Interior-Point SDP Algorithm

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

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

Methodological Philosophy

Interior-point methods 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 interior-point methods is to stay in the interior of the feasible region.

Interior-Point Algorithms for LP

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

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

and

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

We are interested in finding an ϵ-approximate solution for the LP problem: $\mathbf{c}^T\mathbf{x} - \mathbf{b}^T\mathbf{y} \leq \epsilon$. 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.

Interior-Point Algorithms for SDP

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

$$\text{int}\,\mathcal{F}_d = \{(\mathbf{y}, S):\ S = C - \mathcal{A}^T\mathbf{y} \succ \mathbf{0}\} \neq \emptyset,$$

and

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

where

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

We are interested in finding an ϵ-approximate solution for the SDP problem: $C \bullet X - \mathbf{b}^T\mathbf{y} \leq \epsilon$. For simplicity, we assume that an interior-point pair $(X^0, \mathbf{y}^0, S^0)$ is known, and we will use it as our initial point pair.

Barrier Functions for LP

Consider the barrier function optimization

$$\begin{array}{ll} \text{minimize} & -\sum_{j=1}^{n} \log x_j \\ \text{s.t.} & \mathbf{x} \in \operatorname{int} \mathcal{F}_p \end{array}$$

and

$$\begin{array}{ll} \text{maximize} & \sum_{j=1}^{n} \log s_j \\ \text{s.t.} & (\mathbf{y}, \mathbf{s}) \in \operatorname{int} \mathcal{F}_d \end{array}$$

They are linearly constrained convex programs (LCCP).

Barrier Function for SDP

Consider the barrier function optimization

$$\begin{array}{ll} \text{minimize} & -\log\det(X) \\ \text{s.t.} & X \in \operatorname{int}\mathcal{F}_p \end{array}$$

and

$$\begin{array}{ll} \text{maximize} & \log\det(S) \\ \text{s.t.} & (\mathbf{y}, S) \in \operatorname{int}\mathcal{F}_d \end{array}$$

They are also linearly constrained convex programs (LCCP).

LP with Barrier Function

Consider the LP problem with the barrier function

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

and

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

where μ is called the barrier (weight) parameter.

They are again linearly constrained convex programs (LCCP).

SDP with Barrier Function

Consider the LP problem with the barrier function

$$(SDPB) \quad \text{minimize} \quad C \bullet X - \mu \log\det(X)$$
$$\text{s.t.} \quad X \in \operatorname{int} \mathcal{F}_p$$

and

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

where μ is called the barrier (weight) parameter.

They are again linearly constrained convex programs (LCCP).

Common Optimality Conditions for LPB and LDB

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

where

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

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

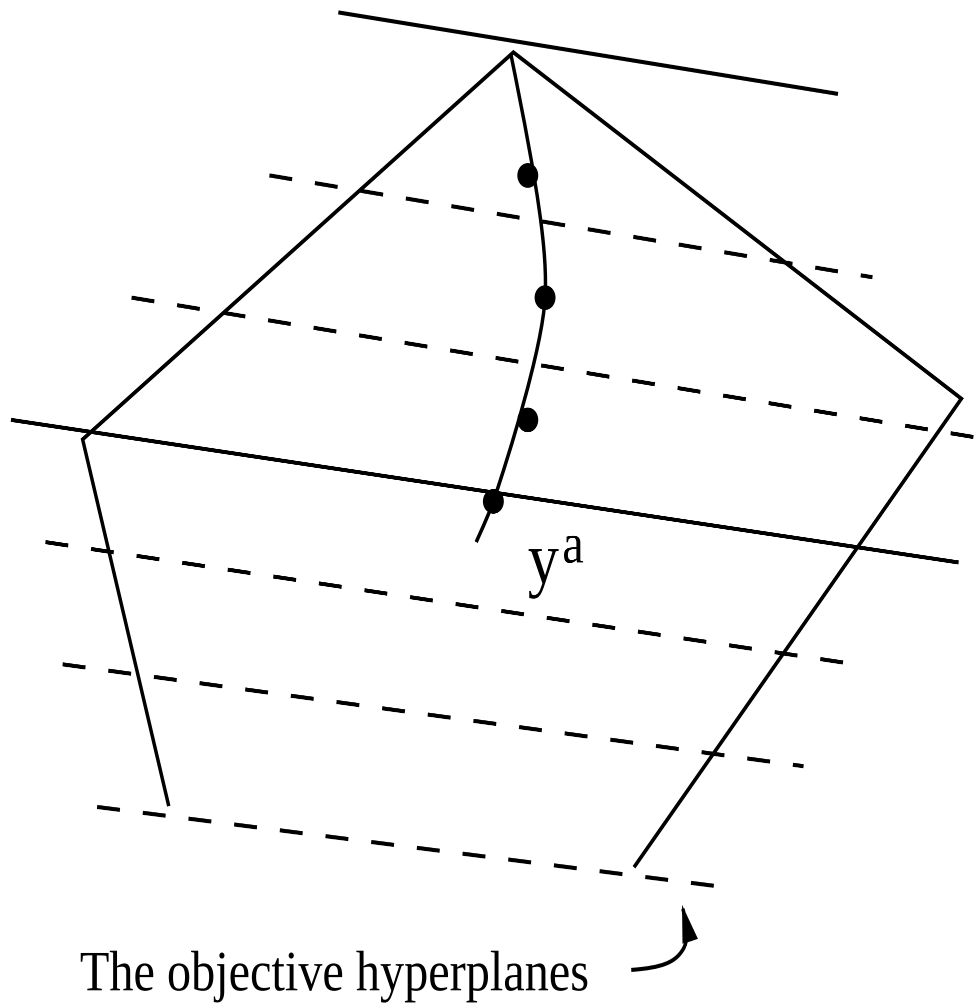

Figure 1: The central path of $\mathbf{y}(\mu)$ in a dual feasible region.

Common Optimality Conditions for SDPB and SDDB

$$\begin{aligned} XS &= \mu I \\ \mathcal{A}X &= \mathbf{b} \\ -\mathcal{A}^T\mathbf{y} - S &= -C; \end{aligned}$$

where we have

$$\mu = \frac{X \bullet S}{n} = \frac{C \bullet X - \mathbf{b}^T\mathbf{y}}{n},$$

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

Central Path for Linear Programming

The path

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

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 $(A, \mathbf{b}, c)$. Then for any $0 < \mu < \infty$, the central path point pair $(\mathbf{x}(\mu), \mathbf{y}(\mu), \mathbf{s}(\mu))$ exists and is unique.*

Central Path for SDP

The path

$$\mathcal{C} = \{(X(\mu), \mathbf{y}(\mu), S(\mu)) \in \operatorname{int} \mathcal{F} : \ XS = \mu I, \ 0 < \mu < \infty\};$$

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

Theorem 2 *Let both (SDP) and (SDD) have interior feasible points for the given data set $(\mathcal{A}, \mathbf{b}, C)$. Then for any $0 < \mu < \infty$, the central path point pair $(X(\mu), \mathbf{y}(\mu), S(\mu))$ exists and is unique.*

Central Path Properties for LP

Theorem 3 *Let $(\mathbf{x}(\mu), \mathbf{y}(\mu), \mathbf{s}(\mu))$ be on the central path of an linear program in standard form.*

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 nontrivial optimal solutions.

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 $\mathbf{s}(0)_{Z^*} > \mathbf{0}$, where (P^*, Z^*) is the strictly complementarity partition of the index set $\{1, 2, ..., n\}$.*

Central Path Properties for SDP

Theorem 4 *Let* $(X(\mu), \mathbf{y}(\mu), S(\mu))$ *be on the central path of an SDP in standard form.*

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* $0 < \mu' < \mu$*,*

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

if both primal and dual have nontrivial optimal solutions.

iii) $(X(\mu), S(\mu))$ *converges to an optimal solution pair for (SDP) and (SDD). Moreover, the limit point is a maximal rank complementarity solution pair.*

Proof Sketch

Let X^* and S^* be max-rank optimal solutions for the primal and dual respectively. Then from

$$(X(\mu) - X^*) \bullet (S(\mu) - S^*) = 0$$

we have

$$X(\mu) \bullet S^* + S(\mu) \bullet X^* = n\mu$$

which further implies

$$S(\mu)^{-1} \bullet S^* + X(\mu)^{-1} \bullet X^* = n.$$

Thus,

$$X(\mu)^{-1} \bullet X^* \leq n$$

or

$$X(\mu)^{-1/2} X^* X(\mu)^{-1/2} \bullet I \leq n.$$

Thus, all eigenvalues of $X(\mu)^{-1/2}X^*X(\mu)^{-1/2}$ must be bounded above by n or

$$n \cdot I \succeq X(\mu)^{-1/2}X^*X(\mu)^{-1/2}$$

or

$$X(\mu) \succeq \frac{1}{n}X^*.$$

Potential and Duality Gap in LP

For $x \in \operatorname{int}\mathcal{F}_p$ and $(y, s) \in \operatorname{int}\mathcal{F}_d$, let parameter $\rho > 0$ and

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

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

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

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

Potential and Duality Gap in SDP

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

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

The Potential Reduction Algorithm

The potential reduction algorithm generates a sequence of $\{X^k, y^k, S^k\} \in \text{int}\,\mathcal{F}$ such that

$$\psi_{n+\sqrt{n}}(X^{k+1}, S^{k+1}) \leq \psi_{n+\sqrt{n}}(X^k, S^k) - .05$$

for $k = 0, 1, 2, ...$.

This indicates that the potential level set shrinks at a constant rate independently of m or n, which leads to the duality gap converging toward zero.

Primal-Dual Potential Reduction 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:

$$\begin{aligned} D^{-1} D_X D^{-1} + D_S &= R := \frac{n}{n+\rho}\mu X^{-1} - S, \\ \mathcal{A} D_X &= \mathbf{0}, \\ -\mathcal{A}^T \mathbf{d}_y - D_S &= \mathbf{0}, \end{aligned} \tag{1}$$

where

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

Note that $D_S \bullet D_X = 0$.

Primal-Dual Scaling

$$\begin{aligned} D_{X'} + D_{S'} &= R', \\ \mathcal{A}' D_{X'} &= \mathbf{0}, \\ -\mathcal{A}'^T \mathbf{d}_y - D_{S'} &= \mathbf{0}, \end{aligned} \tag{2}$$

where

$$D_{X'} = D^{-.5} D_X D^{-.5},\ D_{S'} = D^{.5} D_S D^{.5},\ R' = D^{.5} \left(\frac{n}{n+\rho} \mu X^{-1} - S \right) D^{.5},$$

and

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

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

$$\mathbf{d}_y = (\mathcal{A}'\mathcal{A}'^T)^{-1}\mathcal{A}'R',\ D_{S'} = -\mathcal{A}'^T\mathbf{d}_y, \text{ and } D_{X'} = R' - D_{S'}.$$

Or, we have

$$D_S = -\mathcal{A}^T\mathbf{d}_y \quad \text{and} \quad D_X = D(R - D_S)D.$$

The role of ρ

If $\rho = \infty$, it steps toward the optimal solution characterized by the SDP optimality condition; if $\rho = 0$, it steps toward the central path point $(X(\mu), \mathbf{y}(\mu), S(\mu))$.

If $0 < \rho < \infty$, it steps toward a central path point with a smaller complementarity gap. We will show that when $\rho \geq \sqrt{n}$, then each iterate reduces the primal-dual potential function by at least a constant.

Logarithmic Approximation Lemma

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

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

Proof: Let $\mathbf{d}$ be the vector of eigenvalues of D. Then, $\mathbf{d} \in \mathcal{R}^n$ and $\|\mathbf{d}\|_\infty < 1$, and we proceed to prove

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

The Bound on Potential Reduction for SDP

Let $V^{1/2} = D^{-.5}XD^{-.5} = D^{.5}SD^{.5} \in \operatorname{int} \mathcal{S}^n_+$. Then, one can verify that $S \bullet X = I \bullet V$.

Lemma 2 *Let the direction D_X, $\mathbf{d}_y$ and D_S be generated by equation (1), and let*

$$\theta = \frac{\alpha}{\|V^{-1/2}\|_\infty \|\frac{I \bullet V}{n+\rho} V^{-1/2} - V^{1/2}\|}\ , \tag{3}$$

where α is a positive constant less than 1. Let

$$X^+ = X + \theta D_X, \quad y^+ = y + \theta \mathbf{d}_y, \quad \textit{and} \quad S^+ = S + \theta D_S.$$

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

$$\psi_{n+\rho}(X^+, S^+) - \psi_{n+\rho}(X, S) \le -\alpha \frac{\|V^{-1/2} - \frac{n+\rho}{I \bullet V} V^{1/2}\|}{\|V^{-1/2}\|_\infty} + \frac{\alpha^2}{2(1-\alpha)}\ .$$

Technical Lemmas

Lemma 3 *Let $V \in \operatorname{int} \mathcal{S}^n_+$ and $\rho \geq \sqrt{n}$. Then,*

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

Proof: Let $\mathbf{v}$ be the vector of eigenvalues of V. Then $\mathbf{v} \in \mathcal{R}^n_+$, and for $\rho \geq \sqrt{n}$ we proceed to prove

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

From these lemmas

$$\psi_{n+\rho}(X^+, S^+) - \psi_{n+\rho}(X, S) \leq -\alpha\sqrt{3/4} + \frac{\alpha^2}{2(1-\alpha)}.$$

By carefully choose α, we have a constant potential reduction in each iteration for SDP.

Description of Algorithm for SDP

Given $(X^0, y^0, S^0) \in \operatorname{int} \mathcal{F}$. 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 compute $(D_X, \mathbf{d}_y, D_S)$ from (1).

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

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

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

Complexity of the Algorithm

Theorem 5 *Let* $\rho = \sqrt{n}$ *and* $\psi_{n+\rho}(X^0, S^0) \le \rho \log(X^0 \bullet S^0) + n \log n$. *Then, the SDP Algorithm terminates in at most* $O(\sqrt{n} \log(X^0 \bullet S^0/\epsilon)$ *iterations with*

$$X^k \bullet S^k = C \bullet X^k - \mathbf{b}^T \mathbf{y}^k \le \epsilon.$$

Practical Computational Difficulty:

- The iteration complexity of SDP is in the order of $O(m^3 + mn^3 + m^2n^2)$
- It has to solve a dense system of linear equations at each iteration
- In general, $n = 10000$ is the bottle-neck for practical efficiency, in contrast to linear programming.