Barrier Methods for Conic Linear Optimization

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

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

Conic LP

$$(CLP) \quad \text{minimize} \quad \mathbf{c} \bullet \mathbf{x}$$
$$\text{subject to} \quad \mathbf{a}_i \bullet \mathbf{x} = b_i, i = 1, 2, ..., m, \ \mathbf{x} \in K,$$

where K is a closed and pointed convex cone.

The dual problem is

$$(CLD) \quad \text{maximize} \quad \mathbf{b}^T \mathbf{y}$$
$$\text{subject to} \quad \sum_i^m y_i \mathbf{a}_i + \mathbf{s} = \mathbf{c}, \ \mathbf{s} \in K^*,$$

where $y \in \mathcal{R}^m$, $\mathbf{s}$ is called the dual slack vector/matrix, and K^* is the dual cone of K.

Denote the feasible regions of (CLP) and (CLD) by $\mathcal{F}_p$ and $\mathcal{F}_d$ respectively.

Barrier Functions for CLP

Linear Programming (LP): $\mathbf{c}, \mathbf{a}_i, \mathbf{x} \in \mathcal{R}^n$ and $K = \mathcal{R}^n_+$

$$B_p(\mathbf{x}) = -\sum_{j=1}^{n} \log x_j, \quad B_d(\mathbf{s}) = \sum_{j=1}^{n} \log s_j.$$

Second-Order Cone Programming (SOCP): $\mathbf{c}, \mathbf{a}_i, \mathbf{x} \in \mathcal{R}^n$ and $K = SOC = \{\mathbf{x} :\ x_1 \geq \|\mathbf{x}_{2:n}\|\}$.

$$B_p(\mathbf{x}) = -\log(x_1^2 - \|\mathbf{x}_{2:n}\|^2), \quad B_d(\mathbf{s}) = \log(s_1^2 - \|\mathbf{s}_{2:n}\|^2).$$

Semidefinite Programming (SDP): $\mathbf{c}, \mathbf{a}_i, \mathbf{x} \in \mathcal{S}^n$ and $K = \mathcal{S}^n_+$

$$B_p(\mathbf{x}) = -\log(\det(X)), \quad B_d(\mathbf{s}) = \log(\det(S)).$$

p-Order Cone Programming (POCP): $\mathbf{c}, \mathbf{a}_i, \mathbf{x} \in \mathcal{R}^n$ and $K = POC = \{\mathbf{x} :\ x_1 \geq \|\mathbf{x}_{2:n}\|_p\}$: $\quad B_p(\mathbf{x}) = ???$

Barrier Optimization and Analytic Center

Consider the barrier function optimization

$$
\begin{array}{lll}
(PB) & \text{minimize} & B_p(\mathbf{x}) \\
 & \text{s.t.} & \mathbf{x} \in \operatorname{int} \mathcal{F}_p
\end{array}
$$

and

$$
\begin{array}{lll}
(DB) & \text{maximize} & B_d(\mathbf{s}) \\
 & \text{s.t.} & (\mathbf{y}, \mathbf{s}) \in \operatorname{int} \mathcal{F}_d
\end{array}
$$

They are constrained convex programs.

The optimizers are called "analytic centers" of the primal and dual polyhedrons.

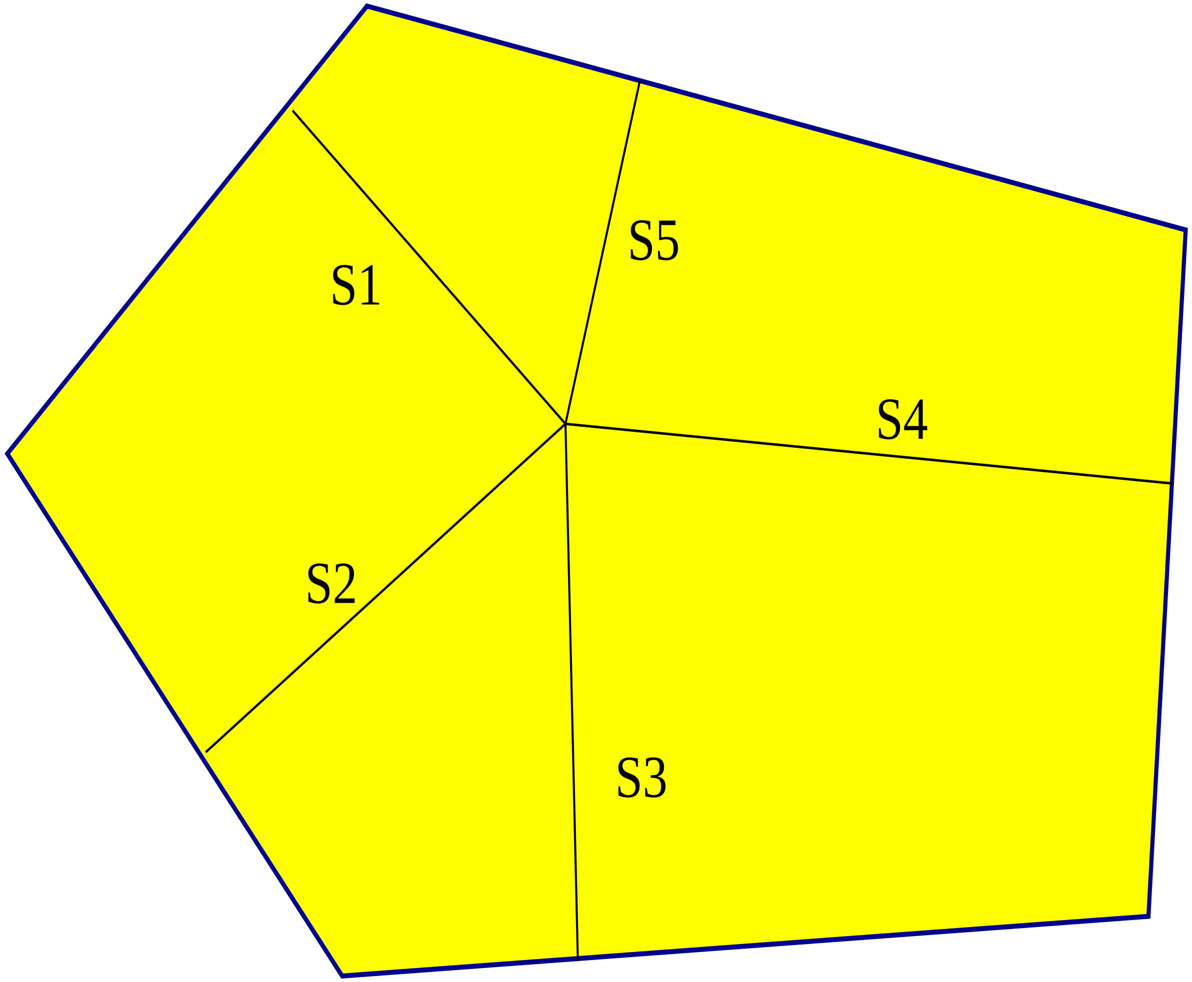

Figure 1: Analytic center maximizes the product of slacks.

Why Analytic

The analytic center of polytope $\mathcal{F}_d$ is an analytic function of input data $A, \mathbf{c}$.

Consider $\Omega = \{y \in R : -y \le 0,\ y \le 1\}$, which is interval $[0, 1]$. The analytic center is $\bar{y} = 1/2$ with $\mathbf{x} = (2, 2)^T$.

Consider

$$\Omega' = \{y \in R : \overbrace{-y \le 0, \cdots, -y \le 0}^{n \text{ times}},\ y \le 1\},$$

which is, again, interval $[0, 1]$ but "$-y \le 0$" is copied n times. The analytic center for this system is $\bar{y} = n/(n + 1)$ with $\mathbf{x} = ((n + 1)/n,\ \cdots,\ (n + 1)/n,\ (n + 1))^T$.

CLP with Barrier Function

Consider the LP problem with the barrier function

$$(CLPB) \quad \text{minimize} \quad \mathbf{c}^T\mathbf{x} + \mu B_p(\mathbf{x})$$
$$\text{s.t.} \quad \mathbf{x} \in \operatorname{int} \mathcal{F}_p$$

and also

$$(CLDB) \quad \text{maximize} \quad \mathbf{b}^T\mathbf{y} + \mu B_d(\mathbf{s})$$
$$\text{s.t.} \quad (\mathbf{y}, \mathbf{s}) \in \operatorname{int} \mathcal{F}_d,$$

where μ is called the barrier (weight) parameter.

They are again constrained convex programs for any fixed μ.

Optimality Conditions for both LPB and LDB

Consider the linear programming (LP) dual case:

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

where we have

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

so that it's the average of complementarity or duality gap.

The minimizers, $\mathbf{y}(\mu), \mathbf{s}(\mu)$, together with $\mathbf{x}(\mu)$, are unique. (Note that $\mathbf{x}(\mu)$ is not the original multipliers of $\mathbf{s} \geq \mathbf{0}$.)

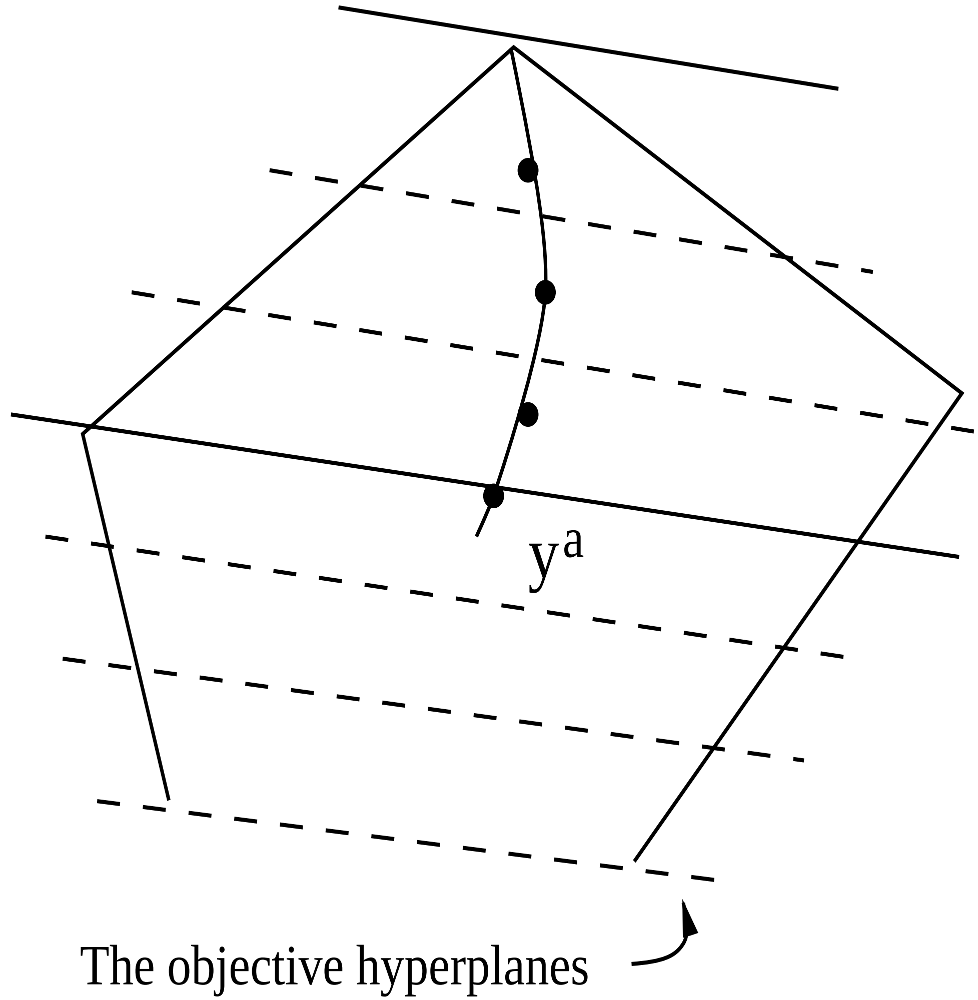

Figure 2: The central path of $\mathbf{y}(\mu)$ in a dual feasible region.

Central Path for Linear Programming

The path

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

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

Theorem 1 *Let both (LP) and (LD) have interior feasible points for the given data set (A, b, 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, and they converge to the analytic centers of the optimal solution sets of the primal and dual problems, respectively.*

The Primal-Dual Path-Following Algorithm

In general, one can start from an (approximate) central path point $\mathbf{x}(\mu^0)$, $(\mathbf{y}(\mu^0), \mathbf{s}(\mu^0))$, or $(\mathbf{x}(\mu^0), \mathbf{y}(\mu^0), \mathbf{s}(\mu^0))$ where μ^0 is sufficiently large.

Then, let μ^1 be a slightly smaller parameter than μ^0. Then, we compute an (approximate) central path point $\mathbf{x}(\mu^1)$, $(\mathbf{y}(\mu^1), \mathbf{s}(\mu^1))$, or $(\mathbf{x}(\mu^1), \mathbf{y}(\mu^1), \mathbf{s}(\mu^1))$. They can be updated from the previous point at μ^0 using the Newton method.

μ might be reduced at each stage by a specific factor, giving $\mu^{k+1} = \gamma\mu^k$ where γ is a fixed positive constant less than one, and k is the stage count.

This is called the primal, dual, or primal-dual path-following method.

LP Primal-Dual Path-Following Algorithm I

More precisely, given a pair $(\mathbf{x}, \mathbf{y}, \mathbf{s}) \in \operatorname{int} \mathcal{F}$ and

$$\|\mu \mathbf{e} - XS\mathbf{e}\| \le \frac{1}{3}\mu, \quad \text{where } \mu = \frac{\mathbf{x}^T\mathbf{s}}{n},$$

we can compute direction vectors $\mathbf{d}_x$, $\mathbf{d}_y$ and $\mathbf{d}_s$ from the Newton system equations:

$$\begin{aligned} S\mathbf{d}_x + X\mathbf{d}_s &= (1 - \tfrac{1}{3\sqrt{n}})\mu\mathbf{e} - XS\mathbf{e}, \\ A\mathbf{d}_x &= \mathbf{0}, \\ -A^T\mathbf{d}_y - \mathbf{d}_s &= \mathbf{0}. \end{aligned} \tag{1}$$

Note that $\mathbf{d}_x^T\mathbf{d}_s = -\mathbf{d}_x^T A^T\mathbf{d}_y = 0$.

LP Primal-Dual Path-Following Algorithm II

Let $\mathbf{x}^+ = \mathbf{x} + \mathbf{d}_x,\ \mathbf{y}^+ = \mathbf{y} + \mathbf{d}_y,\ \mathbf{s}^+ = \mathbf{s} + \mathbf{d}_s$. Then, we have

Theorem 2

$$(\mathbf{x}^+, \mathbf{y}^+, \mathbf{s}^+) \in \operatorname{int} \mathcal{F},$$

and

$$\|\mu^+ \mathbf{e} - X^+ S^+ \mathbf{e}\| \le \frac{1}{3}\mu^+, \quad \textit{where } \mu^+ = \frac{(\mathbf{x}^+)^T \mathbf{s}^+}{n} = (1 - \frac{1}{3\sqrt{n}})\mu.$$

It is easy to see

$$A\mathbf{x}^+ = \mathbf{b} \quad \text{and} \quad A^T \mathbf{y}^+ + \mathbf{s}^+ = \mathbf{c}.$$

Proof Sketch

$$X^{-.5}S^{.5}\mathbf{d}_x + S^{-.5}X^{.5}\mathbf{d}_s = (XS)^{-.5}\left(\mu\mathbf{e} - XS\mathbf{e} - \frac{\mu}{3\sqrt{n}}\mathbf{e}\right).$$

$$\begin{aligned}\|X^{-.5}S^{.5}\mathbf{d}_x + S^{-.5}X^{.5}\mathbf{d}_s\|^2 &\le \|(XS)^{-1}\| \cdot \| \left(\mu\mathbf{e} - XS\mathbf{e} - \tfrac{\mu}{3\sqrt{n}}\mathbf{e}\right) \|^2 \\ &= \frac{1}{\min(XS\mathbf{e})} \cdot \left(\|\mu\mathbf{e} - XS\mathbf{e}\|^2 + \tfrac{\mu^2}{9n}\|\mathbf{e}\|^2\right) \\ &\le \tfrac{3}{2\mu}\left(\tfrac{\mu^2}{9} + \tfrac{\mu^2}{9}\right) = \tfrac{\mu}{3}.\end{aligned}$$

$$\|X^{-.5}S^{.5}\mathbf{d}_x\|^2 + \|S^{-.5}X^{.5}\mathbf{d}_s\|^2 = \|X^{-.5}S^{.5}\mathbf{d}_x + S^{-.5}X^{.5}\mathbf{d}_s\|^2 \le \frac{\mu}{3}.$$

$$\|X^{-1}\mathbf{d}_x\|^2 + \|S^{-1}\mathbf{d}_x\|^2 = \|(XS)^{-.5}X^{-.5}S^{.5}\mathbf{d}_x\|^2 + \|(XS)^{-.5}X^{.5}S^{-.5}\mathbf{d}_s\|^2$$

$$\le \|(XS)^{-1}\|(\|X^{-.5}S^{.5}\mathbf{d}_x\|^2 + \|X^{.5}S^{-.5}\mathbf{d}_s\|^2) \le \frac{1}{\min(XS\mathbf{e})}\cdot\frac{\mu}{3} \le \frac{3}{2\mu}\cdot\frac{\mu}{3} = \frac{1}{2}.$$

Proof Sketch continued

Summing the first set of equations:

$$(\mathbf{x}^+)^T\mathbf{s}^+ = (\mathbf{x}+\mathbf{d}_x)^T(\mathbf{s}+\mathbf{d}_s) = (1-\frac{1}{3\sqrt{n}})\mu.$$

$$\begin{aligned}
\|\mu^+\mathbf{e} - X^+S^+\mathbf{e}\| &\leq \|\mu^+ - XS\mathbf{e} - S\mathbf{d}_x - X\mathbf{d}_s - D_xD_s\mathbf{e}\| \\
&= \|D_xD_s\mathbf{e}\| = \|(X^{-.5}S^{.5}D_x)S^{-.5}X^{.5}D_s\mathbf{e}\| \\
&\leq \tfrac{1}{2}(\|X^{-.5}S^{.5}\mathbf{d}_x + S^{-.5}X^{.5}\mathbf{d}_s\|^2) \\
&\leq \tfrac{1}{2}\tfrac{\mu}{3} \leq \tfrac{1}{3}\mu^+.
\end{aligned}$$

Primal-Dual Potential Function for LP

Yet there is another interior-point algorithm: potential reduction algorithm. For $\mathbf{x} \in \mathrm{int}\,\mathcal{F}_p$ and $(\mathbf{y}, \mathbf{s}) \in \mathrm{int}\,\mathcal{F}_d$, the joint primal-dual potential function is defined by

$$\psi_{n+\rho}(\mathbf{x}, \mathbf{s}) := (n+\rho)\log(\mathbf{x}^T\mathbf{s}) - \sum_{j=1}^{n} \log(x_j s_j),$$

where $\rho \geq 0$.

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

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

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

Primal-Dual Potential Reduction Algorithm for LP

Once have a pair $(\mathbf{x}, \mathbf{y}, \mathbf{s}) \in \mathrm{int}\,\mathcal{F}$, we again compute direction vectors $\mathbf{d}_x$, $\mathbf{d}_y$ and $\mathbf{d}_s$ from the Newton system equations:

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

Note that $\mathbf{d}_x^T\mathbf{d}_s = -\mathbf{d}_x^T A^T\mathbf{d}_y = 0$ here.

Description of Algorithm

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 $\gamma = n/(n+\rho)$ and compute $(\mathbf{d}_x, \mathbf{d}_y, \mathbf{d}_s)$ from (2).

2. Let $\mathbf{x}^{k+1} = \mathbf{x}^k + \bar{\alpha}\mathbf{d}_x$, $\mathbf{y}^{k+1} = \mathbf{y}^k + \bar{\alpha}\mathbf{d}_y$, and $\mathbf{s}^{k+1} = \mathbf{s}^k + \bar{\alpha}\mathbf{d}_s$ where

$$\bar{\alpha} = \arg\min_{\alpha \geq 0} \psi_{n+\rho}(\mathbf{x}^k + \alpha\mathbf{d}_x, \mathbf{s}^k + \alpha\mathbf{d}_s).$$

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

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

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

The role of ρ? And more aggressive step size?

Logarithmic Approximation Lemma

The proof uses a technical lemma:

Lemma 1 *If* $\mathbf{d} \in \mathcal{R}^n$ *such that* $\|\mathbf{d}\|_\infty < 1$ *then*

$$\mathbf{e}^T\mathbf{d} \geq \sum_{i=1}^{n} \log(1+d_i) \geq \mathbf{e}^T\mathbf{d} - \frac{\|\mathbf{d}\|^2}{2(1-\|\mathbf{d}\|_\infty)}.$$

The proof is based on the Taylor expansion of $\ln(1+d_i)$ for $-1 < d_i < 1$.

Lemma 2 *If* $D \in \mathcal{S}^n$ *and* $\|D\|_\infty < 1$, *then,*

$$I \bullet D) \geq \log\det(I+D) \geq I \bullet D - \frac{\|D\|^2}{2(1-\|D\|_\infty)}\ .$$

Semidefinite Programming (SDP)

$$(SDP) \quad \text{Minimize} \quad C \bullet X$$
$$\text{subject to} \quad \mathcal{A}X = \mathbf{b},\ X \succeq 0.$$

The dual problem to (SDP) can be written as:

$$(SDD) \quad \text{Maximize} \quad \mathbf{b}^T\mathbf{y}$$
$$\text{subject to} \quad \mathcal{A}^T\mathbf{y} + S = C,\ S \succeq 0.$$

$$\mathcal{A}X = \begin{pmatrix} A_1 \bullet X \\ \dots \\ A_m \bullet X \end{pmatrix} \quad \text{and} \quad \mathcal{A}^T\mathbf{y} = \sum_{i=1}^{m} y_i A_i.$$

Primal-Dual Potential Functions for SDP

For any $X \in \text{int}\,\mathcal{F}_p$ and $(\mathbf{y}, S) \in \text{int}\,\mathcal{F}_d$,
$\psi_{n+\rho}(X,S) := (n+\rho)\log(X \bullet S) - \log(\det(X)\cdot\det(S))$

$$\psi_n(X,S) \geq n \log n.$$

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

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

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

Primal-Dual (Symmetric) Algorithm for SDP

Once we have a pair $(X, \mathbf{y}, S) \in \text{int}\,\mathcal{F}$ with $\mu = S \bullet X/n$, we can apply the primal-dual Newton method to generate a new iterate X^+ and $(\mathbf{y}^+, S^+)$ as follows: Solve for D_X, $\mathbf{d}_y$ and D_S from the system of linear equations:

$$\begin{array}{rcl} D^{-1}D_X D^{-1} + D_S & = & \frac{X \bullet S}{n+\rho} X^{-1} - S, \\ \mathcal{A}D_X & = & \mathbf{0}, \\ -\mathcal{A}^T \mathbf{d}_y - D_S & = & \mathbf{0}, \end{array} \quad (3)$$

where

$$D = X^{.5}(X^{.5}SX^{.5})^{-.5}X^{.5}.$$

Note that $D_S \bullet D_X = 0$.

Description of Algorithm

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

While $S^k \bullet X^k \geq \epsilon$ **do**

1. Set $(X, S) = (X^k, S^k)$ and $\gamma = n/(n+\rho)$ and compute $(D_X, \mathbf{d}_y, D_S)$ from (3).

2. Let $X^{k+1} = X^k + \bar{\alpha} D_X$, $\mathbf{y}^{k+1} = \mathbf{y}^k + \bar{\alpha}\mathbf{d}_y$, and $S^{k+1} = S^k + \bar{\alpha} D_S$, where

$$\bar{\alpha} = \arg\min_{\alpha \geq 0} \psi_{n+\rho}(X^k + \alpha D_X, S^k + \alpha D_S).$$

3. Let $k := k + 1$ and return to Step 1.

Software Implementation

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`.

hsdLPsolver and more:

`http://www.stanford.edu/~yyye/matlab.html`

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