Final Review

Yinyu Ye

Department of Management Science and Engineering

Stanford University

Stanford, CA 94305, U.S.A.

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

Convex Cones

- A set C is a cone if $\mathbf{x} \in C$ implies $\alpha\mathbf{x} \in C$ for all $\alpha > 0$
- A convex cone is cone plus convex-set.
- Dual cone:

$$C^* := \{\mathbf{y} : \ \mathbf{y} \bullet \mathbf{x} \geq 0 \quad \text{for all} \quad \mathbf{x} \in C\}$$

$-C^*$ is also called the polar of C.

Separating hyperplane theorem

The most important theorem about the convex set is the following separating theorem.

Theorem 1 *(Separating hyperplane theorem) Let* $C \subset \mathcal{E}$*, where* $\mathcal{E}$ *is either* $\mathcal{R}^n$ *or* $\mathcal{M}^n$*, be a closed convex set and let* $\mathbf{y}$ *be a point exterior to* C*. Then there is a vector* $\mathbf{a} \in \mathcal{E}$ *such that*

$$\mathbf{a} \bullet \mathbf{y} < \inf_{\mathbf{x} \in C} \mathbf{a} \bullet \mathbf{x}.$$

Theorem 2 *(LP fundamental theorem) Given (LP) and (LD) where* A *has full row rank* m *and* $\mathbf{b} \in R^m$ *and* $\mathbf{c} \in R^n$*, and consider the classical LP:*

$$(LP) \quad \text{minimize} \quad \mathbf{c}^T\mathbf{x}$$
$$\text{subject to} \quad A\mathbf{x} = \mathbf{b}, \ \mathbf{x} \geq \mathbf{0},$$

where decision vector $\mathbf{x} \in R^n$*.*

i) *if there is a feasible solution, there is a basic feasible solution (Carathéodory's theorem);*

ii) *if there is an optimal solution, there is an optimal basic solution.*

Basic Solution: select m linearly independent columns, denoted by the index set B, from A, and solve

$$A_B\mathbf{x}_B = \mathbf{b}$$

to determine the m-vector $\mathbf{x}_B$. By setting the variables, $\mathbf{x}_N$, of $\mathbf{x}$ corresponding to the remaining columns of A equal to zeros. If $\mathbf{x}_B \geq \mathbf{0}$, then it is a basic feasible solution (BFS).

Theorem 3 *(SDP fundamental theorem) Consider the SDP problem*

$$(SDP) \quad \text{minimize} \quad C \bullet X$$
$$\text{subject to} \quad A_i \bullet X = b_i,\ i = 1, 2, ..., n;\ X \succeq \mathbf{0},$$

where symmetric decision matrix $X \in S^n$.

i) *if there is a feasible solution, there is a feasible solution (Carathéodory's theorem) whose rank* r *meets*

$$\frac{r(r+1)}{2} \leq m.$$

ii) *if there is an optimal solution, there is an optimal basic solution whose rank* r *meets*

$$\frac{r(r+1)}{2} \leq m.$$

Farkas' Lemma

The following results are Farkas' lemma and its variants.

Theorem 4 *The system* $\{\mathbf{x} : \ A\mathbf{x} = \mathbf{b}, \ \mathbf{x} \geq \mathbf{0}\}$ *has a feasible solution* $\mathbf{x}$ *if and only if there is no* $\mathbf{y}$ *to satisfy* $-A^T\mathbf{y} \geq \mathbf{0}$ *and* $\mathbf{b}^T\mathbf{y} > 0$*.*

A vector $\mathbf{y}$, with $-A^T\mathbf{y} \geq \mathbf{0}$ and $\mathbf{b}^T\mathbf{y} > 0$, is called a (primal) infeasibility certificate for the system $\{\mathbf{x} : \ A\mathbf{x} = \mathbf{b}, \ \mathbf{x} \geq \mathbf{0}\}$. Geometrically, Farkas' lemma means that if a vector $\mathbf{b} \in \mathcal{R}^m$ does not belong to the cone generated by $\mathbf{a}_{.1}, ..., \mathbf{a}_{.n}$, then there is a hyperplane separating $\mathbf{b}$ from cone$(\mathbf{a}_{.1}, ..., \mathbf{a}_{.n})$.

Theorem 5 *Let* C *be a (pointed) closed convex cone. Then, the system* $\{\mathbf{x} : \ A\mathbf{x} = \mathbf{b}, \ \mathbf{x} \in C\}$ *has a feasible solution* $\mathbf{x}$ *if and only if there is no* $\mathbf{y}$ *to satisfy* $-A^T\mathbf{y} \in C^*$ *and* $\mathbf{b}^T\mathbf{y} > 0$*, provided that there is* $\mathbf{y}$ *such that* $-A^T\mathbf{y} \in \text{int}\, C^*$*.*

Duality Theory

Consider the linear program in standard form, called the primal problem,

$$(LP) \quad \text{minimize} \quad \mathbf{c}^T\mathbf{x}$$
$$\text{subject to} \quad A\mathbf{x} = \mathbf{b}, \ \mathbf{x} \geq \mathbf{0},$$

where $\mathbf{x} \in \mathcal{R}^n$.

The dual problem can be written as:

$$(LD) \quad \text{maximize} \quad \mathbf{b}^T\mathbf{y}$$
$$\text{subject to} \quad A^T\mathbf{y} + \mathbf{s} = \mathbf{c}, \ \mathbf{s} \geq \mathbf{0},$$

where $\mathbf{y} \in \mathcal{R}^m$ and $\mathbf{s} \in \mathcal{R}^n$. The components of $\mathbf{s}$ are called dual slacks.

Duality Theory

Theorem 6 *(Weak duality theorem) Let $\mathcal{F}_p$ and $\mathcal{F}_d$ be non-empty. Then,*

$$\mathbf{c}^T\mathbf{x} \geq \mathbf{b}^T\mathbf{y} \quad \textit{where} \quad \mathbf{x} \in \mathcal{F}_p, \ (\mathbf{y}, \mathbf{s}) \in \mathcal{F}_d.$$

$$\mathbf{c}^T\mathbf{x} - \mathbf{b}^T\mathbf{y} = \mathbf{c}^T\mathbf{x} - (A\mathbf{x})^T\mathbf{y} = \mathbf{x}^T(\mathbf{c} - A^T\mathbf{y}) = \mathbf{x}^T\mathbf{s} \geq 0.$$

This theorem shows that a feasible solution to either problem yields a bound on the value of the other problem. We call $\mathbf{c}^T\mathbf{x} - \mathbf{b}^T\mathbf{y}$ the duality gap.

From this we have important results in the following.

Theorem 7 *(LP Strong Duality Theorem) Let* $\mathcal{F}_p$ *and* $\mathcal{F}_d$ *be non-empty. Then,* $\mathbf{x}^*$ *is optimal for (LP) and* $(\mathbf{y}^*, \mathbf{s}^*)$ *is optimal for (LD) if any only if*

$$\mathbf{c} \bullet \mathbf{x}^* = \mathbf{b}^T \mathbf{y}^*.$$

If one of (LP) or (LD) is unbounded then the other has no feasible solution. If one of (LP) or (LD) has no feasible solution, then the other is either unbounded or has no feasible solution.

Theorem 8 *(CLP Strong Duality Theorem) Let* $\mathcal{F}_p$ *and* $\mathcal{F}_d$ *be non-empty and at least one of them has an interior. Then,* $\mathbf{x}^*$ *is optimal for (CLP) and* $(\mathbf{y}^*, \mathbf{s}^*)$ *is optimal for (CLD) if any only if*

$$\mathbf{c} \bullet \mathbf{x}^* = \mathbf{b}^T \mathbf{y}^*.$$

There are cases that the duality gap tends to zero but the optimal solution is not attainable.

Complementarity

For feasible $\mathbf{x}$ and $(\mathbf{y}, \mathbf{s})$ of LP, $\mathbf{x}^T\mathbf{s} = \mathbf{x}^T(\mathbf{c} - A^T\mathbf{y}) = \mathbf{c}^T\mathbf{x} - \mathbf{b}^T\mathbf{y}$ is also called the complementarity gap. Note that $\mathbf{x}^T\mathbf{s} = 0$ implies that $x_j s_j = 0$ for all $j = 1, \ldots, n$.

$$
\begin{aligned}
x_j s_j &= 0, \ \forall j \\
A\mathbf{x} &= \mathbf{b} \\
-A^T\mathbf{y} - \mathbf{s} &= -\mathbf{c} \\
\mathbf{x} \geq \mathbf{0} \quad &, \quad \mathbf{s} \geq \mathbf{0}.
\end{aligned}
$$

For SDP:

$$
\begin{aligned}
X_i \cdot S_j &= 0, \ \forall i, j \\
\mathcal{A}X &= \mathbf{b} \\
-\mathcal{A}^T\mathbf{y} - S &= -C \\
X \succeq \mathbf{0} \quad &, \quad S \succeq \mathbf{0}.
\end{aligned}
$$

where X_i (or S_i) is the ith column of X (or S).

Rules to construct the dual

obj. coef. vector	right-hand-side
right-hand-side	obj. coef. vector
A	A^T
Max model	Min model
$x_j \geq 0$	jth constraint $\geq$
$x_j \leq 0$	jth constraint $\leq$
x_j free	jth constraint $=$
ith constraint $\leq$	$y_i \geq 0$
ith constraint $\geq$	$y_i \leq 0$
ith constraint $=$	y_i free

Duality Example

Consider the combinatorial call auction market discussed in the class. This time, the market maker forms the decision problem as:

$$\begin{array}{lll} \max & \sum_{j=1}^{n} x_j & \\ \text{s.t.} & A\mathbf{x} - \mathbf{e} \cdot y & \leq \mathbf{0}, \ (\mathbf{p}) \\ & -\pi^T \mathbf{x} + \alpha \cdot y & \leq 0, \ (\lambda) \\ & \mathbf{x} & \leq \mathbf{q}, \ (\mu) \\ & \mathbf{x} & \geq \mathbf{0}, \end{array}$$

where $(\pi_j, \mathbf{a}_j, q_j)$ are as defined as in our auction problem through out this course, $\mathbf{e}$ is the vector of all ones, and parameter $\alpha \geq 0$. Again, the bidder wins one dollar if the winning state is in his or her selection.

Dual Basic Feasible Solution

For every basis B, a dual vector $\mathbf{y}$ satisfying

$$A_B^T \mathbf{y} = c_B$$

is said to be the corresponding dual basic solution.

If the dual basic solution is also feasible, that is

$$\mathbf{s} = \mathbf{c} - A^T \mathbf{y} \geq \mathbf{0}.$$

If one or more slacks in $\mathbf{c}_N - A_N^T \mathbf{y}$ has value zero, that dual basic feasible solution $\mathbf{y}$ is said to be (dual) degenerate.

The Simplex Algorithm

0. Initialize with a minimization problem in feasible canonical form with respect to a basic index set B. Let N denote the complementary index set.

1. Test for termination: first find

$$r_e = \min_{j \in N}\{r_j\}.$$

   If $r_e \geq 0$, stop. The solution is optimal. Otherwise determine whether the column of $\bar{A}_{.e}$ contains a positive entry. If not, the objective function is unbounded below. Terminate. Let x_e be the entering basic variable.

2. Determine the outgoing: execute the MRT to determine the outgoing variable x_o.

3. Update basis: update B and A_B and transform the problem to canonical form and return to Step 1.

The Ellipsoid Method

The basic ideas of the ellipsoid method stem from research done in the nineteen sixties and seventies mainly in the Soviet Union (as it was then called) by others who preceded Khachiyan. The idea in a nutshell is to enclose the region of interest in each member of a sequence of ellipsoids whose size is decreasing, resembling the bisection method.

The significant contribution of Khachiyan was to demonstrate in two papers—published in 1979 and 1980—that under certain assumptions, the ellipsoid method constitutes a polynomially bounded algorithm for linear programming.

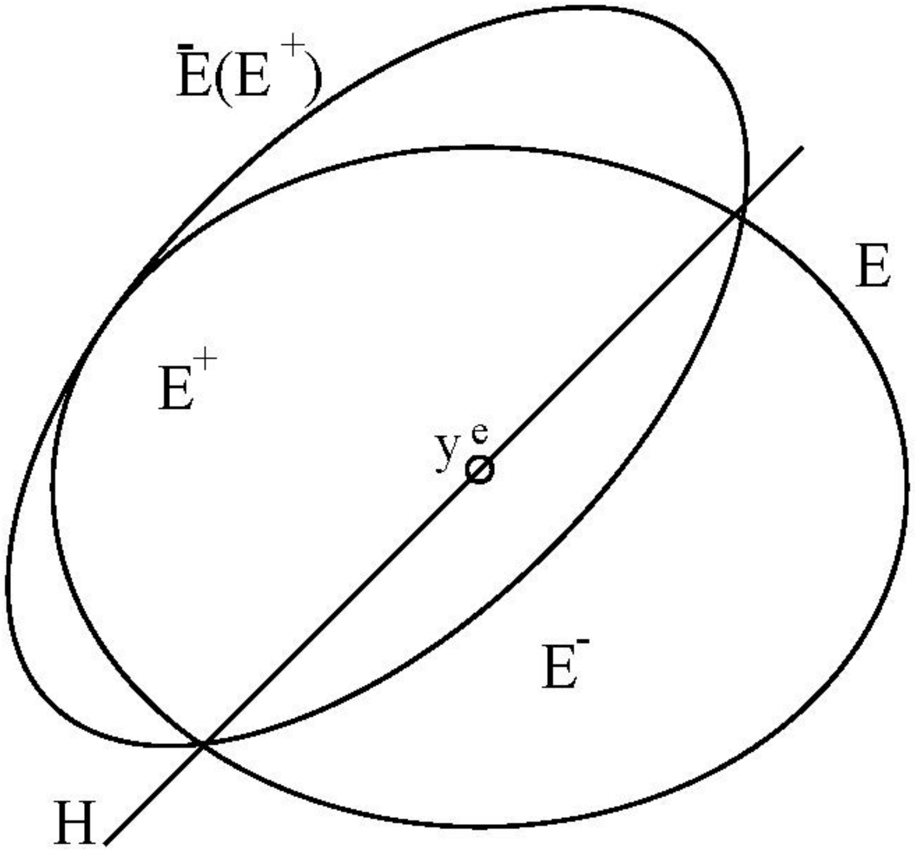

Figure 1: The least volume ellipsoid containing a half ellipsoid

Desired Theoretical Properties

- Separation Problem: Either decide the ellipsoid center $\mathbf{y}^c \in P$, where P is the target set, or find a separating hyperplane $\mathbf{a}$ such that $\mathbf{a}^T\mathbf{y} \leq \mathbf{a}^T\mathbf{y}^c$ for all $\mathbf{y} \in P$.
- Oracle to generate $\mathbf{a}$ without enumerating all hyperplanes.

Theorem 9 *If the separating (oracle) problem can be solved in polynomial time of m and $\log(R/r)$, then we can solve the standard linear programming problem whose running time is polynomial in m and $\log(R/r)$ that is independent of n, the number of inequality constraints.*

The Methodology Concept of Centers

Consider linear program

$$\begin{array}{ll} \text{maximize} & \mathbf{b}^T\mathbf{y} \\ \text{subject to} & A^T\mathbf{y} \le \mathbf{c}. \end{array}$$

Consider an objective level set

$$Y(z^0) := \{\mathbf{y} : \ A^T\mathbf{y} \le \mathbf{c}, \ \mathbf{b}^T\mathbf{y} \ge z^0\},$$

and assume that it is bounded and has an interior.

Compute a "center", $\mathbf{y}^0$, of the level set $Y(z^0)$, then move the objective hyperplane through $\mathbf{y}^0$, and now consider the smaller level set

$$Y(z^1) := \{\mathbf{y} : \ A^T\mathbf{y} \le \mathbf{c}, \ \mathbf{b}^T\mathbf{y} \ge z^1 = \mathbf{b}^T\mathbf{y}^0\}$$

and repeat this process.

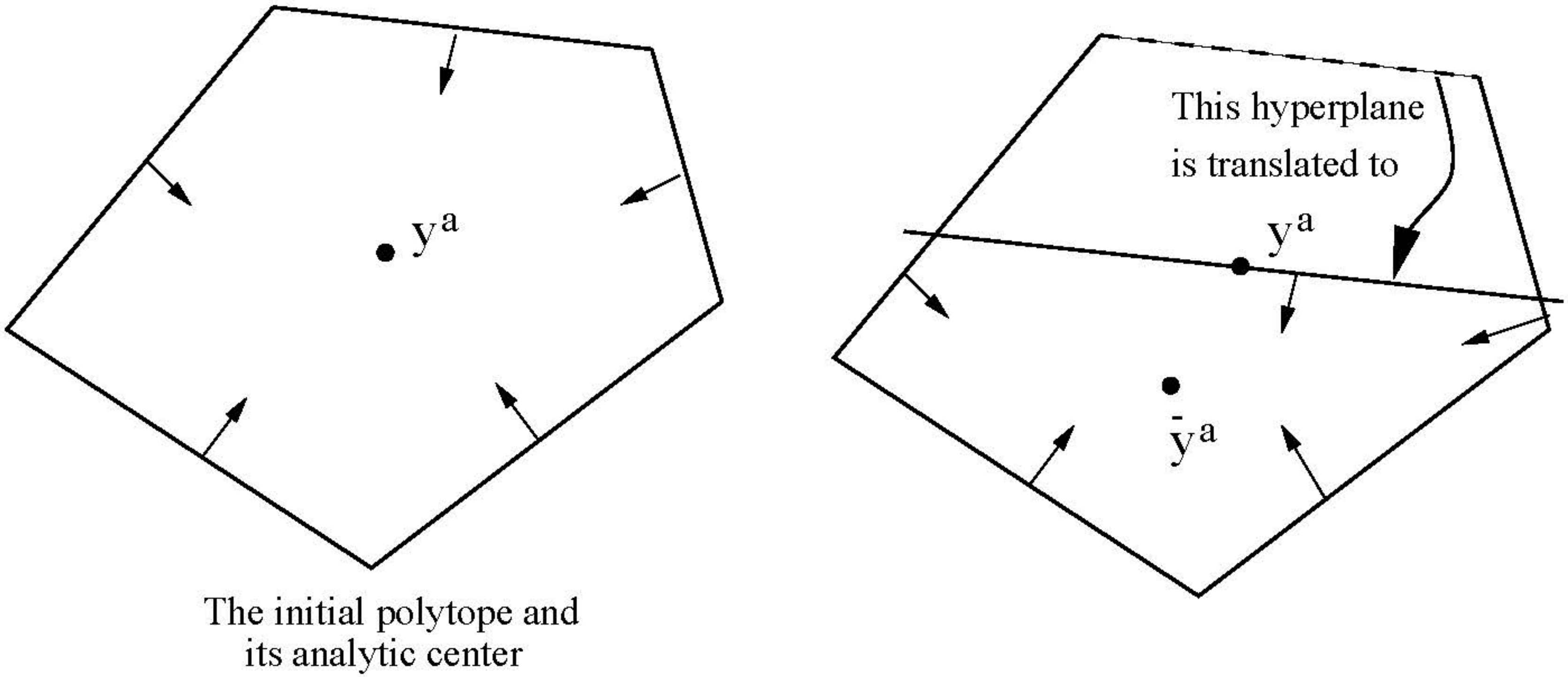

Figure 2: Cur ot translation of a hyperplane through the center.

Analytic Center for the Polytope

One choice of center is the one to maximize the barrier function over the level set:

$$\begin{array}{ll} \text{maximize} & \log s_0 + \sum_j \log s_j \\ \text{subject to} & A^T\mathbf{y} + \mathbf{s} = \mathbf{c}, \\ & \mathbf{b}^T\mathbf{y} - s_0 = z^0. \end{array}$$

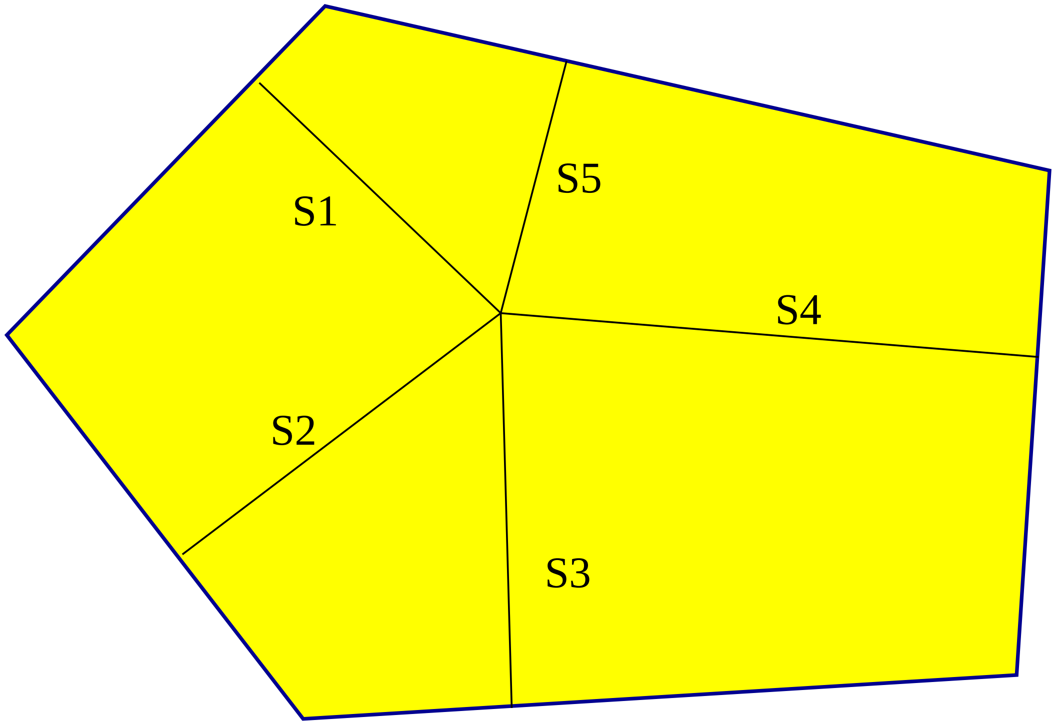

Figure 3: Analytic center maximizes the product of slacks.

LP with Barrier Function

Consider the LP problem with the barrier function

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

and

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

where μ is called the barrier (weight) parameter.

They are again linearly constrained convex programs (LCCP).

Know how to derive the KKT conditions of the problem.

Common Optimality Conditions for LPB and LDB

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

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.

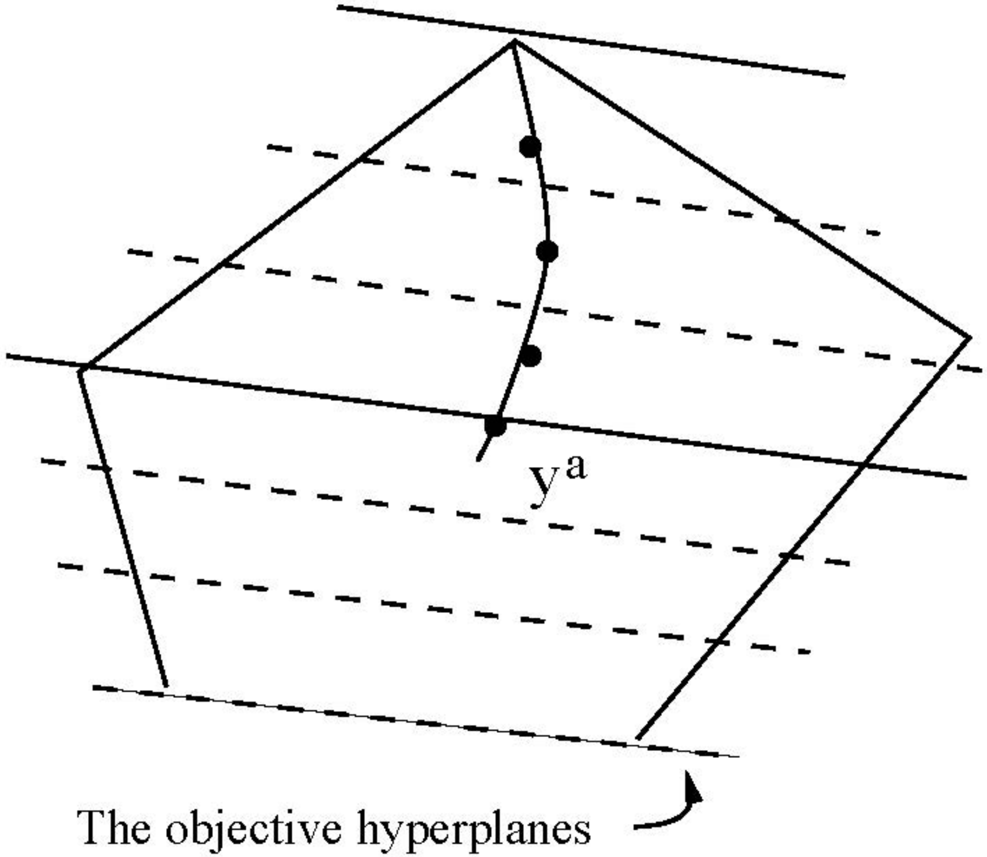

Figure 4: The central path of $\mathbf{y}(\mu)$ in a dual feasible region.

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 10 *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.*

Potential Function for Linear Programming

For $\mathbf{x} \in \operatorname{int} \mathcal{F}_p$ and $(\mathbf{y}, \mathbf{s}) \in \operatorname{int} \mathcal{F}_d$, the 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 we have a pair $(\mathbf{x}, \mathbf{y}, \mathbf{s}) \in \operatorname{int} \mathcal{F}$ with $\mu = \mathbf{x}^T\mathbf{s}/n$, we can generate a new iterate $\mathbf{x}^+$ and $(\mathbf{y}^+, \mathbf{s}^+)$ by solving for $\mathbf{d}_x$, $\mathbf{d}_y$ and $\mathbf{d}_s$ from the system of linear equations:

$$
\begin{aligned}
S\mathbf{d}_x + X\mathbf{d}_s &= \mathbf{r} := \frac{\mathbf{x}^T\mathbf{s}}{n+\rho}\mathbf{e} - X\mathbf{s}, \\
A\mathbf{d}_x &= \mathbf{0}, \\
-A^T\mathbf{d}_y - \mathbf{d}_s &= \mathbf{0}.
\end{aligned}
\tag{1}
$$

Let $\mathbf{d} := (\mathbf{d}_x, \mathbf{d}_y, \mathbf{d}_s)$. To show the dependence of $\mathbf{d}$ on the current pair $(\mathbf{x}, \mathbf{s})$ and the parameter γ, we write $\mathbf{d} = \mathbf{d}(\mathbf{x}, \mathbf{s}, \gamma)$. Note that $\mathbf{d}_x^T\mathbf{d}_s = -\mathbf{d}_x^T A^T\mathbf{d}_y = 0$ here. Th results still hold even if $\mathbf{d}_x^T\mathbf{d}_s \geq 0$.

Lemma 1 *Let the direction* $\mathbf{d} = (\mathbf{d}_x, \mathbf{d}_y, \mathbf{d}_s)$ *be generated by equation (1), and let*

$$\theta = \frac{\alpha\sqrt{\min(X\mathbf{s})}}{\|(XS)^{-1/2}(\frac{\mathbf{x}^T\mathbf{s}}{(n+\rho)}\mathbf{e} - X\mathbf{s})\|}\,, \tag{2}$$

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

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

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

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

A pair of LP has two alternatives

(Solvable)

$$\begin{aligned} A\mathbf{x} - \mathbf{b} &= \mathbf{0} \\ -A^T\mathbf{y} + \mathbf{c} &\geq \mathbf{0}, \\ \mathbf{b}^T\mathbf{y} - \mathbf{c}^T\mathbf{x} &= 0, \\ \mathbf{y} \text{ free}, \ \mathbf{x} &\geq \mathbf{0} \end{aligned}$$

or

(Infeasible)

$$\begin{aligned} A\mathbf{x} &= \mathbf{0} \\ -A^T\mathbf{y} &\geq \mathbf{0}, \\ \mathbf{b}^T\mathbf{y} - \mathbf{c}^T\mathbf{x} &> 0, \\ \mathbf{y} \text{ free}, \ \mathbf{x} &\geq \mathbf{0} \end{aligned}$$

An Integrated Homogeneous System

The two alternative systems can be homogenized as one:

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

where the two alternatives are

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

A HSD linear program

Let's try to add one more constraint to prevent the all-zero solution

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

Note that the constraints of (HSDP) form a skew-symmetric system and the objective coeffcient vector is the negative of the right-hand-side vector, so that it remains a self-dual linear program.

$(\mathbf{y} = \mathbf{0},\ \mathbf{x} = \mathbf{e},\ \tau = 1,\ \theta = 1)$ is a strictly feasible point for (HSDP).

The Augmented Lagrangian Function

But the above method still to find the null space of matrix A. One way to remove it is to construct an augmented Lagrangian function. In general consider

$$\min \quad f(\mathbf{x}) \quad \text{s.t.} \quad A\mathbf{x} = \mathbf{b},\ \mathbf{x} \in \mathcal{X}.$$

Let $\mathbf{y}$ be a temporary multipliers of the equality constraints. Then, the augmented Lagrangian function is

$$L(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}) - \mathbf{y}^T(A\mathbf{x} - \mathbf{b}) + \frac{\beta}{2}\|A\mathbf{x} - \mathbf{b}\|^2,\ \mathbf{x} \in \mathcal{X}.$$

Then, the iterative method would be, starting from a pair of $(\mathbf{x}^0, \mathbf{y}^0)$,

$$\mathbf{x}^{k+1} = \arg\min_{\mathbf{x}\in\mathcal{X}} L(\mathbf{x}, \mathbf{y}^k) \text{ and } \mathbf{y}^{k+1} = \mathbf{y}^k - \beta(A\mathbf{x}^{k+1} - \mathbf{b}).$$

For LP, $f(\mathbf{x}) = \mathbf{c}^T\mathbf{x}$ and $\mathcal{X} = \{\mathbf{x} :\ \mathbf{x} \geq \mathbf{0}\}$. Furthermore, by adding the barrier, we can consider

$$f(\mathbf{x}) = \mathbf{c}^T\mathbf{x} - \mu \sum_j \log(x_j),$$

for some small μ so that $\mathbf{x}$ is free.

Alternating Direction Method of Multipliers

The update of $\mathbf{x}$ still needs to inverse a large matrix. Let us split the $\mathbf{x}$ variables into two blocks, and consider

$$\min\{f_1(\mathbf{x}_1) + f_2(\mathbf{x}_2) \mid A_1\mathbf{x}_1 + A_2\mathbf{x}_2 = \mathbf{b},\ \mathbf{x}_1 \in \mathcal{X}_1, \mathbf{x}_2 \in \mathcal{X}_2.\}$$

where $f_1(\mathbf{x}_1)$ and $f_2(\mathbf{x}_2)$ are convex closed proper functions.

The Original ADMM (Glowinski & Marrocco '75, Gabay & Mercier '76) is provably working:

$$\begin{cases} \mathbf{x}_1^{k+1} = \arg\min\{L(\mathbf{x}_1, \mathbf{x}_2^k, \mathbf{y}^k) \mid \mathbf{x}_1 \in \mathcal{X}_1\}, \\ \mathbf{x}_2^{k+1} = \arg\min\{L(\mathbf{x}_1^{k+1}, \mathbf{x}_2, \mathbf{y}^k) \mid \mathbf{x}_2 \in \mathcal{X}_2\}, \\ \mathbf{y}^{k+1} = \mathbf{y}^k - \beta(A_1\mathbf{x}_1^{k+1} + A_2\mathbf{x}_2^{k+1} - \mathbf{b}), \end{cases}$$

where the augmented Lagrangian function L is defined as

$$L(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}) = \sum_{i=1}^{2} f_i(\mathbf{x}_i) - \mathbf{y}^T\Big(\sum_{i=1}^{2} A_i\mathbf{x}_i - \mathbf{b}\Big) + \frac{\beta}{2}\Big\|\sum_{i=1}^{2} A_i\mathbf{x}_i - \mathbf{b}\Big\|^2.$$

Variables Spliting

$$\min\left\{f_1(\mathbf{x}) + f_2(\mathbf{x}) \mid \mathbf{x} \in X.\right\}$$

can be reformulated as

$$\min\left\{f_1(\mathbf{x}_1) + f_2(\mathbf{x}_2) \mid \mathbf{x}_1 - \mathbf{x}_2 = \mathbf{0},\ \mathbf{x}_1 \in X.\right\}$$

so that the ADMM can be applied.

Often, the minimization of both $\mathbf{x}_1$ and $\mathbf{x}_2$ are easier.

Conic Linear Programming

$$
\begin{array}{lll}
(CLP) & \text{Minimize} & \mathbf{c}^T\mathbf{x} + C \bullet X \\
& \text{subject to} & \mathbf{a}_i^T\mathbf{x} + A_i \bullet X = b_i, i = 1, 2, ..., m, \\
& \mathbf{x} \in K_1,\ X \in K_2. &
\end{array}
$$

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

$$
\begin{array}{lll}
(CLD) & \text{Maximize} & \mathbf{b}^T\mathbf{y} \\
& \text{subject to} & \sum_i^m y_i\mathbf{a}_i + \mathbf{s} = \mathbf{c},\ \mathbf{s} \in K_1^*, \\
& & \sum_i^m y_i A_i + S = C,\ S \in K_2^*.
\end{array}
$$

Convex Cone Example

K	K^*
$\mathbf{R}^n$	$\mathbf{0}$
$\mathbf{R}^n_+$	$\mathbf{R}^n_+$
SOC	SOC
SDP	SDP
p-norm cone	q-norm cone