Interior Point Algorithms

Yinyu Ye
Department of Management Science and Engineering
Stanford University

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

Potential Function

Let Ω be a bounded polytope in $\mathcal{R}^m$ represented by $n\ (> m)$ linear inequalities

$$\Omega = \{y \in \mathcal{R}^m : c - A^T y \geq 0\},$$

where $A \in \mathcal{R}^{m \times n}$ and $c \in \mathcal{R}^n$ are given and A has rank m. Denote the interior of Ω by

$$\text{int}\,\Omega = \{y \in \mathcal{R}^m : c - A^T y > 0\}.$$

$$d(y) = \prod_{j=1}^{n} (c_j - a_j^T y), \quad y \in \Omega,$$

where $a_{.j}$ is the jth column of A.

$$\mathcal{B}(y) := \log d(y) = \sum_{j=1}^{n} \log(c_j - a_{.j}^T y) = \sum_{j=1}^{n} \log s_j, \qquad (1)$$

$-\mathcal{B}(y)$ is the classical logarithmic barrier function.

Analytic Center

The interior point, denoted by y^a and $s^a = c - A^T y^a$, in Ω that maximizes the potential function is called the analytic center of Ω, i.e.,

$$\mathcal{B}(y^a) = \max_{y \in \Omega} \mathcal{B}(y).$$

(y^a, s^a) is uniquely defined, since the potential function is strictly concave in a bounded convex $\operatorname{int} \Omega$.

Analytic Center Condition

Setting $\nabla \mathcal{B}(y, \Omega) = 0$ and letting $x^a = D(s^a)^{-1}e$, the analytic center (y^a, s^a) together with x^a satisfy the following optimality conditions:

$$\begin{aligned} D(x)s &= e \\ Ax &= 0 \\ -A^T y - s &= -c. \end{aligned} \tag{2}$$

Note that adding or deleting a redundant inequality changes the location of the analytic center.

Examples

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

Consider

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

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

Analytic Center for SDP

Let Ω be a bounded convex set in $\mathcal{R}^m$ represented by $n\ (> m)$ a matrix inequality, i.e.,

$$\Omega = \{y \in \mathcal{R}^m : C - \sum_i^m y_i A_i \succeq 0, \}.$$

Let $S = C - \sum_i^m y_i A_i$ and

$$\mathcal{B}(y) := \log\det(S)) = \log\det(C - \sum_i^m y_i A_i). \quad (3)$$

The interior point, denoted by y^a and $S^a = C - \sum_i^m y_i^a A_i$, in Ω that maximizes the potential function is called the analytic center of Ω, i.e.,

$$\mathcal{B}(y^a) = \max_{y \in \Omega} \mathcal{B}(y).$$

Analytic Center Condition for SDP

Setting $\nabla \mathcal{B}(y,\Omega)=0$ and letting $X^a=(S^a)^{-1}$, the analytic center (y^a,S^a) together with X^a satisfy the following optimality conditions:

$$\begin{array}{rcl} XS & = & I \\ \mathcal{A}X & = & 0 \\ -\mathcal{A}^T y - S & = & -C. \end{array} \quad (4)$$

Potential Functions for LP

For $x \in \mathrm{int}\,\mathcal{F}_p$ and $(y, s) \in \mathrm{int}\,\mathcal{F}_d$ it is defined by

$$\psi_{n+\rho}(x, s) := (n+\rho)\log(x \bullet s) - \sum_{j=1}^{n} \log(x_j s_j), \tag{5}$$

where $\rho \geq 0$.

$$\begin{aligned}
\psi_{n+\rho}(x, s) &= (n+\rho)\log(c^T x - b^T y) - \sum_{j=1}^{n} \log x_j - \sum_{j=1}^{n} \log s_j \\
&= \mathcal{P}_{n+\rho}(x, b^T y) - \sum_{j=1}^{n} \log s_j \\
&= \mathcal{B}_{n+\rho}(y, c^T x) - \sum_{j=1}^{n} \log x_j .
\end{aligned}$$

Potential and Duality Gap

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

then, for $\rho > 0$, $\psi_{n+\rho}(x,s) \rightarrow -\infty$ implies that $x^T s \rightarrow 0$.

More precisely, we have

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

A potential reduction algorithm generates sequences $\{x^k, y^k, s^k\} \in \mathrm{int}\,\mathcal{F}$ such that

$$\psi_{n+\sqrt{n}}(x^{k+1}, y^{k+1}, s^{k+1}) \leq \psi_{n+\sqrt{n}}(x^k, y^k, s^k) - .05$$

for $k = 0, 1, 2, ...$. This indicates that the level sets shrink at least a constant rate independently of m or n.

Potential Functions for SDP

For any $X \in \operatorname{int} \mathcal{F}_p$ and $(y, S) \in \operatorname{int} \mathcal{F}_d$,

$$\mathcal{P}_{n+\rho}(X, z) := (n+\rho)\log(C \bullet X - z) - \log\det(X), \quad z \le z^*;$$

$$\mathcal{B}_{n+\rho}(y, z) := (n+\rho)\log(z - b^T y) - \log\det(S), \quad z \ge z^*,$$

where $\rho \ge 0$ and z^* designates the optimal objective value.

$$\begin{aligned}
\psi_{n+\rho}(X, S) &:= (n+\rho)\log(X \bullet S) - \log(\det(X)\cdot\det(S)) \\
&= (n+\rho)\log(C \bullet X - b^T y) - \log\det(X) - \log\det(S) \\
&= \mathcal{P}_{n+\rho}(X, b^T y) - \log\det(S) \\
&= \mathcal{B}_{n+\rho}(S, C \bullet X) - \log\det(X),
\end{aligned}$$

Potential and Duality Gap

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

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

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

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

Central Path for LP

The central path can be expressed as

$$\mathcal{C} = \left\{ (x, y, s) \in \operatorname{int} \mathcal{F} : \; Xs = \frac{x \bullet s}{n} e \right\}$$

in the primal-dual form. We also see

$$\mathcal{C} = \{ (x, y, s) \in \operatorname{int} \mathcal{F} : \; \psi_n(x, s) = n \log n \} .$$

Central Path for SDP

The central path can be expressed as

$$\mathcal{C} = \left\{ (X, y, S) \in \operatorname{int} \mathcal{F} : \; XS = \frac{X \bullet S}{n} I \right\}$$

in the primal-dual form. We also see

$$\mathcal{C} = \{ (X, y, S) \in \operatorname{int} \mathcal{F} : \; \psi_n(X, S) = n \log n \} .$$

Derive Central Pth from Primal

$$
\begin{array}{lll}
(P) & \text{minimize} & c \bullet x - \mu \sum_{j=1}^{n} \log x_j \\
 & \text{s.t.} & Ax = b,\ x \geq 0.
\end{array}
$$

$$
\begin{array}{lll}
(P) & \text{minimize} & C \bullet S - \mu \log \det(X) \\
 & \text{s.t.} & \mathcal{A}X = b,\ X \succeq 0.
\end{array}
$$

Derive Central Path from Dual

$$
\begin{array}{lll}
(D) & \text{maximize} & b^T y + \mu \sum_{j=1}^{n} \log s_j \\
 & \text{s.t.} & A^T y + s = c, \; s \geq 0.
\end{array}
$$

$$
\begin{array}{lll}
(D) & \text{maximize} & b^T y + \mu \log \det(S) \\
 & \text{s.t.} & \mathcal{A}^T y + S = C, \; S \succeq 0.
\end{array}
$$

The Geometric Intepretation

Let $y^a(z)$ be the analytic center of:

$$\Omega(z) = \{y \in \mathcal{R}^m : c - A^T y \geq 0, \quad b^T y \geq z\},$$

where $z < z^*$. Then

$$\{y^a(z) : \; z^* < z < \infty\}$$

is dual side of the central path for LP.

Let $y^a(z)$ be the analytic center of:

$$\Omega(z) = \{y \in \mathcal{R}^m : C - \mathcal{A}^T y \succeq 0, \quad b^T y \geq z\},$$

where $z < z^*$. Then

$$\{y^a(z) : \; z^* < z < \infty\}$$

is dual side of the central path for SDP.

Central Path Property

Theorem 1 *The central path points* $(x(\mu), y(\mu), s(\mu))$ *and* $(X(\mu), y(\mu), S(\mu))$ *exist, and are bounded and unique.*

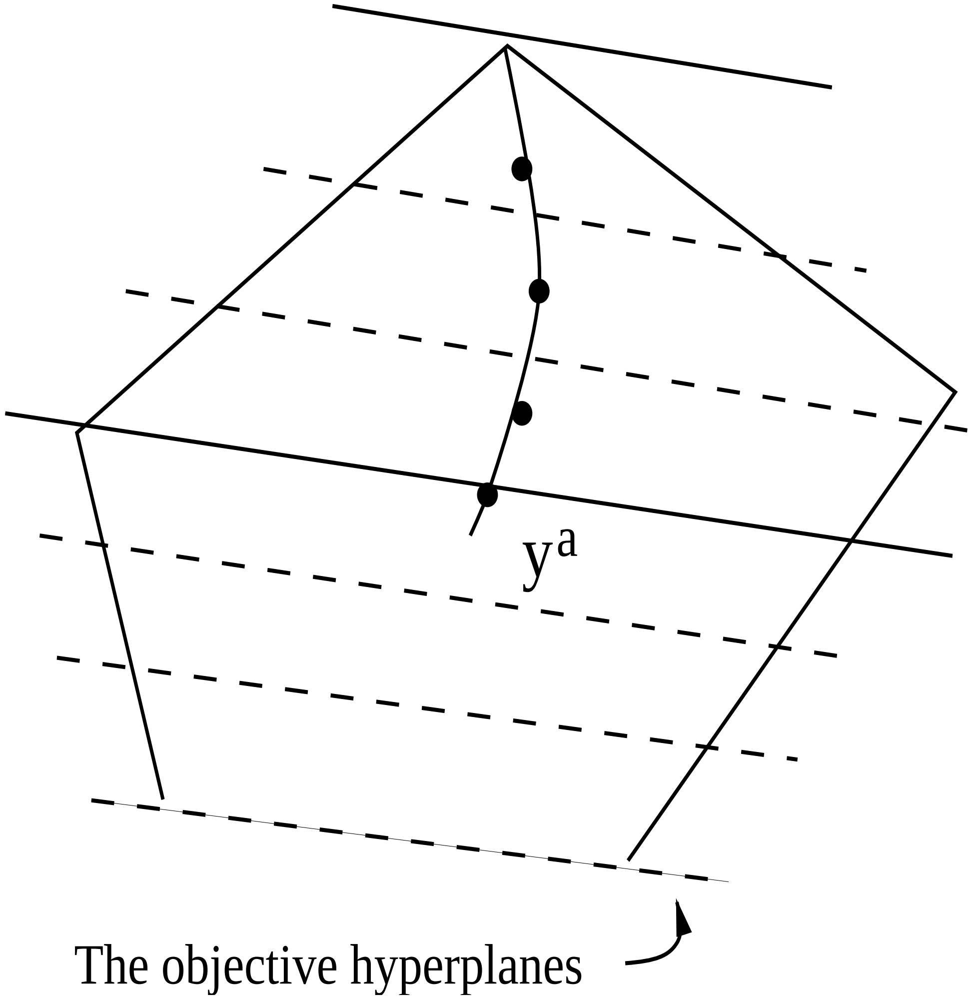

Figure 1: The central path of $y(z)$ in a dual feasible region.

More Properties for LP

Theorem 2 *Let* $(x(\mu), y(\mu), s(\mu))$ *be on the central path.*

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

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

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

if $x(\mu') \neq x(\mu)$ *and* $y(\mu') \neq y(\mu)$.

iii) $(x(\mu), s(\mu))$ *converges to an optimal solution pair for (LP) and (LD). Moreover, the limit point* $x(0)_{P^*}$ *is the analytic center on the primal optimal face, and the limit point* $s(0)_{Z^*}$ *is the analytic center on the dual optimal face, where* (P^*, Z^*) *is the strictly complementarity partition of the index set* $\{1, 2, ..., n\}$.

More Properties for SDP

Corollary 1 *Let* $(X(\mu), y(\mu), S(\mu))$ *be on the central path.*

i) *the central path point* $(X(\mu), S(\mu))$ *is bounded where* $0 < \mu \le \mu^0$ *for any given* $0 < \mu^0 < \infty$*.*

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

$$C \bullet X(\mu') < C \bullet X(\mu) \quad \text{and} \quad b^T y(\mu') > b^T y(\mu),$$

if $X(\mu') \neq X(\mu)$ *and* $y(\mu') \neq y(\mu)$*.*

iii) $(X(\mu), S(\mu))$ *converges to an optimal solution pair for (SDP) and (SDD), and the rank of the limit of* $X(\mu)$ *is maximal among all optimal solutions of (SDP) and the rank of the limit* $S(\mu)$ *is maximal among all optimal solutions of (SDD).*

Important Lemmas

Homework 7:

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

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

Homework 8:

Lemma 2 *Let* $X \in \mathcal{M}^n$ *and* $\|X\|_\infty < 1$*. Then,*

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

$$\operatorname{int} \mathcal{F}_p = \{x : \; Ax = b, \; x > 0\} \neq \emptyset$$

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

Let z^* denote the optimal value and

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

Here, we are interested in finding an ϵ-approximate solution for the LP problem:

$$c^T x - z^* \leq \epsilon \quad \text{and} \quad z^* - b^T y \leq \epsilon.$$

For simplicity, we assume that a central path pair (x^0, y^0, s^0) with $\mu^0 = (x^0)^T s^0 / n$ is known. We will use it as our initial point throughout this chapter.

Primal-Dual (Symmetric) Algorithm for LP

Once we have a pair $(x, y, s) \in \text{int}\,\mathcal{F}$ with $\mu = x^T s/n$, we can generate a new iterate x^+ and (y^+, s^+) by solving for d_x, d_y and d_s from the system of linear equations:

$$\begin{array}{rcl} Sd_x + Xd_s & = & r := \gamma\mu e - Xs, \\ Ad_x & = & 0, \\ -A^T d_y - d_s & = & 0. \end{array} \qquad (6)$$

Let $d := (d_x, d_y, d_s)$. To show the dependence of d on the current pair (x, s) and the parameter γ, we write $d = d(x, s, \gamma)$. Note that $d_x^T d_s = -d_x^T A^T d_y = 0$ here.

$$
\begin{aligned}
d_{x'} + d_{s'} &= r' := (XS)^{-1/2}(\gamma\mu e - Xs), \\
A' d_{x'} &= 0, \\
-(A')^T d_y - d_{s'} &= 0.
\end{aligned}
$$

where

$$D = X^{1/2}X^{-1/2},\ A' = AD,\ d_{x'} = D^{-1}d_x,\ d_{s'} = Dd_s.$$

$$-A'(A')^T d_y = A'r'.$$

If $\gamma = 0$, it steps toward the optimal solution characterized by the optimality condition; if $\gamma = 1$, it steps toward the central path point $(x(\mu), y(\mu), s(\mu))$ characterized by the analytic center condition; if $0 < \gamma < 1$, it steps toward a central path point with a smaller complementarity gap. In the algorithm presented in this section, we choose $\gamma = n/(n + \rho) < 1$. Each iterate reduces the primal-dual potential function by at least a constant δ.

Lemma 3 *Let the direction* $d = (d_x, d_y, d_s)$ *be generated by equation (6) with* $\gamma = n/(n+\rho)$*, and let*

$$\theta = \frac{\alpha\sqrt{\min(Xs)}}{\|(XS)^{-1/2}(\frac{x^Ts}{(n+\rho)}e - Xs)\|}\,, \tag{7}$$

where α *is a positive constant less than 1. Let*

$$x^+ = x + \theta d_x, \quad y^+ = y + \theta d_y, \quad \textit{and} \quad s^+ = s + \theta d_s.$$

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

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

$$\leq -\alpha\sqrt{\min(Xs)}\|(XS)^{-1/2}(e - \frac{(n+\rho)}{x^Ts}Xs)\| + \frac{\alpha^2}{2(1-\alpha)}\,.$$

Let $v = Xs$. Then, we can prove the following lemma Homework 9:

Lemma 4 *Let $v \in \mathcal{R}^n$ be a positive vector and $\rho \geq \sqrt{n}$. Then,*

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

Combining these two lemmas we have

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

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

for a constant δ.

Description of Algorithm

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

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

1. Set $(x, s) = (x^k, s^k)$ and $\gamma = n/(n+\rho)$ and compute (d_x, d_y, d_s) from (6).
2. Let $x^{k+1} = x^k + \bar{\alpha} d_x$, $y^{k+1} = y^k + \bar{\alpha} d_y$, 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.

Theorem 3 *Let* $\rho = O(\sqrt{n})$*. Then, the Algorithm terminates in at most* $O(\sqrt{n}\log((x^0)^T s^0/\epsilon))$ *iterations with*

$$c^T x^k - b^T y^k \leq \epsilon.$$

Primal-Dual (Symmetric) Algorithm for SDP

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

$$\begin{aligned} D^{-1} d_X D^{-1} + d_S &= R := \gamma\mu X^{-1} - S, \\ \mathcal{A} d_X &= 0, \\ -\mathcal{A}^T d_y - d_S &= 0, \end{aligned} \tag{8}$$

where

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

Note that $d_S \bullet d_X = 0$.

Primal-Dual Scaling

$$\begin{aligned} d_{X'} + d_{S'} &= R', \\ \mathcal{A}' d_{X'} &= 0, \\ -\mathcal{A}'^T d_y - d_{S'} &= 0, \end{aligned} \tag{9}$$

where

$$d_{X'} = D^{-.5} d_X D^{-.5}, \quad d_{S'} = D^{.5} d_S D^{.5}, \quad R' = D^{.5}(\gamma\mu X^{-1} - S)D^{.5},$$

and

$$\mathcal{A}' = \begin{pmatrix} A'_1 \\ A'_2 \\ \dots \\ A'_m \end{pmatrix} := \begin{pmatrix} D^{.5} A_1 D^{.5} \\ D^{.5} A_2 D^{.5} \\ \dots \\ D^{.5} A_m D^{.5} \end{pmatrix}.$$

Again, we have $d_{S'} \bullet d_{X'} = 0$, and

$$d_y = (\mathcal{A}'\mathcal{A}'^T)^{-1}\mathcal{A}'R', \; d_{S'} = -\mathcal{A}'^T d_y, \text{ and } d_{X'} = R' - d_{S'}.$$

Or, we have

$$d_S = -\mathcal{A}^T d_y \quad \text{and} \quad d_X = D(R - d_S)D.$$

The Bound on Potential Reduction

$$V^{1/2} = D^{-.5} X D^{-.5} = D^{.5} S D^{.5} \in \operatorname{int} \mathcal{M}_+^n.$$

Then, we can verify that $S \bullet X = I \bullet V$.

Lemma 5 *Let the direction d_X, d_y and d_S be generated by equation (8) with $\gamma = n/(n+\rho)$, and let*

$$\theta = \frac{\alpha}{\|V^{-1/2}\|_\infty \|\frac{I \bullet V}{n+\rho} V^{-1/2} - V^{1/2}\|} , \tag{10}$$

where α is a positive constant less than 1. Let

$$X^+ = X + \theta d_X, \quad y^+ = y + \theta d_y, \quad \textit{and} \quad S^+ = S + \theta d_S.$$

Then, we have $(X^+, y^+, S^+) \in \operatorname{int} \mathcal{F}$ and

$$\psi(X^+, S^+) - \psi(X, S)$$

$$\leq -\alpha \frac{\|V^{-1/2} - \frac{n+\rho}{I \bullet V} V^{1/2}\|}{\|V^{-1/2}\|_\infty} + \frac{\alpha^2}{2(1-\alpha)} .$$

A Technical Lemma and the Convergence

Homework 10:

Lemma 6 *Let* $V \in \operatorname{int} \mathcal{M}_+^n$ *and* $\rho \geq \sqrt{n}$*. Then,*

$$\frac{\|V^{-1/2} - \frac{n+\rho}{I \bullet V} V^{1/2}\|}{\|V^{-1/2}\|_\infty} \geq \sqrt{3/4}.$$

From the two lemmas we have

$$\psi(X^+, S^+) - \psi(X, S)$$

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

for a constant δ.

Description of Algorithm

Given $(X^0, y^0, S^0) \in \text{int}\, \mathcal{F}$. Set $\rho = \sqrt{n}$ 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, d_y, d_S) from (8).

2. Let $X^{k+1} = X^k + \bar{\alpha} d_X$, $y^{k+1} = y^k + \bar{\alpha} d_y$, and $S^{k+1} = S^k + \bar{\alpha} d_S$, where

$$\bar{\alpha} = \arg\min_{\alpha \geq 0} \psi(X^k + \alpha d_X, S^k + \alpha d_S).$$

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

Complexity of the Algorithm

Corollary 2 *Let $\rho = \sqrt{n}$. Then, the Algorithm terminates in at most $O(\sqrt{n}\log(C \bullet X^0 - b^T y^0)/\epsilon)$ iterations with*

$$C \bullet X^k - b^T y^k \leq \epsilon.$$

Dual Scaling Algorithm for SDP

An open question is how to exploit the sparsity structure by polynomial interior-point algorithms so that they can also solve large-scale problems in practice.

1. The computational cost of each iteration in the dual algorithm is less that the cost the primal-dual iterations.

2. In most combinatorial applications, we need only a lower bound for the optimal objective value of (SDP).

3. For large scale problems, S tends to be very sparse and structured since it is the linear combination of C and the A_i's. This sparsity allows considerable savings in both memory and computation time.

Dual Algorithm

$$\phi_{n+\rho}(X, S) = \rho \ln(X \bullet S) - \ln\det X - \ln\det S.$$

Let $\bar{z} = C \bullet X$ for some feasible X and consider the dual potential function

$$\psi(y, \bar{z}) = \rho \ln(\bar{z} - b^T y) - \ln\det S.$$

Its gradient is

$$\nabla\psi(y, \bar{z}) = -\frac{\rho}{\bar{z} - b^T y} b + \mathcal{A}(S^{-1}). \tag{11}$$

Overestimator of Potential

For any given y and $S = C - \mathcal{A}^T(y)$ such that $S \succ 0$ and

$$\|(S^k)^{-.5}\left(\mathcal{A}^T(y-y^k)\right)(S^k)^{-.5}\| < 1,$$

$$\begin{aligned}
&\psi(y,\bar{z}^k) - \psi(y^k,\bar{z}^k) \\
&= \rho\ln(\bar{z}^k - b^Ty) - \rho\ln(\bar{z}^k - b^Ty^k) - \ln\det((S^k)^{-.5}S(S^k)^{-.5}) \\
&\le \rho\ln(\bar{z}^k - b^Ty) - \rho\ln(\bar{z}^k - b^Ty^k) + \mathrm{trace}((S^k)^{-.5}S(S^k)^{-.5} - I) \\
&\quad + \tfrac{\|(S^k)^{-.5}\left(\mathcal{A}^T(y-y^k)\right)(S^k)^{-.5}\|}{2(1-\|(S^k)^{-.5}(\mathcal{A}^T(y-y^k))(S^k)^{-.5}\|_\infty)} \\
&= \rho\ln(\bar{z}^k - b^Ty) - \rho\ln(\bar{z}^k - b^Ty^k) + \mathcal{A}((S^k)^{-1})^T(y-y^k) \\
&\quad + \tfrac{\|(S^k)^{-.5}\left(\mathcal{A}^T(y-y^k)\right)(S^k)^{-.5}\|}{2(1-\|(S^k)^{-.5}(\mathcal{A}^T(y-y^k))(S^k)^{-.5}\|_\infty)} \\
&\le \nabla\psi(y^k,\bar{z}^k)^T(y-y^k) + \tfrac{\|(S^k)^{-.5}\left(\mathcal{A}^T(y-y^k)\right)(S^k)^{-.5}\|}{2(1-\|(S^k)^{-.5}(\mathcal{A}^T(y-y^k))(S^k)^{-.5}\|_\infty)}.
\end{aligned} \tag{12}$$

Ball constrainted sub-problem

$$
\begin{array}{ll}
\text{Minimize} & \nabla\psi^T(y^k,\bar{z}^k)(y-y^k) \\
\text{Subject to} & \|(S^k)^{-.5}\left(\mathcal{A}^T(y-y^k)\right)(S^k)^{-.5}\| \le \alpha,
\end{array}
\tag{13}
$$

where α is a positive constant less than 1. For simplicity, in what follows we let

$$\Delta^k = \bar{z}^k - b^T y^k.$$

Optimality conditions of the sub-problem

The first order optimality conditions state that the minimum point, y^{k+1}, of this convex problem satisfies

$$M^k(y^{k+1} - y^k) + \beta\nabla\psi(y^k, \bar{z}^k) = 0 \tag{14}$$

for a positive value of β, where

$$M^k = \begin{pmatrix} A_1(S^k)^{-1} \bullet (S^k)^{-1}A_1 & \cdots & A_1(S^k)^{-1} \bullet (S^k)^{-1}A_m \\ \vdots & \ddots & \vdots \\ A_m(S^k)^{-1} \bullet (S^k)^{-1}A_1 & \cdots & A_m(S^k)^{-1} \bullet (S^k)^{-1}A_m \end{pmatrix}$$

The matrix M^k is a Gram matrix and is positive definite when $S^k \succ 0$ and the A_i's are linearly independent.

Optimizer of the sub-problem

Using the ellipsoidal constraint, the minimal solution, y^{k+1}, of (13) is given by

$$y^{k+1} - y^k = \frac{\alpha}{\sqrt{\nabla\psi^T(y^k, \bar{z}^k)(M^k)^{-1}\nabla\psi(y^k, \bar{z}^k)}} d(\bar{z}^k)_y \qquad (15)$$

where

$$d(\bar{z}^k)_y = -(M^k)^{-1}\nabla\psi(y^k, \bar{z}^k). \qquad (16)$$

Generating M

Generally, $M^k_{ij} = A_i (S^k)^{-1} \bullet (S^k)^{-1} A_j$. When $A_i = a_i a_i^T$, the Gram matrix can be rewritten in the form

$$M^k = \begin{pmatrix} (a_1^T (S^k)^{-1} a_1)^2 & \cdots & (a_1^T (S^k)^{-1} a_m)^2 \\ \vdots & \ddots & \vdots \\ (a_m^T (S^k)^{-1} a_1)^2 & \cdots & (a_m^T (S^k)^{-1} a_m)^2 \end{pmatrix} \tag{17}$$

and

$$\mathcal{A}((S^k)^{-1}) = \begin{pmatrix} a_1^T (S^k)^{-1} a_1 \\ \vdots \\ a_m^T (S^k)^{-1} a_m \end{pmatrix}.$$

This matrix can be computed very quickly without computing, or saving, $(S^k)^{-1}$.

Potential Reduction

$$\nabla\psi^T(y^k, \bar{z}^k)d(\bar{z}^k)_y = -\|P(\bar{z}^k)\|^2 \tag{18}$$

$$\psi(y^{k+1}, \bar{z}^k) - \psi(y^k, \bar{z}^k) \leq -\alpha\|P(\bar{z}^k)\| + \frac{\alpha^2}{2(1-\alpha)}. \tag{19}$$

Primal Update

To find a feasible primal point X, we solve the least squares problem

$$\begin{array}{ll} \text{Minimize} & \|(S^k)^{.5} X (S^k)^{.5} - \frac{\Delta^k}{\rho} I\| \\ \text{Subject to} & \mathcal{A}(X) = b. \end{array} \tag{20}$$

The answer to (20) is a by-product of computing (16), given explicitly by

$$X(\bar{z}^k) = \frac{\Delta^k}{\rho} (S^k)^{-1} \left(\mathcal{A}^T (d(\bar{z}^k)_y) + S^k \right) (S^k)^{-1}. \tag{21}$$

Primal Objective Value

$$
\begin{aligned}
C \bullet X(\bar{z}^k) &= b^T y^k + X(\bar{z}^k) \bullet S^k \\
&= b^T y^k + \operatorname{trace}\left(\frac{\Delta^k}{\rho} (S^k)^{-1} \left(\mathcal{A}^T (d(\bar{z}^k)_y) + S^k \right) (S^k)^{-1} S^k \right) \\
&= b^T y^k + \frac{\Delta^k}{\rho} \operatorname{trace}\left((S^k)^{-1} \mathcal{A}^T (d(\bar{z}^k)_y) + I \right) \\
&= b^T y^k + \frac{\Delta^k}{\rho} \left(d(\bar{z}^k)_y^T \mathcal{A}((S^k)^{-1}) + n \right)
\end{aligned}
$$

Since the vectors $\mathcal{A}((S^k)^{-1})$ and $d(\bar{z}^k)_y$ were previously found in calculating the dual step direction, the cost of computing a primal objective value is the cost of a vector dot product!

Result for Primal

Defining

$$P(\bar{z}^k) = \frac{\rho}{\Delta^k}(S^k)^{.5}X(\bar{z}^k)(S^k)^{.5} - I, \tag{22}$$

we have the following lemma:

Lemma 7 *Let* $\mu^k = \frac{\Delta^k}{n} = \frac{\bar{z}^k - b^T y^k}{n}$, $\mu = \frac{X(\bar{z}^k)\bullet S^k}{n} = \frac{C\bullet X(\bar{z}^k) - b^T y^k}{n}$, $\rho \geq n + \sqrt{n}$, *and* $\alpha < 1$. *If*

$$\|P(\bar{z}^k)\| < min(\alpha\sqrt{\frac{n}{n+\alpha^2}}, 1-\alpha), \tag{23}$$

then the following three inequalities hold:

1. $X(\bar{z}^k) \succ 0$;
2. $\|(S^k)^{.5}X(\bar{z}^k)(S^k)^{.5} - \mu I\| \leq \alpha\mu$;
3. $\mu \leq (1 - .5\alpha/\sqrt{n})\mu^k$.

Theorem 4 *Either*

$$\psi(X^k, S^{k+1}) \leq \psi(X^k, S^k) - \delta$$

or

$$\psi(X^{k+1}, S^k) \leq \psi(X^k, S^k) - \delta,$$

where $\delta > 1/20$*.*

Description of Algorithm

DUAL ALGORITHM. Given an upper bound $\bar{z}^0$ and a dual point (y^0, S^0) such that $S^0 = C - \mathcal{A}^T y^0 \succ 0$, set $k = 0$, $\rho > n + \sqrt{n}$, $\alpha \in (0, 1)$, and do the following:

while $\bar{z}^k - b^T y^k \geq \epsilon$ **do**

begin

1. Compute $\mathcal{A}((S^k)^{-1})$ and the Gram matrix M^k (17) using Algorithm M or M'.
2. Solve (16) for the dual step direction $d(\bar{z}^k)_y$.
3. Calculate $\|P(\bar{z}^k)\|$ using (18).
4. **If** (23) is true, **then** $X^{k+1} = X(\bar{z}^k)$, $\bar{z}^{k+1} = C \bullet X^{k+1}$, and $(y^{k+1}, S^{k+1}) = (y^k, S^k)$;
   else $y^{k+1} = y^k + \frac{\alpha}{\|P(\bar{z}^k)\|} d(\bar{z}^{k+1})_y$, $S^{k+1} = C - \mathcal{A}^T(y^{k+1})$,

$X^{k+1} = X^k$, and $\bar{z}^{k+1} = \bar{z}^k$.

endif

5. $k := k + 1$.

end

Complexity of the Algorithm

Corollary 3 *Let $\rho = \sqrt{n}$. Then, the Algorithm terminates in at most $O(\sqrt{n}\log(C \bullet X^0 - b^T y^0)/\epsilon)$ iterations with*

$$C \bullet X^k - b^T y^k \le \epsilon.$$

Project: Develop simplified computation procedures of the dual scaling SDP algorithm when each A_i is rank-one, that is, $A_i = a_i a_i^T$, for $i = 1, ..., m$.

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 $O(n\log(R/\epsilon))$.

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.

A HSD linear program

Given any $\mathbf{x}^0 = \mathbf{e} > \mathbf{0}$, $\mathbf{s}^0 = \mathbf{e} > \mathbf{0}$, and $\mathbf{y}^0 = \mathbf{0}$, we formulate

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

where

$$\bar{\mathbf{b}} = \mathbf{b} - A\mathbf{e}, \quad \bar{\mathbf{c}} = \mathbf{c} - \mathbf{e}, \quad \bar{z} = \mathbf{c}^T\mathbf{e} + 1. \tag{24}$$

Denote by $\mathbf{s}$ the slack vector for the second constraint and by κ the slack scalar for the third constraint. Denote by $\mathcal{F}_h$ the set of all points $(\mathbf{y}, \mathbf{x}, \tau, \theta, \mathbf{s}, \kappa)$ that are feasible for (HSDP). Denote by $\mathcal{F}_h^0$ the set of strictly feasible points with $(\mathbf{x}, \tau, \mathbf{s}, \kappa) > \mathbf{0}$ in $\mathcal{F}_h$. Note that by combining the constraints, we can write the last (equality) constraint as

$$\mathbf{e}^T x + \mathbf{e}^T s + \tau + \kappa - (n+1)\theta = (n+1), \tag{25}$$

which serves as a normalizing constraint for (HSDP). Also note that the constraints of (HSDP) form a skew-symmetric system, so that it is a self-dual linear program.

Theorem 5 *Consider problems (HSDP) and (HSDD).*

i) *(HSDD) has the same form as (HSDP), i.e., (HSDD) is simply (HSDP) with* $(\mathbf{y}, \mathbf{x}, \tau, \theta)$ *being replaced by* $(\mathbf{y}', \mathbf{x}', \tau', \theta')$.

ii) *(HSDP) has a strictly feasible point*

$$\mathbf{y} = \mathbf{y}^0, \quad x = x^0 > \mathbf{0}, \quad \tau = 1, \quad \theta = 1, \quad \mathbf{s} = \mathbf{s}^0 > \mathbf{0}, \quad \kappa = 1.$$

iii) *(HSDP) has an optimal solution and its optimal solution set is bounded.*

iv) *The optimal value of (HSDP) is zero, and*

$$(\mathbf{y}, \mathbf{x}, \tau, \theta, \mathbf{s}, \kappa) \in \mathcal{F}_h \quad \textit{implies that} \quad (n+1)\theta = \mathbf{x}^T\mathbf{s} + \tau\kappa.$$

v) *There is an optimal solution* $(\mathbf{y}^*, \mathbf{x}^*, \tau^*, \theta^* = 0, \mathbf{s}^*, \kappa^*) \in \mathcal{F}_h$ *such that*

$$\begin{pmatrix} \mathbf{x}^* + \mathbf{s}^* \\ \tau^* + \kappa^* \end{pmatrix} > \mathbf{0},$$

which we call a strictly self-complementary solution. (Similarly, we sometimes call an optimal solution to (HSDP) a self-complementary solution; the strict inequalities above need not hold.)

Homework 11:

Theorem 6 *Let* $(\mathbf{y}^*, \mathbf{x}^*, \tau^*, \theta^* = 0, \mathbf{s}^*, \kappa^*)$ *be a strictly self complementary solution for (HSDP).*

i) *(LP) has a solution (feasible and bounded) if and only if* $\tau^* > 0$. *In this case,* $\mathbf{x}^*/\tau^*$ *is an optimal solution for (LP) and* $(\mathbf{y}^*/\tau^*, \mathbf{s}^*/\tau^*)$ *is an optimal solution for (LD).*

ii) *(LP) has no solution if and only if* $\kappa^* > 0$. *In this case,* $\mathbf{x}^*/\kappa^*$ *or* $\mathbf{s}^*/\kappa^*$ *or both are certificates for proving infeasibility: if* $\mathbf{c}^T\mathbf{x}^* < 0$ *then (LD) is infeasible; if* $-\mathbf{b}^T\mathbf{y}^* < 0$ *then (LP) is infeasible; and if both* $\mathbf{c}^T\mathbf{x}^* < 0$ *and* $-\mathbf{b}^T\mathbf{y}^* < 0$ *then both (LP) and (LD) are infeasible.*

Theorem 7 **i)** *For any* $\mu > 0$, *there is a unique* $(\mathbf{y}, \mathbf{x}, \tau, \theta, \mathbf{s}, \kappa)$ *in* $\mathcal{F}_h^0$, *such that*

$$\begin{pmatrix} X\mathbf{s} \\ \tau\kappa \end{pmatrix} = \mu\mathbf{e}.$$

ii) *Let* $(\mathbf{d}_y, \mathbf{d}_x, d_\tau, d_\theta, \mathbf{d}_s, d_\kappa)$ *be in the null space of the constraint matrix of (HSDP) after adding surplus variables* $\mathbf{s}$ *and* κ, *i.e.,*

$$\begin{array}{rrrrrrcl} & A\mathbf{d}_x & -\mathbf{b}d_\tau & +\bar{\mathbf{b}}d_\theta & & & = & \mathbf{0}, \\ -A^T\mathbf{d}_y & & +\mathbf{c}d_\tau & -\bar{\mathbf{c}}d_\theta & -\mathbf{d}_s & & = & \mathbf{0}, \\ \mathbf{b}^T\mathbf{d}_y & -\mathbf{c}^T\mathbf{d}_x & & +\bar{z}d_\theta & & -d_\kappa & = & 0, \\ -\bar{\mathbf{b}}^T\mathbf{d}_y & +\bar{\mathbf{c}}^T\mathbf{d}_x & -\bar{z}d_\tau & & & & = & 0. \end{array} \qquad (26)$$

$$(\mathbf{d}_x)^T\mathbf{d}_s + d_\tau d_\kappa = 0.$$

Central Path

We see that Theorem 7 defines an endogenous path within (HSDP):

$$\mathcal{C} = \left\{ (\mathbf{y}, \mathbf{x}, \tau, \theta, \mathbf{s}, \kappa) \in \mathcal{F}_h^0 : \begin{pmatrix} X\mathbf{s} \\ \tau\kappa \end{pmatrix} = \frac{\mathbf{x}^T\mathbf{s} + \tau\kappa}{n+1}\mathbf{e} \right\},$$

which we may call the (self-)*central path* for (HSDP). Obviously, the initial interior feasible point proposed in Theorem 5 is on the path with $\mu = 1$.

Solving (HSDP)

Consider solving the following system of linear equations for $(\mathbf{d}_y, \mathbf{d}_x, d_\tau, d_\theta, \mathbf{d}_s, d_\kappa)$ that satisfies (26) and

$$\begin{pmatrix} X\mathbf{d}_s + S\mathbf{d}_x \\ \tau^k d_\kappa + \kappa^k d_\tau \end{pmatrix} = \gamma\mu\mathbf{e} - \begin{pmatrix} X\mathbf{s} \\ \tau\kappa \end{pmatrix}.$$

Theorem 8 *The* $O(\sqrt{n}\log((\mathbf{x}^0)^T\mathbf{s}^0/\epsilon))$ *interior-point algorithm, coupled with a termination technique described above, generates a strictly self-complementary solution for (HSDP) in* $O(\sqrt{n}(\log(c(A,\mathbf{b},\mathbf{c}))+\log n))$ *iterations and* $O(n^3(\log(c(A,\mathbf{b},\mathbf{c}))+\log n))$ *operations, where* $c(A,\mathbf{b},\mathbf{c})$ *is a positive number depending on the data* $(A,\mathbf{b},\mathbf{c})$*. If (LP) and (LD) have integer data with bit length* L*, then by the construction, the data of (HSDP) remains integral and its length is* $O(L)$*. Moreover,* $c(A,\mathbf{b},\mathbf{c})\le 2^L$*. Thus, the algorithm terminates in* $O(\sqrt{n}L)$ *iterations and* $O(n^3L)$ *operations.*

example

Consider the example where

$$A = \begin{pmatrix} -1 & 0 & 0 \end{pmatrix}, \quad b = 1, \quad \text{and} \quad \mathbf{c} = \begin{pmatrix} 0 & 1 & -1 \end{pmatrix}.$$

Then,

$$y^* = 2, \quad \mathbf{x}^* = (0,2,1)^T, \quad \tau^* = 0, \quad \theta^* = 0, \quad \mathbf{s}^* = (2,0,0)^T, \quad \kappa^* = 1$$

could be a strictly self-complementary solution generated for (HSDP) with

$$\mathbf{c}^T\mathbf{x}^* = 1 > 0, \quad by^* = 2 > 0.$$

Thus $(y^*, \mathbf{s}^*)$ demonstrates the infeasibility of (LP), but $\mathbf{x}^*$ doesn't show the infeasibility of (LD). Of course, if the algorithm generates instead $\mathbf{x}^* = (0,1,2)^T$, then we get demonstrated infeasibility of both.

Homework 12: Develop the homogeneous and self-dual model for solving SDP problems.

Project continued: Down load Sedumi and DSDP5.8, and read the user-guides.