Interior Point Algorithms I

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

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

Methodological Philosophy

Recall that the primal Simplex Algorithm maintains the primal feasibility and complementarity while working toward dual feasibility. (The Dual Simplex Algorithm maintains dual feasibility and complementarity while working toward primal feasibility.)

In contrast, interior-point methods will 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 the simplex method is to make computer see corner points; and the key for interior-point methods is to stay in the interior of the feasible region.

Interior-Point Algorithms for LP

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

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

Let z^* denote the optimal value 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.

Barrier Functions for LP

Consider the barrier function optimization

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

and

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

They are linearly constrained convex programs (LCCP).

LCCP: Descent Direction

Let f be a differentiable function on R^n. If point $\bar{\mathbf{x}} \in R^n$ and there exists a vector $\mathbf{d}$ such that

$$\nabla f(\bar{\mathbf{x}})\mathbf{d} < 0,$$

then there exists a scalar $\bar{\tau} > 0$ such that

$$f(\bar{\mathbf{x}} + \tau\mathbf{d}) < f(\bar{\mathbf{x}}) \text{ for all } \tau \in (0, \bar{\tau}).$$

The vector $\mathbf{d}$ (above) is called a descent direction at $\bar{\mathbf{x}}$. If $\nabla f(\bar{\mathbf{x}}) \neq 0$, then $\nabla f(\bar{\mathbf{x}})$ is the direction of steepest ascent and $-\nabla f(\bar{\mathbf{x}})$ is the direction of steepest descent at $\bar{\mathbf{x}}$.

Denote by $\mathcal{D}^d_{\bar{\mathbf{x}}}$ the set of descent directions at $\bar{\mathbf{x}}$, that is,

$$\mathcal{D}^d_{\bar{\mathbf{x}}} = \{\mathbf{d} \in R^n : \nabla f(\bar{\mathbf{x}})\mathbf{d} < 0\}.$$

LCCP: Feasible Direction

At feasible point $\bar{\mathbf{x}}$, a feasible direction is

$$\mathcal{D}^f_{\bar{\mathbf{x}}} := \{\mathbf{d} \in R^n : \mathbf{d} \neq \mathbf{0},\ \bar{\mathbf{x}} + \lambda \mathbf{d} \in \mathcal{F} \text{ for all small } \lambda > 0\}.$$

Examples:

$$\mathcal{F} = R^n \Rightarrow \mathcal{D}^f_{\bar{\mathbf{x}}} = R^n.$$

$$\mathcal{F} = \{\mathbf{x} :\ A\mathbf{x} = \mathbf{b}\} \Rightarrow \mathcal{D}^f_{\bar{\mathbf{x}}} = \{\mathbf{d} : A\mathbf{d} = \mathbf{0}\}.$$

$$\mathcal{F} = \{\mathbf{x} :\ A\mathbf{x} = \mathbf{b},\ \mathbf{x} \geq \mathbf{0}\} \Rightarrow \mathcal{D}^f_{\bar{\mathbf{x}}} = \{\mathbf{d} : A\mathbf{d} = \mathbf{0},\ d_j \geq 0,\ \forall j \in \mathcal{A}(\bar{\mathbf{x}})\},$$

where the active or binding constraint set $\mathcal{A}(\bar{\mathbf{x}}) := \{j :\ \bar{x}_j = 0\}$.

Optimality Conditions for LCCP

What are the necessary conditions in order to have $\bar{\mathbf{x}}$ as a local optimizer for LCCP?

A general answer: the intersection of the descent and feasible direction sets at $\bar{x}$, $\mathcal{D}^d_{\bar{\mathbf{x}}}$ and $\mathcal{D}^f_{\bar{\mathbf{x}}}$, must be empty.

This condition is also sufficient if $f(\mathbf{x})$ is a convex function.

Unconstrained Problems

Consider the unconstrained problem, where f is differentiable on R^n,

$$(\text{UP}) \quad \begin{array}{ll} \text{minimize} & f(\mathbf{x}) \\ \text{subject to} & \mathbf{x} \in R^n. \end{array}$$

$\mathcal{D}^f_{\bar{\mathbf{x}}} = R^n$, so that $\mathcal{D}^d_{\bar{\mathbf{x}}} = \{\mathbf{d} \in R^n : \nabla f(\bar{\mathbf{x}})\mathbf{d} < 0\} = \emptyset$:

Theorem 1 *Let $\bar{x}$ be a (local) minimizer of (UP). If the functions f is continuously differentiable at $\bar{x}$, then*

$$\nabla f(\bar{\mathbf{x}}) = \mathbf{0}.$$

Linear Equality-Constrained Problems

Consider the linear equality-constrained problem, where f is differentiable on R^n,

$$(\text{LEP}) \quad \begin{array}{ll} \text{minimize} & f(\mathbf{x}) \\ \text{subject to} & A\mathbf{x} = \mathbf{b}. \end{array}$$

Theorem 2 *Let $\bar{x}$ be a (local) minimizer of (LEP). If the functions f is continuously differentiable at $\bar{x}$, then*

$$\nabla f(\bar{\mathbf{x}})^T - A^T\mathbf{y} = \mathbf{0}$$

for some $\mathbf{y} = (\bar{y}_1; \ldots; \bar{y}_m) \in R^m$, which are called Lagrange or dual multipliers.

The geometric interpretation: the objective gradient vector is perpendicular to the constraint hyperplanes.

Proof

$$F = \{\mathbf{x} : \ A\mathbf{x} = \mathbf{b}\} \Rightarrow \mathcal{D}^f_{\bar{\mathbf{x}}} = \{\mathbf{d} : A\mathbf{d} = 0\}.$$

If $\bar{\mathbf{x}}$ is a local optimizer, then the intersection of the descent and feasible direction sets at $\bar{x}$ must be empty or

$$\bar{A}\mathbf{d} = \mathbf{0}, \ \nabla f(\bar{\mathbf{x}})\mathbf{d} \neq 0$$

has no feasible solution. By the Alternative System Theorem it must be true that its alternative system has a solution, that is, there is $\mathbf{y} \in R^m$ such that

$$\nabla f(\bar{\mathbf{x}}) = \mathbf{y}^T A = \sum_{i=1}^{m} \bar{y}_i A_i.$$

Analytic Center for the Primal Polytope

The maximizer $\mathbf{x}^a$ of (PB) is called the analytic center of polytope $\mathcal{F}_p$. From the optimality condition theorem, we have

$$-(X^a)^{-1}\mathbf{e} - A^T\mathbf{y} = \mathbf{0},\ A\mathbf{x}^a = \mathbf{b},\ \mathbf{x}^a > \mathbf{0}.$$

or

$$\begin{array}{rcl} X^a\mathbf{s} & = & \mathbf{e} \\ A\mathbf{x}^a & = & \mathbf{b} \\ -A^T\mathbf{y} - \mathbf{s} & = & \mathbf{0} \\ \mathbf{x}^a & > & \mathbf{0}. \end{array} \tag{1}$$

Analytic Center for the Dual Polytope

The maximizer $(\mathbf{y}^a, \mathbf{s}^a)$ of (DB) is called the analytic center of polytope $\mathcal{F}_d$, and we have

$$\begin{array}{rcl} S^a\mathbf{x} & = & \mathbf{e} \\ A\mathbf{x} & = & \mathbf{0} \\ -A^T\mathbf{y}^a - \mathbf{s}^a & = & -\mathbf{c} \\ \mathbf{s}^a & > & \mathbf{0}. \end{array} \qquad (2)$$

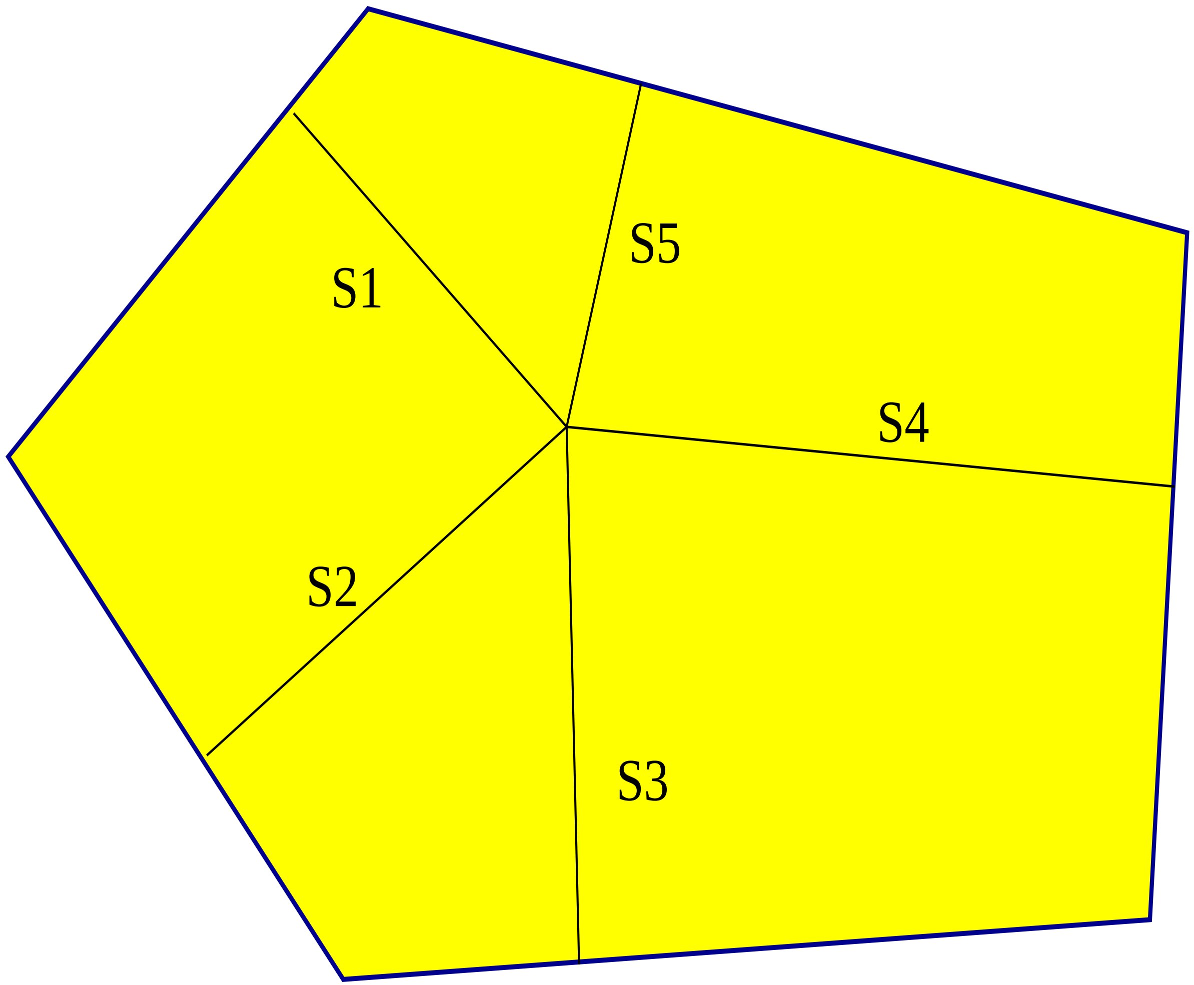

Figure 1: Analytic center maximizes the barrier function.

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).

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.

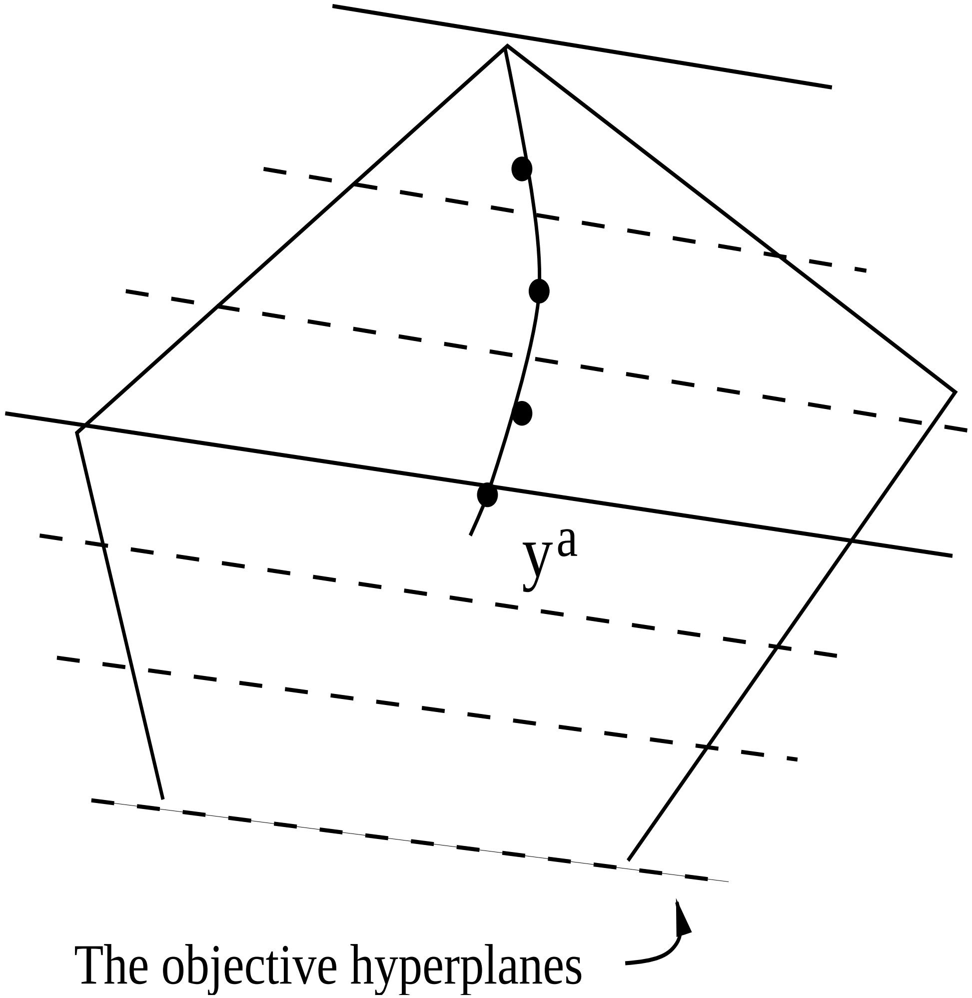

Figure 2: The central path of $\mathbf{y}(\mu)$ in a dual feasible region.

Central Path for Linear Programming

The path

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

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

Theorem 3 *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.*

Central Path Properties

Theorem 4 *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 $\{1, 2, ..., n\}$.*

Proof of (i)

$$(\mathbf{x}(\mu^0) - \mathbf{x}(\mu))^T(\mathbf{s}(\mu^0) - \mathbf{s}(\mu)) = 0,$$

since $(\mathbf{x}(\mu^0) - \mathbf{x}(\mu)) \in \mathcal{N}(A)$ and $(\mathbf{s}(\mu^0) - \mathbf{s}(\mu)) \in \mathcal{R}(A^T)$. This can be rewritten as

$$\sum_j^n \left(s(\mu^0)_j x(\mu)_j + x(\mu^0)_j s(\mu)_j \right) = n(\mu^0 + \mu) \leq 2n\mu^0,$$

or

$$\sum_j^n \left(\frac{x(\mu)_j}{x(\mu^0)_j} + \frac{s(\mu)_j}{s(\mu^0)_j} \right) \leq 2n.$$

Thus, $\mathbf{x}(\mu)$ and $\mathbf{s}(\mu)$ are bounded, which proves (i).

Proof of (iii)

Since $\mathbf{x}(\mu)$ and $\mathbf{s}(\mu)$ are both bounded, they have at least one limit point which we denote by $\mathbf{x}(0)$ and $\mathbf{s}(0)$. Let $\mathbf{x}^*_{P^*}$ ($\mathbf{x}^*_{Z^*} = \mathbf{0}$) and $\mathbf{s}^*_{Z^*}$ ($\mathbf{s}^*_{P^*} = \mathbf{0}$), respectively, be any strictly complementary solution pair on the primal and dual optimal faces: $\{\mathbf{x}_{P^*} : \; A_{P^*}\mathbf{x}_{P^*} = \mathbf{b}, \; \mathbf{x}_{P^*} \geq \mathbf{0}\}$ and $\{\mathbf{s}_{Z^*} : \; \mathbf{s}_{Z^*} = \mathbf{c}_{Z^*} - A^T_{Z^*}\mathbf{y} \geq \mathbf{0}, \; \mathbf{c}_{P^*} - A^T_{P^*}\mathbf{y} = \mathbf{0}\}$. Again, we have

$$\sum_j^n \left(s^*_j x(\mu)_j + x^*_j s(\mu)_j\right) = n\mu,$$

or

$$\sum_{j\in P^*} \left(\frac{x^*_j}{x(\mu)_j}\right) + \sum_{j\in Z^*} \left(\frac{s^*_j}{s(\mu)_j}\right) = n.$$

Thus, we have

$$x(\mu)_j \geq x^*_j/n > 0, \; j \in P^*$$

and

$$s(\mu)_j \geq s_j^*/n > 0, \ j \in Z^*.$$

This implies that

$$x(\mu)_j \to 0, \ j \in Z^*$$

and

$$s(\mu)_j \to 0, \ j \in P^*.$$

The Path-Following Algorithm

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

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

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

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

Primal-Dual Potential Function for LP

For $\mathbf{x} \in \text{int}\,\mathcal{F}_p$ and $(\mathbf{y}, \mathbf{s}) \in \text{int}\,\mathcal{F}_d$, the joint primal-dual potential function is defined by

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

where $\rho \geq 0$.

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

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

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

Primal-Dual Potential Reduction Algorithm for LP

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

$$\begin{aligned} S\mathbf{d}_x + X\mathbf{d}_s &= \mathbf{r} := \gamma\mu\mathbf{e} - X\mathbf{s}, \\ A\mathbf{d}_x &= \mathbf{0}, \\ -A^T\mathbf{d}_y - \mathbf{d}_s &= \mathbf{0}. \end{aligned} \tag{3}$$

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.

$$\begin{aligned}\mathbf{d}_{x'} + \mathbf{d}_{s'} &= \mathbf{r}' := (XS)^{-1/2}(\gamma\mu\mathbf{e} - X\mathbf{s}),\\ A'\mathbf{d}_{x'} &= \mathbf{0},\\ -(A')^T\mathbf{d}_y - \mathbf{d}_{s'} &= \mathbf{0}.\end{aligned}$$

where

$$D = X^{1/2}X^{-1/2},\; A' = AD,\; \mathbf{d}_{x'} = D^{-1}\mathbf{d}_x,\; \mathbf{d}_{s'} = D\mathbf{d}_s.$$

One can first compute

$$-A'(A')^T\mathbf{d}_y = A'\mathbf{r}'$$

and then solve for $\mathbf{d}_x$ and $\mathbf{d}_s$.

The role of γ

If $\gamma = 0$, it steps toward the optimal solution characterized by the LP optimality condition; if $\gamma = 1$, it steps toward the central path point $(\mathbf{x}(\mu), \mathbf{y}(\mu), \mathbf{s}(\mu))$; 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.

Logarithmic Approximation Lemma

We first present a technical lemma:

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

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

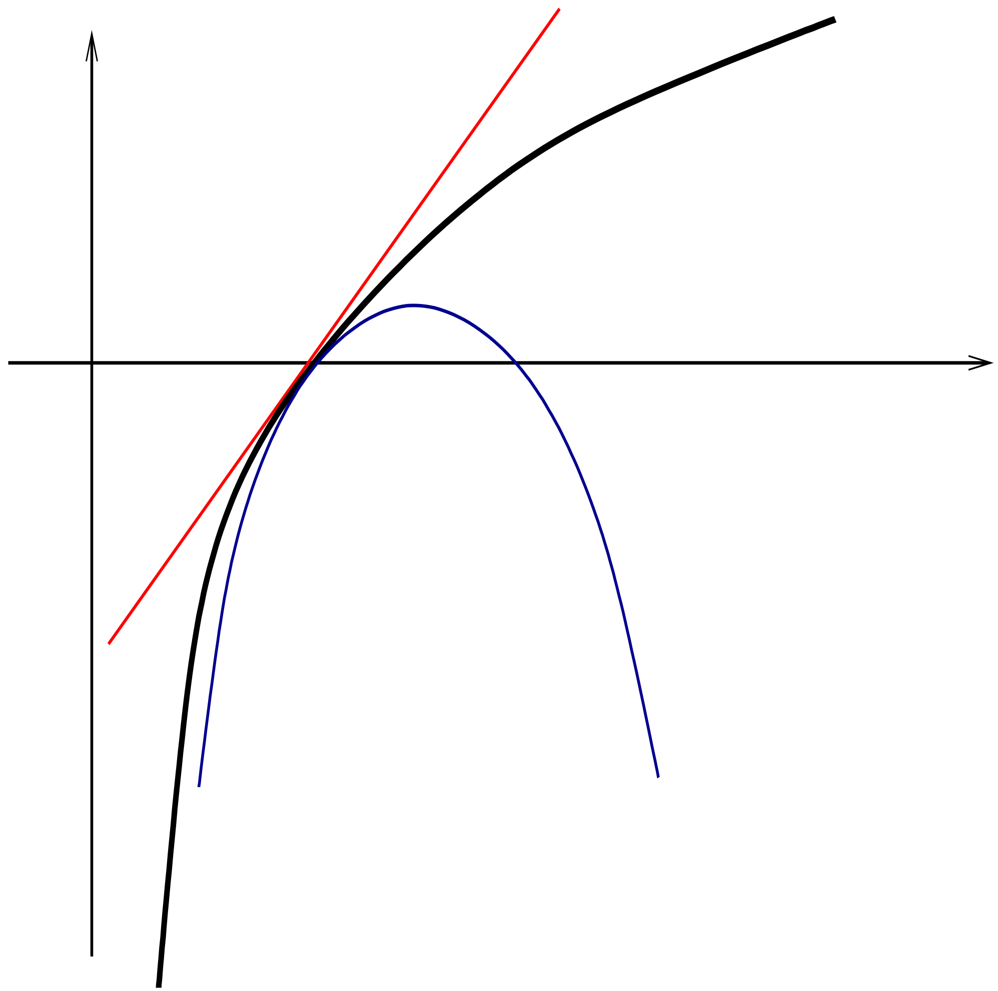

Figure 3: Logarithmic approximation by linear and quadratic functions

Lemma 2 *Let the direction vector* $\mathbf{d} = (\mathbf{d}_x, \mathbf{d}_y, \mathbf{d}_s)$ *be generated by equation (3) with* $\gamma = n/(n+\rho)$*, and let*

$$\theta = \frac{\alpha\sqrt{\min(X\mathbf{s})}}{\|(XS)^{-1/2}(\frac{\mathbf{x}^T\mathbf{s}}{(n+\rho)}\mathbf{e} - X\mathbf{s})\|}\ , \tag{4}$$

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

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

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

Let $\mathbf{v} = X\mathbf{s}$. Then, we can prove the following technical lemma:

Lemma 3 *Let* $\mathbf{v} \in \mathcal{R}^n$ *be a positive vector and* $\rho \geq \sqrt{n}$. *Then,*

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

Combining these two lemmas we have

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

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

for a constant δ.

Description of Algorithm

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

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

1. Set $(\mathbf{x}, \mathbf{s}) = (\mathbf{x}^k, \mathbf{s}^k)$ and $\gamma = n/(n+\rho)$ and compute $(\mathbf{d}_x, \mathbf{d}_y, \mathbf{d}_s)$ from (3).

2. Let $\mathbf{x}^{k+1} = \mathbf{x}^k + \bar{\alpha}\mathbf{d}_x$, $\mathbf{y}^{k+1} = \mathbf{y}^k + \bar{\alpha}\mathbf{d}_y$, and $\mathbf{s}^{k+1} = \mathbf{s}^k + \bar{\alpha}\mathbf{d}_s$ where
$$\bar{\alpha} = \arg\min_{\alpha \geq 0} \psi_{n+\rho}(\mathbf{x}^k + \alpha \mathbf{d}_x, \mathbf{s}^k + \alpha \mathbf{d}_s).$$

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

Theorem 5 *Let* $\rho \geq \sqrt{n}$ *and* $\psi_{n+\rho}(\mathbf{x}^0, \mathbf{s}^0) \leq \rho \log((\mathbf{x}^0)^T \mathbf{s}^0) + n \log n$. *Then, the Algorithm terminates in at most* $O(\rho \log((\mathbf{x}^0)^T \mathbf{s}^0 / \epsilon))$ *iterations with*

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

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

The role of ρ? And aggressive step size?

Adaptive Path-Following Algorithms

Here we describe and analyze a predictor-corrector interior-point algorithm for linear programming.

Consider the neighborhood

$$\mathcal{N}_2(\eta) = \left\{ (\mathbf{x}, \mathbf{s}) \in \operatorname{int} \mathcal{F} : \|X\mathbf{s} - \mu\mathbf{e}\| \le \eta\mu \quad \text{where} \quad \mu = \frac{\mathbf{x}^T\mathbf{s}}{n} \right\}$$

for some $\eta \in (0, 1)$. We will first analyze an adaptive-step path-following algorithm that generates a sequence of iterates in $\mathcal{N}_2(1/4)$. Actually, it also generates intermediate iterates in $\mathcal{N}_2(1/2)$.

Having obtained the search direction d, we let

$$
\begin{aligned}
\mathbf{x}(\theta) &:= \mathbf{x} + \theta \mathbf{d}_x, \\
\mathbf{y}(\theta) &:= \mathbf{y} + \theta \mathbf{d}_y, \\
\mathbf{s}(\theta) &:= \mathbf{s} + \theta \mathbf{d}_s.
\end{aligned} \tag{5}
$$

We will frequently let the next iterate be $(\mathbf{x}^+, \mathbf{s}^+) = (\mathbf{x}(\bar{\theta}), \mathbf{s}(\bar{\theta}))$, where $\bar{\theta}$ is as large as possible so that $(\mathbf{x}(\theta), \mathbf{s}(\theta))$ remains in the neighborhood $\mathcal{N}$ for $\theta \in [0, \bar{\theta}]$.

Let $\mu(\theta) = \mathbf{x}(\theta)^T \mathbf{s}(\theta)/n$ and $X(\theta) = \mathrm{diag}(\mathbf{x}(\theta))$. In order to get bounds on $\bar{\theta}$, we first note that

$$\begin{array}{rcl} \mu(\theta) & = & (1-\theta)\mu + \theta\gamma\mu, \\ X(\theta)\mathbf{s}(\theta) - \mu(\theta)\mathbf{e} & = & (1-\theta)(X\mathbf{s} - \mu\mathbf{e}) + \theta^2 D_x \mathbf{d}_s, \end{array}$$

where $D_x = \mathrm{diag}(\mathbf{d}_x)$.

Hence we can usually choose a larger $\bar{\theta}$ (and get a larger decrease in the duality gap) if $D_x \mathbf{d}_s$ is smaller.

First, it is helpful to re-express $D_x\mathbf{d}_s$. Let

$$\begin{aligned} \mathbf{p} &:= X^{-.5}S^{.5}\mathbf{d}_x, \\ \mathbf{q} &:= X^{.5}S^{-.5}\mathbf{d}_s, \\ \mathbf{r} &:= (XS)^{-.5}(\gamma\mu\mathbf{e} - X\mathbf{s}), \end{aligned} \tag{6}$$

Note that $\mathbf{p}+\mathbf{q}=\mathbf{r}$ and $\mathbf{p}^T\mathbf{q}=0$ so that $\mathbf{p}$ and $\mathbf{q}$ represent an orthogonal decomposition of $\mathbf{r}$.

Lemma 4 $\mathbf{p}$ *and* $\mathbf{q}$ *of (6) have:*

i)

$$\|P\mathbf{q}\| \leq \frac{\sqrt{2}}{4}\|\mathbf{r}\|^2;$$

ii)

$$-\frac{\|\mathbf{r}\|^2}{4} \leq p_j q_j \leq \frac{r_j^2}{4} \text{ for each } j.$$

Lemma 5 *Let* $\mathbf{r}$ *be as above.*

i) *If* $\gamma = 0$*, then* $\|\mathbf{r}\|^2 = n\mu$*.*

ii) *If* $\eta \in (0, 1)$*,* $\gamma = 1$ *and* $(\mathbf{x}, \mathbf{s}) \in \mathcal{N}_2(\eta)$*, then* $\|\mathbf{r}\|^2 \leq \eta^2\mu/(1 - \eta)$*.*

Predictor-corrector algorithm

The Algorithm takes a single "corrector" step to the central path after each "predictor" step to decrease μ. Although it is possible to use more general values of η, we will work with nearly-centered pairs in $\mathcal{N}_2(\eta)$ with $\eta = 1/4$ (iterates after the corrector step), and intermediate pairs in $\mathcal{N}_2(2\eta)$ (iterates after a predictor step).

Given $(\mathbf{x}^0, \mathbf{s}^0) \in \mathcal{N}_2(\eta)$ with $\eta = 1/4$. Set $k := 0$.

While $(\mathbf{x}^k)^T\mathbf{s}^k > \epsilon$ **do**:

1. Predictor step: set $(\mathbf{x}, \mathbf{s}) = (\mathbf{x}^k, \mathbf{s}^k)$ and compute $\mathbf{d} = \mathbf{d}(\mathbf{x}, \mathbf{s}, 0)$ from (3); compute the largest $\bar{\theta}$ so that

$$(\mathbf{x}(\theta), \mathbf{s}(\theta)) \in \mathcal{N}_2(2\eta) \;\text{ for }\; \theta \in [0, \bar{\theta}].$$

2. Corrector step: set $(\mathbf{x}', \mathbf{s}') = (\mathbf{x}(\bar{\theta}), \mathbf{s}(\bar{\theta}))$ and compute $\mathbf{d}' = \mathbf{d}(\mathbf{x}', \mathbf{s}', 1)$ from (3); set $(\mathbf{x}^{k+1}, \mathbf{s}^{k+1}) = (\mathbf{x}' + \mathbf{d}'_x, \mathbf{s}' + \mathbf{d}'_s)$.

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

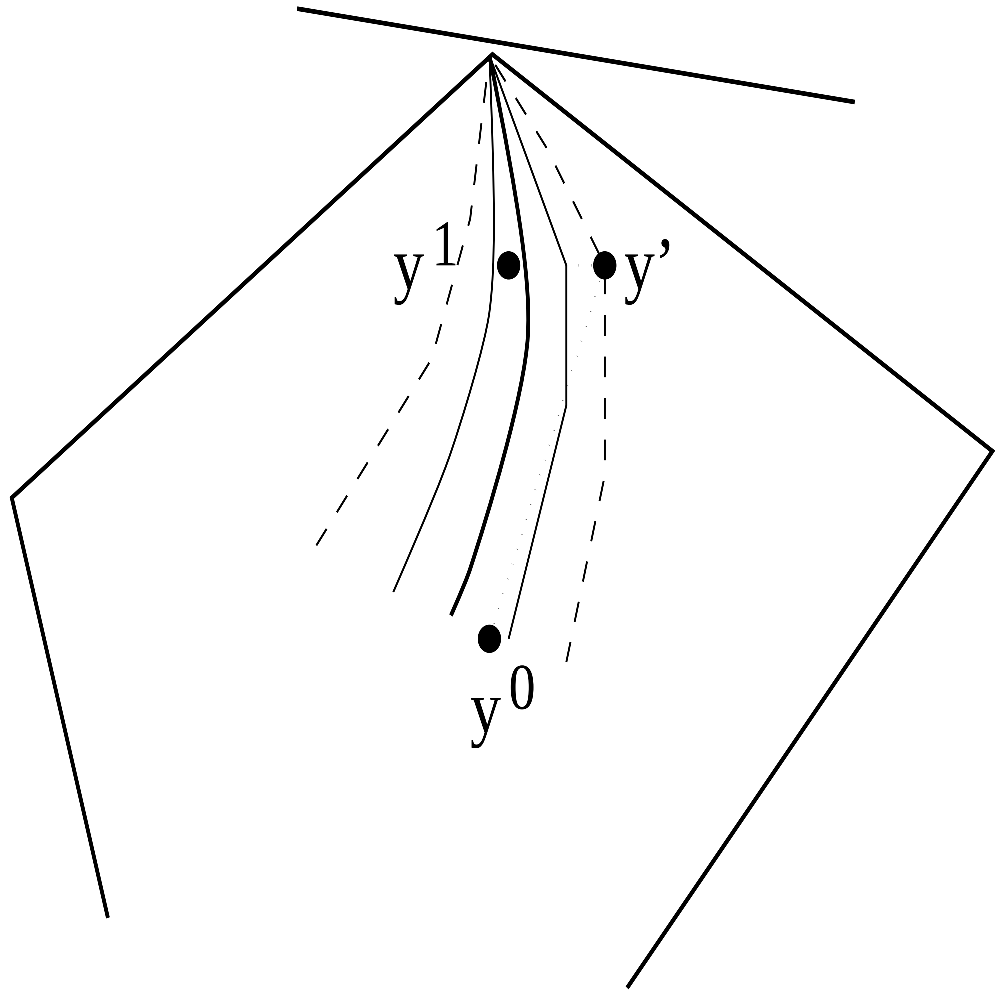

Figure 4: Illustration of the predictor-corrector algorithm.

Lemma 6 *For each k, $(\mathbf{x}^k, \mathbf{s}^k) \in \mathcal{N}_2(\eta)$.*

Now let $(\mathbf{x}, \mathbf{s}) = (\mathbf{x}^k, \mathbf{s}^k)$, $\mathbf{d} = \mathbf{d}(\mathbf{x}, \mathbf{s}, 0)$, $\mu = \mu^k = \mathbf{x}^T\mathbf{s}/n$, and $\mathbf{p}$, $\mathbf{q}$ and $\mathbf{r}$ be as in (6); these quantities all refer to the predictor step at iteration k. By (6),

$$\mu' = (1 - \bar{\theta})\mu, \quad \text{or}$$

$$\mu^{k+1} = (1 - \bar{\theta})\mu^k. \tag{7}$$

Hence the improvement in the duality gap at the kth iteration depends on the size of $\bar{\theta}$.

Lemma 7 *With the notation above, the step-size in the predictor step satisfies*

$$\bar{\theta} \geq \frac{2}{1 + \sqrt{1 + 4\|P\mathbf{q}/\mu\|/\eta}}.$$

Theorem 6 *Let $\eta = 1/4$. Then the Algorithm will terminate in at most $O(\sqrt{n}\log((\mathbf{x}^0)^T\mathbf{s}^0/\epsilon))$ iterations with*

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

Moreover, there is fixed positive number $\bar{\epsilon} < 1$, depending on $A, \mathbf{b}, \mathbf{c}$ only, such that, after $(\mathbf{x}^k)\mathbf{s}^k \le \bar{\epsilon}$,

$$(\mathbf{x}^{k+1})\mathbf{s}^{k+1} \le \left((\mathbf{x}^k)\mathbf{s}^k\right)^2.$$

Proof: Using Lemma 4(i) and Lemma 5(i), we have

$$\|P\mathbf{q}\| \le \frac{\sqrt{2}}{4}\|\mathbf{r}\|^2 = \frac{\sqrt{2}}{4}n\mu,$$

so that

$$\bar{\theta} \ge \frac{2}{1+\sqrt{1+\sqrt{2}n/\eta}} = \frac{2}{1+\sqrt{1+4\sqrt{2}n}}$$

at each iteration. Then (7) and Lemma 7 imply that

$$\mu^{k+1} \leq \left(1 - \frac{2}{1+\sqrt{1+4\sqrt{2}n}}\right)\mu^k$$

for each k. This yields the desired result of the complexity bound.

The proof of the quadratic convergence is omitted.