Conic Linear Programming Algorithms

Yinyu Ye

Department of Management Science and Engineering

Stanford University

Stanford, CA 94305, U.S.A.

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

Chapter 6.6

Recall 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 convex cone.

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

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

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

Note that cone K can be a product of many (different) convex cones.

Dual of Conic LP

The dual problem to

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

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

Theorem 1 *(Weak duality theorem)*

$$\mathbf{c} \bullet \mathbf{x} - \mathbf{b}^T \mathbf{y} = \mathbf{x} \bullet \mathbf{s} \geq 0$$

for any feasible $\mathbf{x}$ *of (CLP) and* $(\mathbf{y}, \mathbf{s})$ *of (CLD).*

Self-Dual Cones Again

Frequently, $K^* = K$, that is, they are self-dual.

The dual of the n-dimensional non-negative orthant, $\mathcal{R}^n_+ = \{\mathbf{x} \in \mathcal{R}^n : \mathbf{x} \geq \mathbf{0}\}$, is $\mathcal{R}^n_+$; it is self-dual.

The dual of the positive semi-definite symmetric matrix cone in $\mathcal{S}^n$, $\mathcal{S}^n_+$, is $\mathcal{S}^n_+$; it is self-dual.

The dual of the second-order cone, $\{\mathbf{x} \in \mathcal{R}^n : x_1 \geq \|\mathbf{x}_{-1}\|\}$, is also the second-order cone; it is self-dual.

SOCP Examples

$$\text{minimize} \quad \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} \bullet \mathbf{x}$$

$$\text{subject to} \quad \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \bullet \mathbf{x} = 1, \ \mathbf{x} \in SOC.$$

Dual:

$$\text{maximize} \quad y$$

$$\text{subject to} \quad \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} - y \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \mathbf{s} \in SOC.$$

SDP Examples

$$\text{minimize} \quad \begin{pmatrix} 2 & .5 \\ .5 & 1 \end{pmatrix} \bullet X$$

$$\text{subject to} \quad \begin{pmatrix} 1 & .5 \\ .5 & 1 \end{pmatrix} \bullet X = 1, \;\; X \succeq \mathbf{0}.$$

Dual:

$$\text{maximize} \quad y$$

$$\text{subject to} \quad \begin{pmatrix} 2 & .5 \\ .5 & 1 \end{pmatrix} - y \cdot \begin{pmatrix} 1 & .5 \\ .5 & 1 \end{pmatrix} = S \succeq \mathbf{0},$$

Conic Linear Programming in Compact Form

$$
\begin{array}{lll}
(CLP) & \text{minimize} & \mathbf{c} \bullet \mathbf{x} \\
& \text{subject to} & \mathcal{A}\mathbf{x} = \mathbf{b}, \\
& & \mathbf{x} \in K.
\end{array}
$$

$$
\begin{array}{lll}
(CLD) & \text{maximize} & \mathbf{b}^T \mathbf{y} \\
& \text{subject to} & \mathcal{A}^T \mathbf{y} + \mathbf{s} = \mathbf{c}, \\
& & \mathbf{s} \in K^*.
\end{array}
$$

Denote by $\mathcal{F}_p$ and $\mathcal{F}_d$ the primal and dual feasible sets, respectively.

Optimality Conditions for CLP

$$
\begin{array}{rcl}
\mathbf{c} \bullet \mathbf{x} - \mathbf{b}^T \mathbf{y} & = & 0 \\
\mathcal{A}\mathbf{x} & = & \mathbf{b} \\
-\mathcal{A}^T \mathbf{y} - \mathbf{s} & = & -\mathbf{c} \\
\mathbf{x} \in K, \ \mathbf{s} \in K^* & & .
\end{array}, \qquad (1)
$$

Or

$$
\begin{array}{rcl}
\mathbf{x} \bullet \mathbf{s} & = & 0 \\
\mathcal{A}\mathbf{x} & = & \mathbf{b} \\
-\mathcal{A}^T \mathbf{y} - \mathbf{s} & = & -\mathbf{c} \\
\mathbf{x} \in K, \ \mathbf{s} \in K^* & & .
\end{array} \qquad (2)
$$

Barrier Functions for Convex Cones

A differentiable function $B(\mathbf{x})$ is called barrier function for a closed convex cone K if for the sequence $\{\mathbf{x}^k \in \text{int}\, K\}$, $k = 1, \ldots$,

$$\mathbf{x}^k \to \partial K \quad \Rightarrow \quad B(\mathbf{x}^k) \to \infty,$$

where ∂K represents the boundary of K. $\mathbf{x} \bullet (-\nabla B(\mathbf{x}))$ is called the barrier-coefficient of $B(\mathbf{x})$, denoted by ν; and a point in $\text{int}\, K$ is called the central point if it is a fixed point of

$$\mathbf{x} = -\nabla B(\mathbf{x}),$$

denoted by $\mathbf{e}^c$.

Logarithmic Barrier Functions

- $\mathcal{R}^n_+$:

$$B(\mathbf{x}) = -\sum_{j=1}^{n} \ln(x_j), \; \nabla B(\mathbf{x}) = -\Delta(\mathbf{x})^{-1}\mathbf{e}, \; \nabla^2 B(\mathbf{x}) = \Delta(\mathbf{x})^{-2} \in \mathcal{S}^n.$$

  The central point is $\mathbf{e}$, the vector of all ones, and the barrier-coefficient is $\mathbf{x} \bullet (-\nabla B(\mathbf{x})) = \mathbf{x} \bullet \Delta(\mathbf{x})^{-1}\mathbf{e} = n$.

- $\mathcal{S}^n_+$:

$$B(X) = -\ln\det(X), \; \nabla B(X) = -X^{-1},$$

$$\nabla^2 B(X) = \{\partial^2 B(X)/\partial X_{ij}\partial X_{kl} = X^{-1}_{ik} X^{-1} jl\} = X^{-1} \otimes X^{-1} \in \mathcal{S}^{n^2},$$

  where $\otimes$ stands for matrix Kronecker product. The central point is I, the identity matrix, and the barrier-coefficient is $X \bullet (-\nabla B(X)) = X \bullet X^{-1} = n$.

- $\mathcal{N}_2^n$:

$$B(\mathbf{x}) = -\frac{1}{2}\ln(x_1^2 - \|\mathbf{x}_{-1}\|^2), \quad \nabla B(\mathbf{x}) = \frac{1}{\delta(\mathbf{x})^2}\begin{pmatrix} -x_1 \\ \mathbf{x}_{-1} \end{pmatrix},$$

$$\nabla^2 B(\mathbf{x}) = \frac{1}{\delta(\mathbf{x})^2}\begin{pmatrix} -1 & 0 \\ 0 & I \end{pmatrix} + \frac{2}{\delta(\mathbf{x})^4}\begin{pmatrix} x_1 \\ -\mathbf{x}_{-1} \end{pmatrix}\begin{pmatrix} x_1 \\ -\mathbf{x}_{-1} \end{pmatrix}^T,$$

where $\delta(\mathbf{x}) = \sqrt{x_1^2 - \|\mathbf{x}_{-1}\|^2}$. The central point is $\mathbf{e}_1$, the unit vector with 1 as its first element and zero everywhere else, and the barrier-coefficient is

$$\mathbf{x} \bullet (-\nabla B(\mathbf{x})) = \mathbf{x} \bullet \frac{-1}{\delta(\mathbf{x})^2}\begin{pmatrix} -x_1 \\ \mathbf{x}_{-1} \end{pmatrix} = 1.$$

- The mixed cone $K = K_1 \oplus K_2$, that is, $\mathbf{x} = [\mathbf{x}_1; \mathbf{x}_2]$ where $\mathbf{x}_1 \in K_1$ and $X_2 \in K_2$:

$$B(\mathbf{x}) = B_1(\mathbf{x}_1) + B_2(\mathbf{x}_2)$$

  where $B_1(\cdot)$ and $B_2(\cdot)$ are barrier functions for K_1 and K_2, respectively. The barrier-coefficient is the sum of the barrier-coefficients of the two cones.

The Central Path

Consider (CLP) with the μ-weighted barrier function added in the objective:

$$\begin{array}{lll} (CLPB) & \text{minimize} & \mathbf{c} \bullet \mathbf{x} + \mu B(\mathbf{x}) \\ & \text{s.t.} & \mathcal{A}\mathbf{x} = \mathbf{b}, \\ & & \mathbf{x} \in K; \end{array}$$

or (CLD) with the μ-weighted barrier function added in the objective:

$$\begin{array}{lll} (CLDB) & \text{maximize} & \mathbf{b}^T\mathbf{y} - \mu B(\mathbf{s}) \\ & \text{s.t.} & \mathcal{A}^T\mathbf{y} + \mathbf{s} = \mathbf{c}, \\ & & \mathbf{s} \in K^*. \end{array}$$

Theorem 2 *Let both (CLP) and (CLD) have interior feasible solutions. Then, for any given* $0 < \mu < \infty$*, the optimizers of (CLPB) and (CLDB) exist and they are unique and in the interior of cone* K *and* K^**, respectively. As* μ *continuously varies toward zero, they form a path (called the central path) converging to an interior point in the optimal face.*

The Central Path Equations

For any given $\mu > 0$, the optimizers of (CLPB) have necessary and sufficient conditions:

$$\begin{aligned} \mathbf{c} + \mu \nabla B(\mathbf{x}) - \mathcal{A}^T \mathbf{y} &= \mathbf{0} \\ \mathcal{A}\mathbf{x} &= \mathbf{b} \end{aligned}$$

Let $\mathbf{s} = \mathbf{c} - \mathcal{A}^T \mathbf{y}$. Then the conditions become

$$\begin{aligned} \mathbf{s} + \mu \nabla B(\mathbf{x}) &= \mathbf{0} \\ \mathcal{A}\mathbf{x} &= \mathbf{b} \\ -\mathcal{A}^T \mathbf{y} - \mathbf{s} &= -\mathbf{c}. \end{aligned} \tag{3}$$

One can verify that $\mathbf{s} = -\mu \nabla B(\mathbf{x}) \in \text{int } K^*$.

Similarly, the optimizers of (CLDB) have necessary and sufficient conditions:

$$
\begin{aligned}
\mathbf{x} + \mu \nabla B(\mathbf{s}) &= \mathbf{0} \\
\mathcal{A}\mathbf{x} &= \mathbf{b} \\
-\mathcal{A}^T \mathbf{y} - \mathbf{s} &= -\mathbf{c}.
\end{aligned}
\tag{4}
$$

One can verify that $\mathbf{x} = -\mu \nabla B(\mathbf{s}) \in \operatorname{int} K$.

Symmetric Central Path Equations for Self-dual Cones

Linear Programming:

$$\begin{aligned} \mathbf{x}.\cdot\mathbf{s} &= \mu\mathbf{e} \\ A\mathbf{x} &= \mathbf{b} \\ -A^T\mathbf{y}-\mathbf{s} &= -\mathbf{c} \end{aligned} \qquad \text{where } \mu = \frac{\mathbf{x}^T\mathbf{s}}{n}.$$

Second-Order Cone Programming:

$$\begin{aligned} \mathbf{x}.\cdot\mathbf{s} &= \mu\mathbf{e}_1 \\ A\mathbf{x} &= \mathbf{b} \\ -A^T\mathbf{y}-\mathbf{s} &= -\mathbf{c} \end{aligned} \qquad \text{where } \mu = \mathbf{x}^T\mathbf{s}.$$

Semidefinite Programming:

$$\begin{aligned} XS &= \mu I \\ \mathcal{A}X &= \mathbf{b} \\ -\mathcal{A}^T\mathbf{y}-S &= -C \end{aligned} \qquad \text{where } \mu = \frac{X\bullet S}{n}.$$

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* $0 < \mu \le \mu^0$ *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* $0 < \mu \le \mu^0$ *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

$$n \cdot I \succeq X(\mu)^{-1/2} X^* X(\mu)^{-1/2} \quad \text{or} \quad X(\mu) \succeq \frac{1}{n} X^*.$$

Path Following Algorithms

Suppose we have an approximate central path point $(\mathbf{x}, \mathbf{y}, \mathbf{s})$ in a neighborhood of $(\mathbf{x}(\mu), \mathbf{y}(\mu), \mathbf{s}(\mu))$ for a given $\mu > 0$. Then we consider to compute a new approximate central-path point $(X^+, \mathbf{y}^+, S^+)$ corresponding to a chosen μ^+ where $\mu > \mu^+ > 0$. If one repeats this process, then a sequence of approximate central-path points $(\mathbf{x}^k, \mathbf{y}^k, \mathbf{s}^k)$, corresponding to $\mu^0 > \mu^1 > ... > \mu^k, ...$, would be generated, and it converges to the optimal solution set as $\mu^k \to 0$.

If μ^+ is close to μ, we expect $(\mathbf{x}(\mu^+), \mathbf{y}(\mu^+), \mathbf{s}(\mu^+))$ is also close to $(\mathbf{x}, \mathbf{y}, \mathbf{s})$, so that $(\mathbf{x}, \mathbf{y}, \mathbf{s})$ would be a good initial point for computing $(\mathbf{x}^+, \mathbf{y}^+, \mathbf{s}^+)$ by numerical procedures such as Newton's method. Such an algorithm is called the path following algorithm.

Potential Reduction Algorithms

In practical computation, it is more efficient to generate iterative solutions in a large neighborhood as long as a merit function is monotonically decreasing, so that the greater the reduction of the function, the faster convergence of the iterative solutions to optimality. Such an algorithm is said function-driven. If the merit function is the objective function itself, a function-driven algorithm is likely to generate iterative solutions being prematurely too close to the boundary, and the convergence would be slow down in future iterations. A better driven function should balance the reduction of the objective function as well as a good position in the (interior) of the feasible region – we now present a potential function logarithmically combining the objective function and the barrier function.

Potential and Duality Gap in LP

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

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

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

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

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

Potential Function 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

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

The Potential Reduction Algorithm

The potential reduction algorithm generates a sequence of $\{X^k, \mathbf{y}^k, S^k\} \in \mathrm{int}\,\mathcal{F}$ 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:

$$\begin{aligned} D^{-1}D_X D^{-1} + D_S &= R := \tfrac{n}{n+\rho}\mu X^{-1} - S, \\ \mathcal{A}D_X &= \mathbf{0}, \\ -\mathcal{A}^T\mathbf{d}_y - D_S &= \mathbf{0}, \end{aligned} \tag{5}$$

where

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

Note that $D_S \bullet D_X = 0$.

Primal-Dual Scaling

$$\begin{array}{rcl} D_{X'} + D_{S'} & = & R', \\ \mathcal{A}' D_{X'} & = & \mathbf{0}, \\ -\mathcal{A}'^T \mathbf{d}_y - D_{S'} & = & \mathbf{0}, \end{array} \tag{6}$$

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 for SDP

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

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

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

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

The Bound on Potential Reduction for SDP

Let $V^{1/2} = D^{-.5} X D^{-.5} = D^{.5} S D^{.5} \in \operatorname{int} \mathcal{S}^n_+$. 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 (5), and let*

$$\theta = \frac{\alpha}{\|V^{-1/2}\|_\infty \|\frac{I \bullet V}{n+\rho} V^{-1/2} - V^{1/2}\|}, \tag{7}$$

where α is a positive constant less than 1. Let

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

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* $V \in \text{int}\, \mathcal{S}^n_+$ *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 (5).
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.

Dual Interior-Point 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: an Alternating Descent Method

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

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

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

Its gradient is

$$\nabla\psi(\mathbf{y}, \bar{z}) = -\frac{\rho}{\bar{z} - \mathbf{b}^T\mathbf{y}}\mathbf{b} + \mathcal{A}S^{-1}. \tag{8}$$

We minimize over $\mathbf{y}$ first, then over X second. Recall

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

Over-Estimator of Potential

For any given $\mathbf{y}$ and $S = C - \mathcal{A}^T\mathbf{y} \succ \mathbf{0}$ and

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

$$\begin{aligned}
&\psi(\mathbf{y}, \bar{z}^k) - \psi(\mathbf{y}^k, \bar{z}^k) \\
&= \rho\ln(\bar{z}^k - \mathbf{b}^T\mathbf{y}) - \rho\ln(\bar{z}^k - \mathbf{b}^T\mathbf{y}^k) - \ln\det((S^k)^{-.5}S(S^k)^{-.5}) \\
&\le -\frac{\rho}{\bar{z}^k - \mathbf{b}^T\mathbf{y}^k}\mathbf{b}^T(\mathbf{y}-\mathbf{y}^k) + I \bullet ((S^k)^{-.5}S(S^k)^{-.5} - I) \\
&\quad + \frac{\|(S^k)^{-.5}\left(\mathcal{A}^T(\mathbf{y}-\mathbf{y}^k)\right)(S^k)^{-.5}\|}{2(1-\|(S^k)^{-.5}(\mathcal{A}^T(\mathbf{y}-\mathbf{y}^k))(S^k)^{-.5}\|_\infty)} \\
&= -\frac{\rho}{\bar{z}^k - \mathbf{b}^T\mathbf{y}^k}\mathbf{b}^T(\mathbf{y}-\mathbf{y}^k) + (\mathcal{A}(S^k)^{-1})^T(\mathbf{y}-\mathbf{y}^k) \\
&\quad + \frac{\|(S^k)^{-.5}\left(\mathcal{A}^T(\mathbf{y}-\mathbf{y}^k)\right)(S^k)^{-.5}\|}{2(1-\|(S^k)^{-.5}(\mathcal{A}^T(\mathbf{y}-\mathbf{y}^k))(S^k)^{-.5}\|_\infty)} \\
&= \nabla\psi(\mathbf{y}^k, \bar{z}^k)^T(\mathbf{y}-\mathbf{y}^k) + \frac{\|(S^k)^{-.5}\left(\mathcal{A}^T(\mathbf{y}-\mathbf{y}^k)\right)(S^k)^{-.5}\|}{2(1-\|(S^k)^{-.5}(\mathcal{A}^T(\mathbf{y}-\mathbf{y}^k))(S^k)^{-.5}\|_\infty)}.
\end{aligned} \tag{9}$$

Solve the Ball Constrained Problem

$$\begin{array}{ll} \text{Minimize} & \nabla\psi^T(\mathbf{y}^k, \bar{z}^k)(\mathbf{y} - \mathbf{y}^k) \\ \text{subject to} & \|(S^k)^{-.5}\left(\mathcal{A}^T(\mathbf{y} - \mathbf{y}^k)\right)(S^k)^{-.5}\| \le \alpha, \end{array} \tag{10}$$

where α is a positive constant less than 1 that would be determined later.

For simplicity, in what follows we let the current duality gap be

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

Optimality Conditions

The first order KKT conditions state that the minimum point, $\mathbf{y}^{k+1}$, of this convex minimization problem satisfies

$$M^k(\mathbf{y}^{k+1} - \mathbf{y}^k) + \beta \nabla \psi(\mathbf{y}^k, \bar{z}^k) = 0 \tag{11}$$

for a positive multiplier 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 it is positive definite when $S^k \succ 0$ and A_i's are linearly independent.

Close-Form Solution

Using the ellipsoidal constraint being tight, the minimal solution, $\mathbf{y}^{k+1}$, of (10) is given by a close form

$$\mathbf{y}^{k+1} - \mathbf{y}^k = \frac{\alpha}{\sqrt{\nabla\psi^T(\mathbf{y}^k, \bar{z}^k)(M^k)^{-1}\nabla\psi(\mathbf{y}^k, \bar{z}^k)}}\mathbf{d}(\bar{z}^k)_y \tag{12}$$

where

$$\mathbf{d}(\bar{z}^k)_y = -(M^k)^{-1}\nabla\psi(\mathbf{y}^k, \bar{z}^k). \tag{13}$$

Potential Reduction

We can derive

$$\nabla\psi^T(\mathbf{y}^k,\bar{z}^k)\mathbf{d}(\bar{z}^k)_y = -\nabla\psi^T(\mathbf{y}^k,\bar{z}^k)(M^k)^{-1}\nabla\psi(\mathbf{y}^k,\bar{z}^k) = -\|P(\bar{z}^k)\|^2 \tag{14}$$

where

$$P(\bar{z}^k) = \frac{\rho}{\Delta^k}(S^k)^{.5}X(\bar{z}^k)(S^k)^{.5} - I, \tag{15}$$

and

$$X(\bar{z}^k) = \frac{\Delta^k}{\rho}(S^k)^{-1}\left(\mathcal{A}^T\mathbf{d}(\bar{z}^k)_y + S^k\right)(S^k)^{-1}. \tag{16}$$

Thus,

$$\psi(\mathbf{y}^{k+1},\bar{z}^k) - \psi(\mathbf{y}^k,\bar{z}^k) \le -\alpha\|P(\bar{z}^k)\| + \frac{\alpha^2}{2(1-\alpha)}. \tag{17}$$

Potential Primal Feasible Solution and its Objective Value

$X(\bar{z}^k)$ is actually the minimizer of the least squares problem

$$\begin{array}{ll} \text{Minimize} & \|(S^k)^{.5}X(S^k)^{.5} - \frac{\Delta^k}{\rho}I\| \\ \text{subject to} & \mathcal{A}X = \mathbf{b}. \end{array} \tag{18}$$

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

Since the vectors $\mathcal{A}(S^k)^{-1}$ and $\mathbf{d}(\bar{z}^k)_y$ were calculated, the cost of computing a primal objective value is the cost of a vector dot product!

But $X(\bar{z}^k)$ may not be PSD...

When the Primal is Feasible

We have the following lemma:

Lemma 4 *Let* $\mu^k = \frac{\Delta^k}{n} = \frac{\bar{z}^k - \mathbf{b}^T\mathbf{y}^k}{n}$, $\mu = \frac{X(\bar{z}^k)\bullet S^k}{n} = \frac{C\bullet X(\bar{z}^k) - \mathbf{b}^T\mathbf{y}^k}{n}$, $\rho \geq n + \sqrt{n}$, *and* $\alpha < 1$. *If*

$$\|P(\bar{z}^k)\| < \min\left(\alpha\sqrt{\frac{n}{n+\alpha^2}}, 1-\alpha\right), \tag{19}$$

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. $\mu \leq (1 - \frac{\alpha}{2\sqrt{n}})\mu^k$.

Alternating Potential Reduction

Thus, if $\|P(\bar{z}^k)\| \geq \min\left(\alpha\sqrt{\frac{n}{n+\alpha^2}}, 1-\alpha\right)$, we update $\mathbf{y}^k$ to $\mathbf{y}^{k+1}$; otherwise, we let $X^{k+1} = X(\bar{z}^k)$. In such alternating moves, we have

Theorem 6 *Either the primal-dual potential*

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

or

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

where $\delta > 1/20$.

Description of Algorithm

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

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

begin

1. Compute $\mathcal{A}(S^k)^{-1}$ and formulate the Gram matrix M^k.
2. Solve (13) for the dual step direction $\mathbf{d}(\bar{z}^k)_y$.
3. Calculate $\|P(\bar{z}^k)\|$ using (14).
4. **If** (19) is true, **then** $X^{k+1} = X(\bar{z}^k)$, $\bar{z}^{k+1} = C \bullet X^{k+1}$, and $(\mathbf{y}^{k+1}, S^{k+1}) = (\mathbf{y}^k, S^k)$;

   else $\mathbf{y}^{k+1} = \mathbf{y}^k + \frac{\alpha}{\|P(\bar{z}^k)\|}\mathbf{d}(\bar{z}^{k+1})_y$, $S^{k+1} = C - \mathcal{A}^T(\mathbf{y}^{k+1})$, $X^{k+1} = X^k$, and $\bar{z}^{k+1} = \bar{z}^k$.

   endif
5. $k := k + 1$.

end Note that we do not need eigenvalue computation in evaluate $\|P(\bar{z}^k)\|$, but use

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

Corollary 1 *Let $\rho = \sqrt{n}$. Then, the Algorithm terminates in at most $O(\sqrt{n}\log(C \bullet X^0 - \mathbf{b}^T\mathbf{y}^0)/\epsilon)$ iterations with*

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

Formulation Work of M^k

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

$$M^k = \begin{pmatrix} (a_1^T(S^k)^{-1}a_1)^2 & \cdots & (a_1^T(S^k)^{-1}a_m)^2 \\ \vdots & \ddots & \vdots \\ (a_m^T(S^k)^{-1}a_1)^2 & \cdots & (a_m^T(S^k)^{-1}a_m)^2 \end{pmatrix} \qquad (20)$$

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, $(S^k)^{-1}$.

Quick Computation with the Rank-One Structure

Let $A^T = [\mathbf{a}_1\ \mathbf{a}_2\ ...\ \mathbf{a}_m]$ and $A' = A(S^k)^{-1/2}$. Then we have

$$M^k = [A(S^k)^{-1}A^T]^2 = [A'(A')^T]^2$$

and

$$A(S^k)^{-1} = \text{diag}(A(S^k)^{-1}A^T) = \text{diag}(A'(A')^T).$$

Thus, M^k and the gradient vector can be computed in $O(n^3 + n^2m + nm^2)$ arithmetic operations.

Then, the dual direction $\mathbf{d}(\cdot)_y$ can be computed in $O(m^3)$ operations.

The norm of $P(\cdot)$ can be checked in $O(m^2)$ operations and the new upper bound can be updated in $O(m)$ operations.

If needed, $X(\cdot)$ can be computed in $O(n^3 + n^2m)$ operations.

Primal-Dual SDP Alternative Systems

A pair of SDP has two alternatives under mild conditions

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

or

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

An Integrated Homogeneous and Self-Dual System

The two alternative systems can be homogenized as one:

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

where the three alternatives are

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