility. (The Dual Simplex Algorithm maintains dual feasibility and complementarity while working toward primal feasibility.)

In contrast, interior-point methods will move in the interior of the feasible region, hoping to by-pass many corner points on the boundary of the region. The primal-dual interior-point method maintains both primal and dual feasibility while working toward complementarity.

The key for the simplex method is to make computer see corner points; and the key for interior-point methods is to stay in the interior of the feasible region.

Interior-Point Algorithms for LP

$$(LP)\ \min\ \mathbf{c}^T\mathbf{x}\ \text{s.t.}\ A\mathbf{x}=\mathbf{b},\ \mathbf{x}\geq\mathbf{0}\quad <=>\quad (LD)\ \max\ \mathbf{b}^T\mathbf{y}\ \text{s.t.}A^T\mathbf{y}+\mathbf{s}=\mathbf{c},\ \mathbf{s}\geq\mathbf{0}.$$

$$\text{int}\,\mathcal{F}_p=\{\mathbf{x}:\ A\mathbf{x}=\mathbf{b},\ \mathbf{x}>\mathbf{0}\}\neq\emptyset$$

$$\text{int}\,\mathcal{F}_d=\{(\mathbf{y},\mathbf{s}):\ \mathbf{s}=\mathbf{c}-A^T\mathbf{y}>\mathbf{0}\}\neq\emptyset.$$

Let z^* denote the optimal value and

$$\mathcal{F}=\mathcal{F}_p\times\mathcal{F}_d.$$

We are interested in finding an ϵ-approximate solution for the LP problem:

$$\mathbf{x}^T\mathbf{s}=\mathbf{c}^T\mathbf{x}-\mathbf{b}^T\mathbf{y}\leq\epsilon.$$

For simplicity, we assume that an interior-point pair $(\mathbf{x}^0,\mathbf{y}^0,\mathbf{s}^0)$ is known, and we will use it as our initial point pair.

Barrier Functions and Analytic Center

Consider the barrier function optimization problems:

$$(PB) \quad \text{minimize} \quad -\sum_{j=1}^{n} \log x_j \quad \text{s.t.} \quad \mathbf{x} \in \text{int}\,\mathcal{F}_p$$

and

$$(DB) \quad \text{maximize} \quad \sum_{j=1}^{n} \log s_j \quad \text{s.t.} \quad (\mathbf{y}, \mathbf{s}) \in \text{int}\,\mathcal{F}_d$$

The maximizer $\mathbf{x}$ (or $(\mathbf{y}, \mathbf{s})$) of (PB) (or (BD)) is called the analytic center of bounded polyhedron $\mathcal{F}_p$ (or $\mathcal{F}_d$). Applying the KKT conditions and using $X = \text{diag}(\mathbf{x})$, we have

$$-X^{-1}\mathbf{e} - A^T\mathbf{y} = \mathbf{0} \quad \text{or} \quad -\mathbf{e} - XA^T\mathbf{y} = \mathbf{0},\ A\mathbf{x} = \mathbf{b},\ \mathbf{x} > \mathbf{0}.$$

After introducing auxiliary vector $\mathbf{s} = X^{-1}\mathbf{e}$, the conditions become

$$\begin{array}{rcl} X\mathbf{s} & = & \mathbf{e} \\ A\mathbf{x} & = & \mathbf{b} \\ -A^T\mathbf{y} - \mathbf{s} & = & \mathbf{0} \\ \mathbf{x} & > & \mathbf{0}. \end{array} \qquad \left(\text{or} \quad \begin{array}{rcl} S\mathbf{x} & = & \mathbf{e} \\ A\mathbf{x} & = & \mathbf{0} \\ -A^T\mathbf{y} - \mathbf{s} & = & -\mathbf{c} \\ \mathbf{s} & > & \mathbf{0}. \end{array} \right)$$

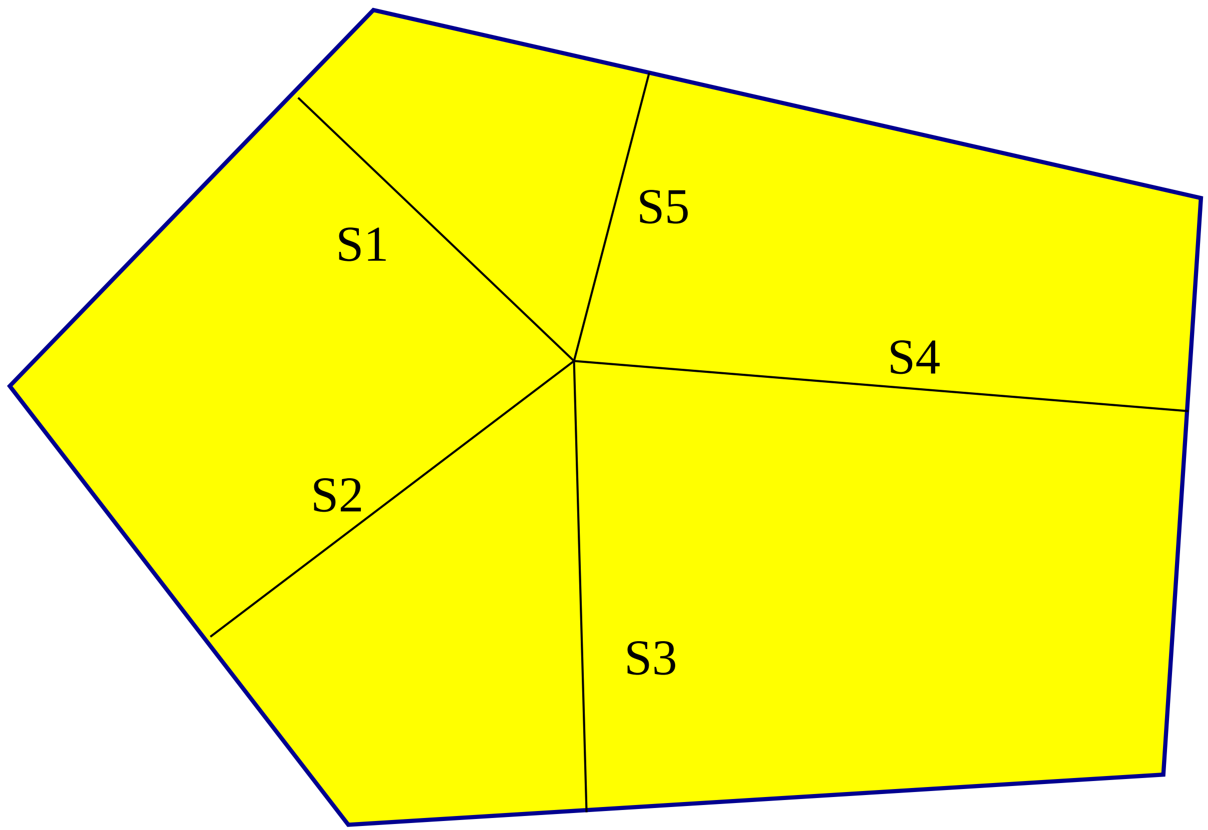

Figure 1: The dual analytic center maximizes the product of slacks.

Examples

$$\mathcal{F}_p = \{\mathbf{x} : \sum_j \mathbf{x}_j = 1, \ \mathbf{x} \geq \mathbf{0}\}.$$

The analytic center of $\mathcal{F}_p$ would be

$$\mathbf{x}^c = (\frac{1}{n}; \ ...; \ \frac{1}{n}), \ y = -n, \ \mathbf{s} = (n; \ ...; \ n).$$

$$\mathcal{F}_d = \{\mathbf{y} : \mathbf{0} \leq \mathbf{y} \leq \mathbf{e}\}.$$

The analytic center of $\mathcal{F}_d$ would be

$$\mathbf{y}^c = \arg\max \sum_i (\log(y_i) + \log(1 - y_i)) = \arg\max \sum_i \log(y_i(1 - y_i))$$

that is

$$\mathbf{y}^c = (\frac{1}{2}; \ ...; \ \frac{1}{2}), \ \mathbf{s} = \frac{1}{2}\mathbf{e}, \ \mathbf{x} = 2\mathbf{e}.$$

Why “analytic”: depending on the analytical representation data.

Logarithmic Function and Scaled Concordant Lipschitz

Lemma 1 *Let* $B(\mathbf{x}) = -\sum_{j=1}^{n} \log(x_j)$*. Then, for any point* $\mathbf{x} > \mathbf{0}$ *and direction vector* $\mathbf{d}$ *such that* $\|X^{-1}\mathbf{d}\|_\infty \leq \alpha(< 1)$,

$$-\mathbf{e}^T X^{-1}\mathbf{d} \leq B(\mathbf{x}+\mathbf{d}) - B(\mathbf{x}) \leq -\mathbf{e}^T X^{-1}\mathbf{d} + \frac{\|X^{-1}\mathbf{d}\|^2}{2(1-\alpha)}.$$

The Barrier function property can be generalized to the so-called Second-Order Scaled Concordant Lipschitz Condition: for any $\mathbf{x} > \mathbf{0}$ and $\mathbf{x} + \mathbf{d}$ in the function domain:

$$\|X\left(\nabla f(\mathbf{x}+\mathbf{d}) - \nabla f(\mathbf{x}) - \nabla^2 f(\mathbf{x})\mathbf{d}\right)\| \leq \beta_\alpha \mathbf{d}^T \nabla^2 f(\mathbf{x})\mathbf{d}, \text{ whenever } \|X^{-1}\mathbf{d}\| \leq \alpha(< 1).$$

Such condition can be verified using Taylor Expansion Series; basically, the scaled third derivative of the function is bounded by its (unscaled) second derivative.

- All quadratic functions are scaled concordant Lipschitz with $\beta_\alpha = 0$.
- Convex function $-\log(x)$ is scaled concordant Lipschitz with $\beta_\alpha = \frac{1}{(1-\alpha)}$.
- All power functions $\{x^p : x > 0\}$ with integer p are scaled concordant Lipschitz with $\beta_\alpha = \frac{O(p)}{(1-\alpha)}$.

Affine-Scaling Gradient Projection

To compute the analytic center, we consider the affine-scaling GPM from any feasible $\mathbf{x} > \mathbf{0}$:

$$\begin{array}{ll} \text{minimize} & -\mathbf{e}^T X^{-1}\mathbf{d} \\ \text{s.t.} & A\mathbf{d} = \mathbf{0},\ \|X^{-1}\mathbf{d}\| \le \alpha \end{array} \quad \text{or} \quad \begin{array}{ll} \text{minimize} & -\mathbf{e}^T \mathbf{d}' \\ \text{s.t.} & AX\mathbf{d}' = \mathbf{0},\ \|\mathbf{d}'\| \le \alpha \end{array}$$

which has a close-form solution

$$\mathbf{d}' = \alpha(I - XA^T(AX^2A^T)^{-1}AX)\mathbf{e}/\|(I - XA^T(AX^2A^T)^{-1}AX)\mathbf{e}\|.$$

Note that $\mathbf{d} = X\mathbf{d}'$ so that we let $\mathbf{x}^+ = \mathbf{x} + \mathbf{d}$, which should remain positive:

$$\mathbf{x}^+ = \mathbf{x} + \mathbf{d} = \mathbf{x} + X\mathbf{d}' = X(\mathbf{e} + \mathbf{d}') > \mathbf{0}$$

as long as $\mathbf{x} > \mathbf{0}$ and $\|\mathbf{d}'\| < 1$. Then, from Lemma 1 the Barrier function value would be decreased at least by

$$B(\mathbf{x}^+) - B(\mathbf{x}) \le -\alpha\|(I - XA^T(AX^2A^T)^{-1}AX)\mathbf{e}\| + \frac{\alpha^2}{2(1-\alpha)}.$$

Convergence Speed Analysis

For simplicity, let $\mathbf{y}(\mathbf{x}) = (AX^2A^T)^{-1}AX\mathbf{e}$ and $\mathbf{s}(\mathbf{x}) = A^T\mathbf{y}(\mathbf{s})$ so that

$$(I - XA^T(AX^2A^T)^{-1}AX)\mathbf{e} = \mathbf{e} - X\mathbf{s}(\mathbf{x}).$$

Note that $\mathbf{y}(\mathbf{x})$ minimizes $\min_{\mathbf{y}} \ \|\mathbf{e} - XA^T\mathbf{y}\|^2$.

Thus, as long as $\|\mathbf{e} - X\mathbf{s}(\mathbf{x})\| \geq 1$, the Barrier function can be decreased by a universal constant $-\alpha + \frac{\alpha^2}{2(1-\alpha)} = -3/4$ when we set $\alpha = 1/2$.

If the quantity $\|\mathbf{e} - X\mathbf{s}(\mathbf{x})\| < 1$, then we simply let $\mathbf{x}^+ = \mathbf{x} + X(\mathbf{e} - X\mathbf{s}(\mathbf{x}))$, in which case we now prove $\|\mathbf{e} - X^+\mathbf{s}(\mathbf{x}^+)\| \leq \|\mathbf{e} - X\mathbf{s}(\mathbf{x})\|^2$ (quadratic convergence)!

$$\begin{aligned}
\|\mathbf{e} - X^+\mathbf{s}(\mathbf{x}^+)\|^2 &\leq \|\mathbf{e} - X^+\mathbf{s}(\mathbf{x})\|^2, \quad \text{(because } \mathbf{y}(\mathbf{x}^+) \text{ minimizes the squares)} \\
&= \|\mathbf{e} - (2X - X^2S(\mathbf{x})\mathbf{s}(\mathbf{x})\|^2 \\
&= \sum_{j=1}^n \left(1 - 2x_js_j(\mathbf{x}) + x_j^2(s_j(\mathbf{x}))^2\right)^2 \\
&= \sum_{j=1}^n (1 - x_js_j(\mathbf{x}))^4 \\
&\leq \left(\sum_{j=1}^n (1 - x_js_j(\mathbf{x}))^2\right)^2 = \|\mathbf{e} - X\mathbf{s}(\mathbf{x})\|^4.
\end{aligned}$$

Analytic Volume and Cutting Plane for LP: Geometric Interpretation

$$AV(\mathcal{F}_d) := \prod_{j=1}^{n} \bar{s}_j = \prod_{j=1}^{n} (c_j - \mathbf{a}_j^T \bar{\mathbf{y}})$$

can be viewed as the analytic volume of polytope $\mathcal{F}_d$ or simply $\mathcal{F}$ in the rest of discussions.

If one inequality in $\mathcal{F}$, say the first one, needs to be translated, change $\mathbf{a}_1^T \mathbf{y} \le c_1$ to $\mathbf{a}_1^T y \le \mathbf{a}_1^T \bar{\mathbf{y}}$; i.e., the first inequality is parallelly moved and it now cuts through $\bar{\mathbf{y}}$ and divides $\mathcal{F}$ into two bodies. Analytically, c_1 is replaced by $\mathbf{a}_1^T \bar{\mathbf{y}}$ and the rest of data are unchanged. Let

$$\mathcal{F}^+ := \{\mathbf{y} : \ \mathbf{a}_j^T \mathbf{y} \le c_j^+, \ j = 1, ..., n\},$$

where $c_j^+ = c_j$ for $j = 2, ..., n$ and $c_1^+ = \mathbf{a}_1^T \bar{\mathbf{y}}$.

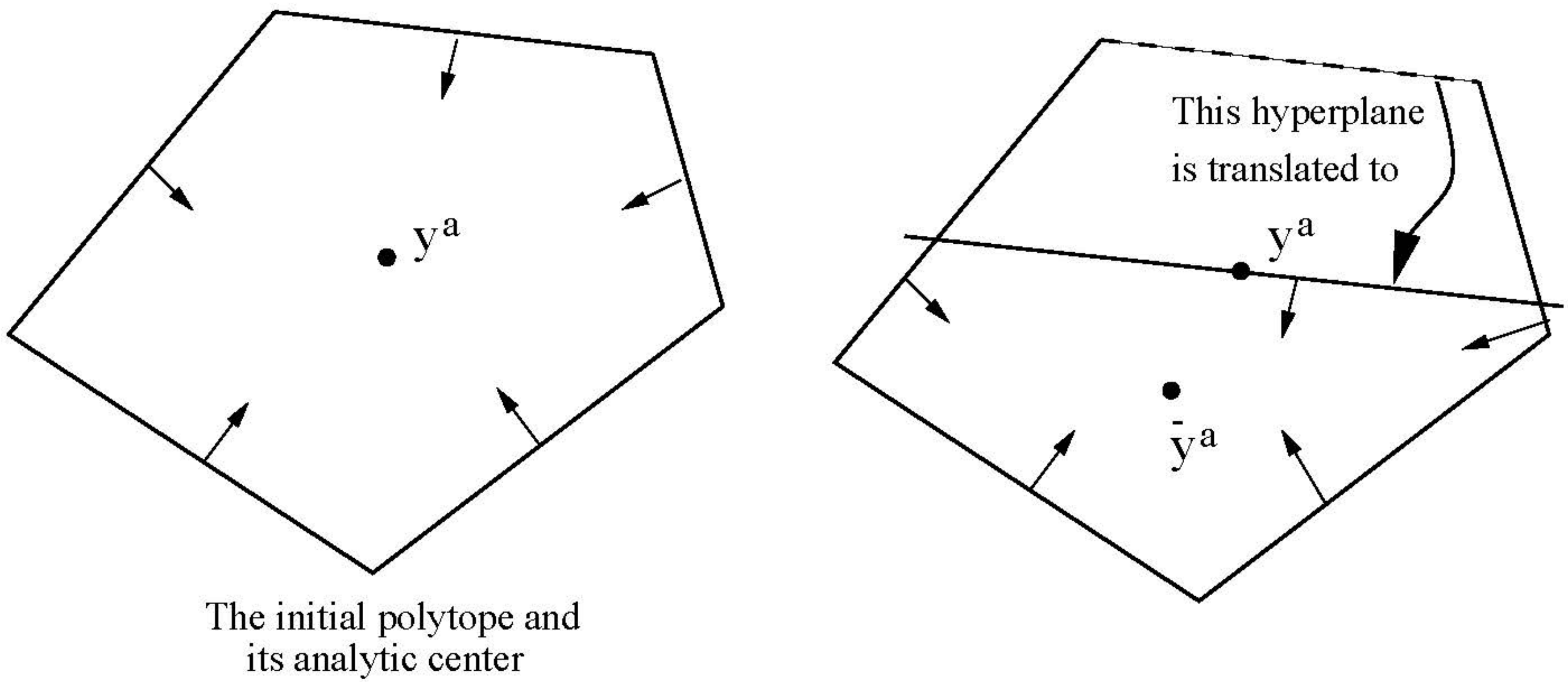

Figure 2: Translation of a hyperplane to the AC.

Analytic Volume Reduction of the New Polytope

Let $\bar{\mathbf{y}}^+$ be the analytic center of $\mathcal{F}^+$. Then, the analytic volume of $\mathcal{F}^+$

$$AV(\mathcal{F}^+) = \prod_{j=1}^{n}(c_j^+ - \mathbf{a}_j^T\bar{\mathbf{y}}^+) = (\mathbf{a}_1^T\bar{\mathbf{y}} - \mathbf{a}_1^T\bar{\mathbf{y}}^+)\prod_{j=2}^{n}(c_j - \mathbf{a}_j^T\bar{\mathbf{y}}^+).$$

We have the following volume reduction theorem:

Theorem 1

$$\frac{AV(\mathcal{F}^+)}{AV(\mathcal{F})} \le \exp(-1).$$

Proof

Since $\bar{\mathbf{y}}$ is the analytic center of $\mathcal{F}$, there exists $\bar{\mathbf{x}} > \mathbf{0}$ such that

$$\bar{X}\bar{\mathbf{s}} = \bar{X}(\mathbf{c} - A^T\bar{\mathbf{y}}) = \mathbf{e} \quad \text{and} \quad A\bar{\mathbf{x}} = \mathbf{0}.$$

Thus,

$$\bar{\mathbf{s}} = (\mathbf{c} - A^T\bar{\mathbf{y}}) = \bar{X}^{-1}\mathbf{e} \quad \text{and} \quad \mathbf{c}^T\bar{\mathbf{x}} = (\mathbf{c} - A^T\bar{\mathbf{y}})^T\bar{\mathbf{x}} = \mathbf{e}^T\mathbf{e} = n.$$

We have

$$\begin{aligned} \mathbf{e}^T\bar{X}\bar{\mathbf{s}}^+ &= \mathbf{e}^T\bar{X}(\mathbf{c}^+ - A^T\bar{\mathbf{y}}^+) = \mathbf{e}^T\bar{X}\mathbf{c}^+ \\ &= \mathbf{c}^T\bar{\mathbf{x}} - \bar{x}_1(c_1 - \mathbf{a}_1^T\bar{\mathbf{y}}) = n - 1. \end{aligned}$$

$$
\begin{aligned}
\frac{AV(\mathcal{F}^+)}{AV(\mathcal{F})} &= \prod_{j=1}^{n} \frac{\bar{s}_j^+}{\bar{s}_j} \\
&= \prod_{j=1}^{n} \bar{x}_j \bar{s}_j^+ \\
&\le \left(\frac{1}{n} \sum_{j=1}^{n} \bar{x}_j \bar{s}_j^+ \right)^n \\
&= \left(\frac{1}{n} \mathbf{e}^T \bar{X} \bar{\mathbf{s}}^+ \right)^n \\
&= \left(\frac{n-1}{n} \right)^n \le \exp(-1).
\end{aligned}
$$

Analytic Volume of Polytope and Multiple Cutting Planes

Now suppose we translate $k(< n)$ hyperplanes, say $1, 2, ..., k$, moved to cut the analytic center $\bar{\mathbf{y}}$ of $\mathcal{F}$, that is,

$$\mathcal{F}^+ := \{\mathbf{y} : \mathbf{a}_j^T \mathbf{y} \le c_j^+, \; j = 1, ..., n\},$$

where $c_j^+ = c_j$ for $j = k+1, ..., n$ and $c_j^+ = \mathbf{a}_j^T \bar{\mathbf{y}}$ for $j = 1, ..., k$.

Corollary 1

$$\frac{AV(\mathcal{F}^+)}{AV(\mathcal{F})} \le \exp(-k).$$

Barrier Regularization Function for LP: Algebraic Implementation

Consider the LP pair with the barrier function

$$(LPB) \quad \text{minimize} \quad \mathbf{c}^T\mathbf{x} - \mu\sum_{j=1}^n \log x_j \qquad\qquad (LDB) \quad \text{maximize} \quad \mathbf{b}^T\mathbf{y} + \mu\sum_{j=1}^n \log s_j$$
$$\text{s.t.} \quad \mathbf{x} \in \text{int}\,\mathcal{F}_p \qquad <=> \qquad \text{s.t.} \quad (\mathbf{y}, \mathbf{s}) \in \text{int}\,\mathcal{F}_d,$$

and they are primal-dual to each other and share a common set of KKT Optimality Conditions:

$$\begin{aligned} X\mathbf{s} &= \mu\mathbf{e} \\ A\mathbf{x} &= \mathbf{b} \\ -A^T\mathbf{y} - \mathbf{s} &= -\mathbf{c}; \end{aligned} \tag{1}$$

where barrier parameter

$$\mu = \frac{\mathbf{x}^T\mathbf{s}}{n} = \frac{\mathbf{c}^T\mathbf{x} - \mathbf{b}^T\mathbf{y}}{n},$$

so that it's the average of complementarity or duality gap. As μ varies, the optimizers form the LP central paths in the primal and dual feasible regions, respectively.

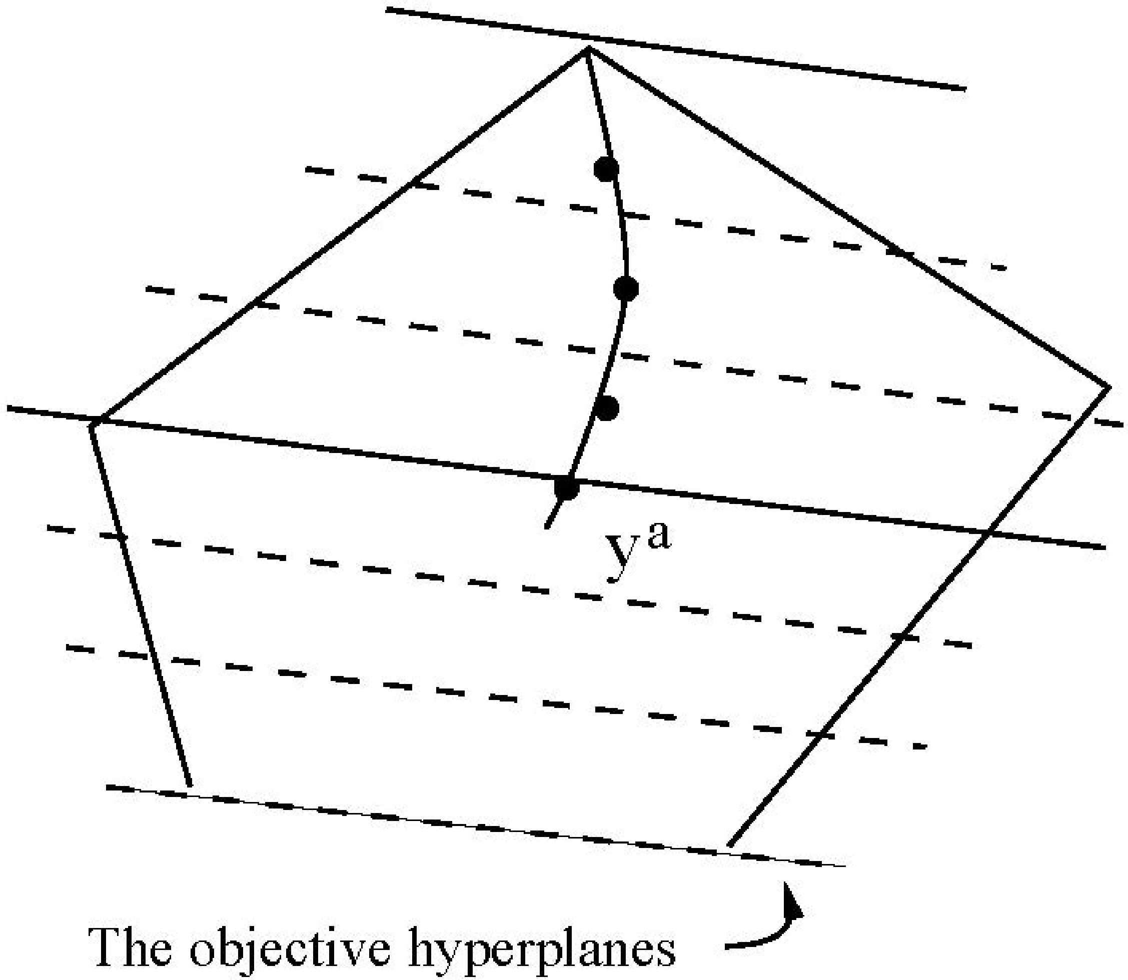

Figure 3: The central path of $\mathbf{y}(\mu)$ in a dual feasible region.

Examples

$$\min \sum_j c_j \mathbf{x}_j - \mu \sum_j \log(x_j) \text{ s.t. } \sum_j x_j = 1.$$

$$c_j - \frac{\mu}{x_j} = y, \; x_j > 0, \; \forall j,$$

thus, $x_j = \frac{\mu}{c_j - y}, \; \forall j$. Then, from

$$\sum_j \frac{\mu}{c_j - y} = 1, \; c_j - y > 0, \; \forall j,$$

we can solve $y(\mu)$ and $\mathbf{x}(\mu)$ as the roots of polynomials.

Central Path for Linear Programming

$$\mathcal{C} = \{(\mathbf{x}(\mu), \mathbf{y}(\mu), \mathbf{s}(\mu)) \in \text{int}\,\mathcal{F} : \; X\mathbf{s} = \mu\mathbf{e}, \; 0 < \mu < \infty\} \,;$$

is called the (primal and dual) central path of linear programming.

Theorem 2 *Let both (LP) and (LD) have interior feasible points for the given data set* $(A, \mathbf{b}, \mathbf{c})$*. Then for any* $0 < \mu < \infty$*, the central path point pair* $(\mathbf{x}(\mu), \mathbf{y}(\mu), \mathbf{s}(\mu))$ *exists and is unique. Moreover, the followings hold.*

i) *The central path point* $(\mathbf{x}(\mu), \mathbf{s}(\mu))$ *is bounded for* $0 < \mu \leq \mu^0$ *and any given* $0 < \mu^0 < \infty$*.*

ii) *For* $0 < \mu' < \mu$*,*

$$\mathbf{c}^T\mathbf{x}(\mu') < \mathbf{c}^T\mathbf{x}(\mu) \quad \textit{and} \quad \mathbf{b}^T\mathbf{y}(\mu') > \mathbf{b}^T\mathbf{y}(\mu)$$

if both primal and dual have no constant objective values.

iii) $(\mathbf{x}(\mu), \mathbf{s}(\mu))$ *converges to an optimal solution pair for (LP) and (LD). Moreover, the limit point* $\mathbf{x}(0)_{P^*} > \mathbf{0}$ *and the limit point* $\mathbf{y}(0), \mathbf{s}(0)_{Z^*} > \mathbf{0}$ *are the analytic centers of the optimal solution sets of primal and dual, respectively; where* (P^*, Z^*) *is the strictly complementarity partition if variable index set* $\{1, 2, ..., n\}$*.*

Proof of (iii)

Since $\mathbf{x}(\mu)$ and $\mathbf{s}(\mu)$ are both bounded, they have at least one limit point which we denote by $\mathbf{x}(0)$ and $\mathbf{s}(0)$. Let $\mathbf{x}^*_{P^*} > \mathbf{0}$ ($\mathbf{x}^*_{Z^*} = \mathbf{0}$) and $\mathbf{s}^*_{Z^*} > \mathbf{0}$ ($\mathbf{s}^*_{P^*} = \mathbf{0}$), be the analytic centers on the optimal sets of on the primal and dual optimal faces, respectively, that is, they are the maximizers of
$\{\prod_{j\in P^*} x_j : \ A_{P^*}\mathbf{x}_{P^*} = \mathbf{b}, \ \mathbf{x}_{P^*} \ge \mathbf{0}\}$ and
$\{\prod_{j\in Z^*} s_j : \ \mathbf{s}_{Z^*} = \mathbf{c}_{Z^*} - A^T_{Z^*}\mathbf{y} \ge \mathbf{0}, \ \mathbf{c}_{P^*} - A^T_{P^*}\mathbf{y} = \mathbf{0}\}$, respectively. Note
$(\mathbf{x}(\mu) - \mathbf{x}^*)^T(\mathbf{s}(\mu) - \mathbf{s}^*) = 0$, so that

$$\sum_j^n \left(s^*_j x(\mu)_j + x^*_j s(\mu)_j\right) = n\mu, \quad \text{or} \quad \sum_{j\in P^*}\left(\frac{x^*_j}{x(\mu)_j}\right) + \sum_{j\in Z^*}\left(\frac{s^*_j}{s(\mu)_j}\right) = n.$$

Therefore, from the arithmetic-geometric mean inequality we have

$$\prod_{j\in P^*}\frac{x^*_j}{x(\mu)_j}\prod_{j\in Z^*}\frac{s^*_j}{s(\mu)_j} \le 1, \quad \text{or} \quad \left(\prod_{j\in P^*} x(\mu)_j\right)\left(\prod_{j\in Z^*} s(\mu)_j\right) \ge \left(\prod_{j\in P^*} x^*_j\right)\left(\prod_{j\in Z^*} s^*_j\right)$$

The limit points must also satisfy the inequality which implies $\prod_{j\in P^*} x(0)_j \ge \prod_{j\in P^*} x^*_j$ and $\prod_{j\in Z^*} s(0)_j \ge \prod_{j\in Z^*} s^*_j$. But the analytic center is unique so that the claim is true.

The Primal-Dual Path-Following Algorithm for LP

In general, we start from an (approximate) central path point $(\mathbf{x}^k, \mathbf{y}^k, \mathbf{s}^k) \in \mathcal{F}$ such that

$$\|X^k\mathbf{s}^k - \mu^k\mathbf{e}\| \le \sigma\mu^k, \quad \text{for some } \sigma \in [0, 1).$$

Then, let $\mu^{k+1} = (1-\eta)\mu^k$ for some $\eta \in (0,\ 1]$, we aim to find a new pair $(\mathbf{x}, \mathbf{y}, \mathbf{s}) \in \mathcal{F}$ such that

$$X\mathbf{s} = \mu^{k+1}\mathbf{e}.$$

We start from $(\mathbf{x}^k, \mathbf{y}^k, \mathbf{s}^k) \in \mathcal{F}$ and apply the Newton iteration for direction vectors $(\mathbf{d}_x, \mathbf{d}_y, \mathbf{d}_s)$:

$$\begin{array}{rcl} S^k\mathbf{d}_x + X^k\mathbf{d}_s & = & \mu^{k+1}\mathbf{e} - X^k\mathbf{s}^k \\ A\mathbf{d}_x & = & \mathbf{0} \\ A^T\mathbf{d}_y + \mathbf{d}_s & = & \mathbf{0} \end{array},$$

then let $\mathbf{x}^{k+1} = \mathbf{x}^k + \mathbf{d}_x, \quad \mathbf{y}^{k+1} = \mathbf{y}^k + \mathbf{d}_y, \quad \mathbf{s}^{k+1} = \mathbf{s}^k + \mathbf{d}_s$. Carefully choosing $\sigma = O(1)$ and $\eta = O(\frac{1}{\sqrt{n}})$ guarantees $(\mathbf{x}^{k+1}, \mathbf{s}^{k+1}) > \mathbf{0}$ and

$$\|X^{k+1}\mathbf{s}^{k+1} - \mu^{k+1}\mathbf{e}\| \le \sigma\mu^{k+1}, \quad \text{for the same } \sigma \in [0, 1).$$

Too many restrictions when following a path... Is a function-driven interior-point algorithm?

Primal-Dual Potential Function for LP

For $(\mathbf{x}, \mathbf{y}, \mathbf{s}) \in \text{int}\,\mathcal{F}$, the joint primal-dual potential function is defined by

$$\psi_{n+\rho}(\mathbf{x}, \mathbf{s}) := (n+\rho)\log(\mathbf{x}^T\mathbf{s}) - \sum_{j=1}^{n}\log(x_j s_j), \quad \text{for some } \rho > 0.$$

$$\psi_{n+\rho}(\mathbf{x}, \mathbf{s}) = \rho\log(\mathbf{x}^T\mathbf{s}) + \psi_n(\mathbf{x}, \mathbf{s}) \geq \rho\log(\mathbf{x}^T\mathbf{s}) + n\log n,$$

then, for $\rho > 0$, $\psi_{n+\rho}(\mathbf{x}, \mathbf{s}) \to -\infty$ implies that $\mathbf{x}^T\mathbf{s} \to 0$. More precisely, we have

$$\mathbf{x}^T\mathbf{s} \leq \exp\left(\frac{\psi_{n+\rho}(\mathbf{x}, \mathbf{s}) - n\log n}{\rho}\right).$$

Given a pair $(\mathbf{x}^k, \mathbf{y}^k, \mathbf{s}^k) \in \text{int}\,\mathcal{F}$, compute direction vectors $(\mathbf{d}_x, \mathbf{d}_y, \mathbf{d}_s)$ from the Newton iteration:

$$\begin{aligned} S^k\mathbf{d}_x + X^k\mathbf{d}_s &= \tfrac{(\mathbf{x}^k)^T\mathbf{s}^k}{n+\rho}\mathbf{e} - X^k\mathbf{s}^k, \\ A\mathbf{d}_x &= \mathbf{0}, \\ A^T\mathbf{d}_y + \mathbf{d}_s &= \mathbf{0}. \end{aligned} \tag{2}$$

How to solve the equation system efficiently using the block structures?

Block Structure in the KKT System

$$\begin{aligned} S^k\mathbf{d}_x + X^k\mathbf{d}_s &= \mathbf{r}^k, \\ A\mathbf{d}_x &= \mathbf{0}, \\ A^T\mathbf{d}_y + \mathbf{d}_s &= \mathbf{0}. \end{aligned}$$

Scale the first block to: $\mathbf{d}_x + (S^k)^{-1}X^k\mathbf{d}_s = (S^k)^{-1}\mathbf{r}^k$.

Multiplying A to both sides and using the second block equations: $A(S^k)^{-1}X^k\mathbf{d}_s = A(S^k)^{-1}\mathbf{r}^k$.

Applying the third block equations: $-A(S^k)^{-1}X^kA^T\mathbf{d}_y = A(S^k)^{-1}\mathbf{r}^k$.

This is an $m \times m$ positive definite system, and solve it for $\mathbf{d}_y$; then $\mathbf{d}_s$ from the third block; then $\mathbf{d}_x$ from the first block.

Positive Definite System Equation Solver: $Q\mathbf{d} = \mathbf{r}$ where Q is a PD matrix.

Matrix Factorization:

- Cholesky: $R^TR = Q$, where R is a Right-Triangle matrix
- $LDL^T = Q$, where L is a Left-Triangle matrix.

Description of Algorithm for LP

Given $(\mathbf{x}^0, \mathbf{y}^0, \mathbf{s}^0) \in \text{int}\, \mathcal{F}$. Set $\rho \geq \sqrt{n}$ and $k := 0$.

While $(\mathbf{x}^k)^T \mathbf{s}^k \geq \epsilon$ **do**

1. Set $(\mathbf{x}, \mathbf{s}) = (\mathbf{x}^k, \mathbf{s}^k)$ and compute $(\mathbf{d}_x, \mathbf{d}_y, \mathbf{d}_s)$ from (2).
2. Let $\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha^k \mathbf{d}_x$, $\mathbf{y}^{k+1} = \mathbf{y}^k + \alpha^k \mathbf{d}_y$, and $\mathbf{s}^{k+1} = \mathbf{s}^k + \alpha^k \mathbf{d}_s$ where
$$\alpha^k = \arg\min_{\alpha \geq 0} \psi_{n+\rho}(\mathbf{x}^k + \alpha \mathbf{d}_x, \mathbf{s}^k + \alpha \mathbf{d}_s).$$
3. Let $k := k + 1$ and return to Step 1.

Theorem 3 *Let $\rho \geq \sqrt{n}$. Then, the potential reduction algorithm generates the (interior) feasible solution sequence $\{\mathbf{x}^k, \mathbf{y}^k, \mathbf{s}^k\}$ such that*

$$\psi_{n+\rho}(\mathbf{x}^{k+1}, \mathbf{s}^{k+1}) - \psi_{n+\rho}(\mathbf{x}^k, \mathbf{s}^k) \leq -0.15.$$

Thus, if $\psi_{n+\rho}(\mathbf{x}^0, \mathbf{s}^0) \leq \rho \log((\mathbf{x}^0)^T \mathbf{s}^0) + n \log n$, the algorithm terminates in at most $O(\rho \log((\mathbf{x}^0)^T \mathbf{s}^0/\epsilon))$ iterations with $(\mathbf{x}^k)^T \mathbf{s}^k = \mathbf{c}^T \mathbf{x}^k - \mathbf{b}^T \mathbf{y}^k \leq \epsilon$.

The proof used a key fact: $\mathbf{d}_x^T \mathbf{d}_s = -\mathbf{d}_x^T A^T \mathbf{d}_y = 0$ for the directions. Also

$$\begin{aligned}
(\mathbf{x}^k)^T \mathbf{s}^k &\leq \exp\left(\frac{\psi_{n+\rho}(\mathbf{x}^k, \mathbf{s}^k) - n \log n}{\rho}\right) \\
&\leq \exp\left(\frac{\psi_{n+\rho}(\mathbf{x}^0, \mathbf{s}^0) - n \log n - \rho \log((\mathbf{x}^0)^T \mathbf{s}^0/\epsilon)}{\rho}\right) \\
&\leq \exp\left(\frac{\rho \log(\mathbf{x}^0, \mathbf{s}^0) - \rho \log((\mathbf{x}^0)^T \mathbf{s}^0/\epsilon)}{\rho}\right) \\
&= \exp(\log(\epsilon)) = \epsilon.
\end{aligned}$$

The role of ρ? And more aggressive step size?

Proof Sketch of the Reduction Theorem

We first have the following lemma:

Lemma 2 *Let the direction vector* $\mathbf{d} = (\mathbf{d}_x, \mathbf{d}_y, \mathbf{d}_s)$ *be computed by (2), and let* $\theta = \frac{\alpha\sqrt{\min(XS\mathbf{e})}}{\|(XS)^{-1/2}\mathbf{r}\|}$ *where* α *is a positive constant less than* 1*. Let*

$$\mathbf{x}^+ = \mathbf{x} + \theta\mathbf{d}_x, \quad \mathbf{y}^+ = \mathbf{y} + \theta\mathbf{d}_y, \quad \textit{and} \quad \mathbf{s}^+ = \mathbf{s} + \theta\mathbf{d}_s.$$

Then, we have $(\mathbf{x}^+, \mathbf{y}^+, \mathbf{s}^+) \in \text{int}\,\mathcal{F}$ *and*

$$\begin{array}{l} \psi_{n+\rho}(\mathbf{x}^+, \mathbf{s}^+) - \psi_{n+\rho}(\mathbf{x}, \mathbf{s}) \\ \le -\alpha\sqrt{\min(XS\mathbf{e})}\|(XS)^{-1/2}(\mathbf{e} - \frac{(n+\rho)}{\mathbf{x}^T\mathbf{s}}X\mathbf{s})\| + \frac{\alpha^2}{2(1-\alpha)} \end{array}.$$

$$
\begin{array}{ll}
& \psi(\mathbf{x}^+,\mathbf{s}^+) - \psi(\mathbf{x},\mathbf{s}) \\
= & (n+\rho)\log\left(1+\frac{\theta\mathbf{d}_s^T\mathbf{x}+\theta\mathbf{d}_x^T\mathbf{s}}{\mathbf{x}^T\mathbf{s}}\right) - \sum_{j=1}^n \left(\log(1+\frac{\theta d_{s_j}}{s_j}) + \log(1+\frac{\theta d_{x_j}}{x_j})\right) \\
\le & (n+\rho)\left(\frac{\theta\mathbf{d}_s^T\mathbf{x}+\theta\mathbf{d}_x^T\mathbf{s}}{\mathbf{x}^T\mathbf{s}}\right) - \sum_{j=1}^n \left(\log(1+\frac{\theta d_{s_j}}{s_j}) + \log(1+\frac{\theta d_{x_j}}{x_j})\right) \\
\le & (n+\rho)\left(\frac{\theta\mathbf{d}_s^T\mathbf{x}+\theta\mathbf{d}_x^T\mathbf{s}}{\mathbf{x}^T\mathbf{s}}\right) - \theta\mathbf{e}^T(S^{-1}\mathbf{d}_s + X^{-1}\mathbf{d}_x) + \frac{||\theta S^{-1}\mathbf{d}_s||^2+||\theta X^{-1}\mathbf{d}_x||^2}{2(1-\alpha)} \\
\le & \frac{n+\rho}{\mathbf{x}^T\mathbf{s}}\theta(\mathbf{d}_s^T\mathbf{x}+\mathbf{d}_x^T\mathbf{s}) - \theta\mathbf{e}^T(S^{-1}\mathbf{d}_s + X^{-1}\mathbf{d}_x) + \frac{\alpha^2}{2(1-\alpha)} \\
= & \theta\left(\frac{n+\rho}{\mathbf{x}^T\mathbf{s}}\mathbf{e}^T(X\mathbf{d}_s+S\mathbf{d}_x) - \mathbf{e}^T(S^{-1}\mathbf{d}_s + X^{-1}\mathbf{d}_x)\right) + \frac{\alpha^2}{2(1-\alpha)} \\
= & \theta\left(\frac{n+\rho}{\mathbf{x}^T\mathbf{s}}\mathbf{e}^T(X\mathbf{d}_s+S\mathbf{d}_x) - \mathbf{e}^T(XS)^{-1}(X\mathbf{d}_s + S\mathbf{d}_x)\right) + \frac{\alpha^2}{2(1-\alpha)} \\
= & \theta\left(\frac{n+\rho}{\mathbf{x}^T\mathbf{s}}XS\mathbf{e}-\mathbf{e}\right)^T(XS)^{-1}(X\mathbf{d}_s + S\mathbf{d}_x) + \frac{\alpha^2}{2(1-\alpha)} \\
= & \theta\left(\frac{n+\rho}{\mathbf{x}^T\mathbf{s}}XS\mathbf{e}-\mathbf{e}\right)^T(XS)^{-1}\left(\frac{\mathbf{x}^T\mathbf{s}}{n+\rho}\mathbf{e} - XS\mathbf{e}\right) + \frac{\alpha^2}{2(1-\alpha)} \\
= & -\theta\cdot\frac{n+\rho}{\mathbf{x}^T\mathbf{s}}\cdot\|(XS)^{-1/2}\mathbf{r}\|^2 + \frac{\alpha^2}{2(1-\alpha)} \\
= & -\alpha\sqrt{\min(XS\mathbf{e})}\cdot\|\frac{n+\rho}{\mathbf{x}^T\mathbf{s}}(XS)^{-1/2}\mathbf{r}\| + \frac{\alpha^2}{2(1-\alpha)}.
\end{array}
$$

Let $\mathbf{v} = XS\mathbf{e}$. Then, we can prove the following technical lemma:

Lemma 3 *Let $\mathbf{v} \in \mathcal{R}^n$ be a positive vector and $\rho \geq \sqrt{n}$. Then,*

$$\sqrt{\min(\mathbf{v})}\|V^{-1/2}(\mathbf{e} - \frac{(n+\rho)}{\mathbf{e}^T\mathbf{v}}\mathbf{v})\| \geq \sqrt{3/4}\,.$$

Combining these Lemmas 2 and 3 we have

$$\psi_{n+\rho}(\mathbf{x}^+, \mathbf{s}^+) - \psi_{n+\rho}(\mathbf{x}, \mathbf{s})$$

$$\leq -\alpha\sqrt{3/4} + \frac{\alpha^2}{2(1-\alpha)} = -\delta$$

for a constant δ.

Initialization

- Combining the primal and dual into a single linear feasibility problem, then applying LP algorithms to find a feasible point of the problem. Theoretically, this approach can retain the currently best complexity result.
- The big M method, i.e., add one or more artificial column(s) and/or row(s) and a huge penalty parameter M to force solutions to become feasible during the algorithm.
- Phase I-then-Phase II method, i.e., first try to find a feasible point (and possibly one for the dual problem), and then start to look for an optimal solution if the problem is feasible and bounded.
- Combined Phase I-Phase II method, i.e., approach feasibility and optimality simultaneously. To our knowledge, the "best" complexity of this approach is $O(n\log((\mathbf{x}^0)^T\mathbf{s}^0/\epsilon))$.

Homogeneous and Self-Dual Algorithm

- It solves the linear programming problem without any regularity assumption concerning the existence of optimal, feasible, or interior feasible solutions, while it retains the currently best complexity result
- It can start at any positive primal-dual pair, feasible or infeasible, near the central ray of the positive orthant (cone), and it does not use any big M penalty parameter or lower bound.
- Each iteration solves a system of linear equations whose dimension is almost the same as that solved in the standard (primal-dual) interior-point algorithms.
- If the LP problem has a solution, the algorithm generates a sequence that approaches feasibility and optimality simultaneously; if the problem is infeasible or unbounded, the algorithm will produce an infeasibility certificate for at least one of the primal and dual problems.

Primal-Dual Alternative Systems

Recall that a pair of LP has two alternatives

$$
\text{(Solvable)}\quad
\begin{aligned}
A\mathbf{x}-\mathbf{b} &= \mathbf{0}\\
-A^T\mathbf{y}+\mathbf{c} &\ge \mathbf{0},\\
\mathbf{b}^T\mathbf{y}-\mathbf{c}^T\mathbf{x} &= 0,\\
\mathbf{y}\ \text{free},\ \mathbf{x} &\ge \mathbf{0}
\end{aligned}
\qquad \text{or} \qquad
\text{(Infeasible)}\quad
\begin{aligned}
A\mathbf{x} &= \mathbf{0}\\
-A^T\mathbf{y} &\ge \mathbf{0},\\
\mathbf{b}^T\mathbf{y}-\mathbf{c}^T\mathbf{x} &> 0,\\
\mathbf{y}\ \text{free},\ \mathbf{x} &\ge \mathbf{0}
\end{aligned}
$$

$$
(HP)\qquad
\begin{aligned}
A\mathbf{x}-\mathbf{b}\tau &= \mathbf{0}\\
-A^T\mathbf{y}+\mathbf{c}\tau &= \mathbf{s}\ge \mathbf{0},\\
\mathbf{b}^T\mathbf{y}-\mathbf{c}^T\mathbf{x} &= \kappa \ge 0,\\
\mathbf{y}\ \text{free},\ (\mathbf{x};\tau) &\ge \mathbf{0}
\end{aligned}
$$

where the two alternatives are:

$$
\text{(Solvable)}:\ (\tau>0,\kappa=0)\quad \text{or}\quad \text{(Infeasible)}:\ (\tau=0,\kappa>0)
$$

Let's Find a Feasible Solution of (HP)

Given $\mathbf{x}^0 = \mathbf{e} > \mathbf{0}$, $\mathbf{s}^0 = \mathbf{e} > \mathbf{0}$, and $\mathbf{y}^0 = \mathbf{0}$, we formulate a self-dual LP problem:

$$
\begin{array}{llrrrrl}
(HS-DP) & \min & & & & (n+1)\theta & \\
 & \text{s.t.} & & A\mathbf{x} & -\mathbf{b}\tau & +\bar{\mathbf{b}}\theta & = \mathbf{0}, \\
 & & -A^T\mathbf{y} & & +\mathbf{c}\tau & -\bar{\mathbf{c}}\theta & \geq \mathbf{0}, \\
 & & \mathbf{b}^T\mathbf{y} & -\mathbf{c}^T\mathbf{x} & & +\bar{z}\theta & \geq 0, \\
 & & -\bar{\mathbf{b}}^T\mathbf{y} & +\bar{\mathbf{c}}^T\mathbf{x} & -\bar{z}\tau & & = -(n+1), \\
 & & \mathbf{y} \text{ free}, & \mathbf{x} \geq \mathbf{0}, & \tau \geq 0, & \theta \text{ free}. &
\end{array}
$$

Note that $(\mathbf{y} = \mathbf{0},\ \mathbf{x} = \mathbf{e},\ \tau = 1,\ \theta = 1)$ is a strictly feasible point for (HSDP). Moreover, one can show that the constraints imply

$$
\mathbf{e}^T x + \mathbf{e}^T s + \tau + \kappa - (n+1)\theta = (n+1),
$$

which serves as a normalizing constraint for (HSDP) to prevent the all-zero solution.

Main Result

Theorem 4 *The interior-point algorithm solves (HS-DP) in* $O(\sqrt{n}\log\frac{n}{\epsilon})$ *steps and each step solves a system of linear equations as the same size as in feasible algorithms, and it always produces an optimal solution* $(\mathbf{y}^*, \mathbf{x}^*, \tau^*, \mathbf{s}^*, \kappa^*, \theta^* = 0)$ *where* $\tau^* + \kappa^* > 0$*. If* $\tau^* > 0$ *then it produces an optimal solution pair for the original LP problem; if* $\kappa^* > 0$*, then it produces a certificate to prove (at least) one of the pair is infeasible.*

Extensions to Solving SDP: Potential Function

For any $X \in \operatorname{int} \mathcal{F}_p$ and $(\mathbf{y}, S) \in \operatorname{int} \mathcal{F}_d$, let parameter $\rho > 0$ and

$$\psi_{n+\rho}(X, S) := (n+\rho)\log(X \bullet S) - \log(\det(X)\cdot\det(S)),$$

$$\psi_{n+\rho}(X, S) = \rho\log(X \bullet S) + \psi_n(X, S) \geq \rho\log(X \bullet S) + n\log n.$$

Then, $\psi_{n+\rho}(X, S) \to -\infty$ implies that $X \bullet S \to 0$. More precisely, we have

$$X \bullet S \leq \exp(\frac{\psi_{n+\rho}(X, S) - n\log n}{\rho}).$$

Primal-Dual SDP Alternative Systems

A pair of SDP has two alternatives under mild conditions

(Solvable)
$$\begin{array}{rl} \mathcal{A}X - \mathbf{b} & = \mathbf{0} \\ -\mathcal{A}^T\mathbf{y} + C & \succeq \mathbf{0}, \\ \mathbf{b}^T\mathbf{y} - C \bullet X & = 0, \\ \mathbf{y} \text{ free},\ X & \succeq \mathbf{0} \end{array}$$

or

(Infeasible)
$$\begin{array}{rl} \mathcal{A}X & = \mathbf{0} \\ -\mathcal{A}^T\mathbf{y} & \succeq \mathbf{0}, \\ \mathbf{b}^T\mathbf{y} - C \bullet X & > 0, \\ \mathbf{y} \text{ free},\ X & \succeq \mathbf{0} \end{array}$$

An Integrated Homogeneous System

The two alternative systems can be homogenized as one:

$$
\begin{array}{rrl}
(HSDP) & \mathcal{A}X - \mathbf{b}\tau & = \mathbf{0} \\
 & -\mathcal{A}^T\mathbf{y} + C\tau & = \mathbf{s} \geq \mathbf{0}, \\
 & \mathbf{b}^T\mathbf{y} - C \bullet X & = \kappa \geq 0, \\
 & \mathbf{y} \text{ free}, \; X \succeq \mathbf{0}, & \tau \geq 0,
\end{array}
$$

where the three alternatives are

$$
\begin{array}{rl}
\text{(Solvable)}: & (\tau > 0, \kappa = 0) \\
\text{(Infeasible)}: & (\tau = 0, \kappa > 0) \\
\text{(All others)}: & (\tau = \kappa = 0).
\end{array}
$$

Software Implementation

Cplex-Barrier IBM, GUROBI, COPT

SEDUMI: http://sedumi.mcmaster.ca/

MOSEK: http://www.mosek.com/products_mosek.html

SDDPT3: http://www.math.nus.edu.sg/~mattohkc/sdpt3.html

DSDP (Dual Semidefinite Programming Algorithm):

http://www.stanford.edu/~yyye/Col.html

CVX/ECOS: http://www.stanford.edu/~boyd/cvx

hsdLPsolver and more: http://www.stanford.edu/~yyye/matlab.html