Interior Point Algorithms I: Geometric Explanation

Yinyu Ye

Department of Management Science and Engineering

Stanford University

Stanford, CA 94305, U.S.A.

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

Chapter 5.4-5.5

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

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

and

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

They are linearly constrained convex programs (LCCP).

Analytic Center for the Primal Polytope

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

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

or

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

Analytic Center for the Dual Polytope

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

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

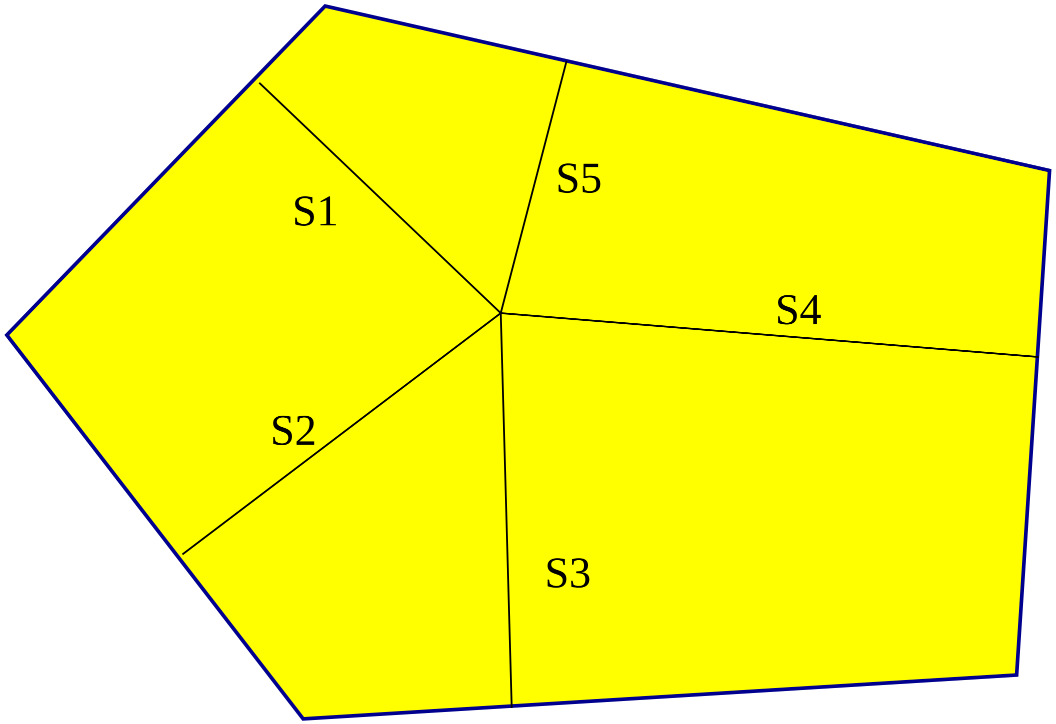

Figure 1: Analytic center maximizes the product of slacks.

Why Analytic

The analytic center of polytope $\mathcal{F}_d$ is an analytic function of input data $A, \mathbf{c}$.

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

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 $\bar{y} = n/(n+1)$ with $\mathbf{x} = ((n+1)/n,\ \cdots,\ (n+1)/n,\ (n+1))^T$.

Analytic Volume of Polytope and Cutting Plane

$$AV(\mathcal{F}_d) := \prod_{j=1}^{n} \bar{s}_j = \prod_{j=1}^{n} (c_j - \mathbf{a}_j^T \bar{\mathbf{y}})$$

can be viewed as the analytic volume of polytope $\mathcal{F}_d$ or simply $\mathcal{F}$ in the rest of discussions.

If one inequality in $\mathcal{F}$, say the first one, needs to be translated, change $\mathbf{a}_1^T \mathbf{y} \le c_1$ to $\mathbf{a}_1^T y \le \mathbf{a}_1^T \bar{\mathbf{y}}$; i.e., the first inequality is parallelly moved and it now cuts through $\bar{\mathbf{y}}$ and divides $\mathcal{F}$ into two bodies. Analytically, c_1 is replaced by $\mathbf{a}_1^T \bar{\mathbf{y}}$ and the rest of data are unchanged. Let

$$\mathcal{F}^+ := \{\mathbf{y} : \ \mathbf{a}_j^T \mathbf{y} \le c_j^+, \ j = 1, ..., n\},$$

where $c_j^+ = c_j$ for $j = 2, ..., n$ and $c_1^+ = \mathbf{a}_1^T \bar{\mathbf{y}}$.

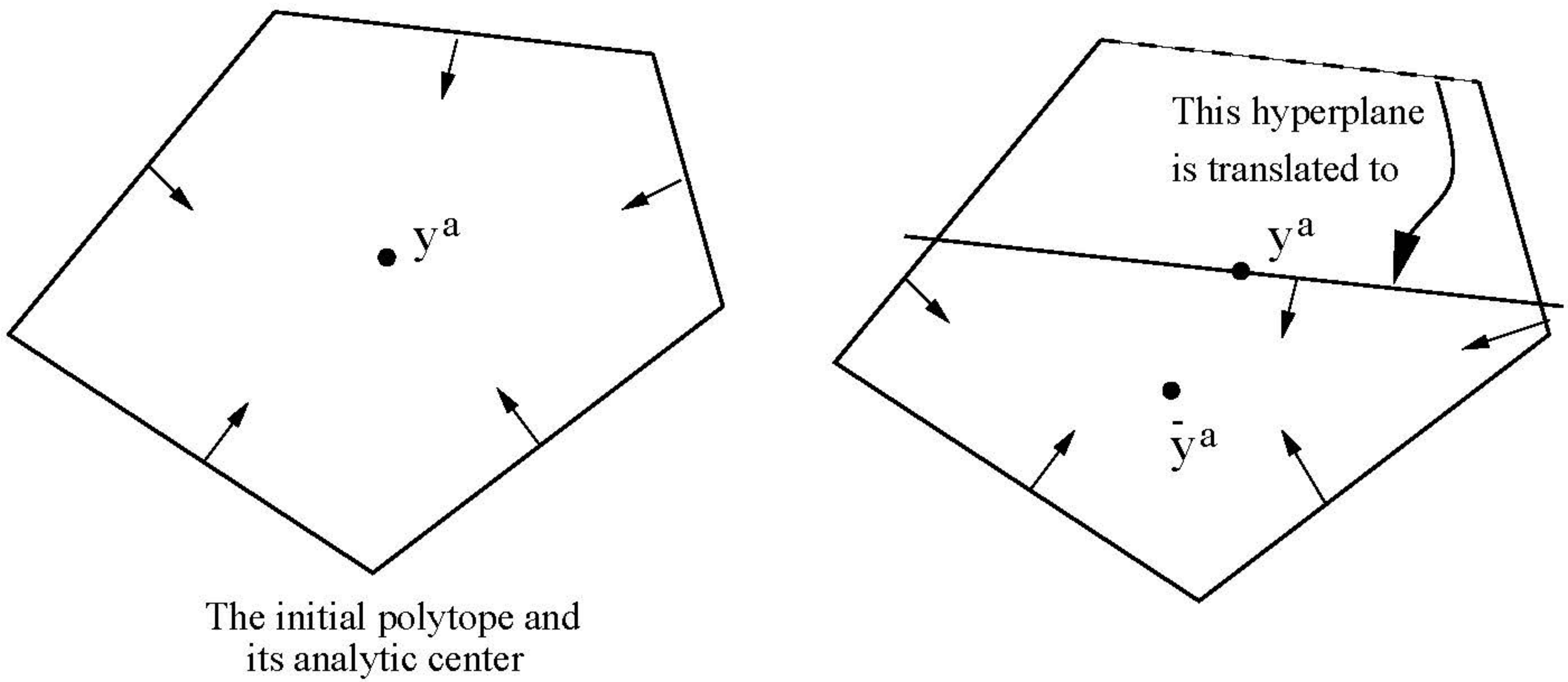

Figure 2: Translation of a hyperplane to the AC.

Analytic Volume Reduction of the New Polytope

Let $\bar{\mathbf{y}}^+$ be the analytic center of $\mathcal{F}^+$. Then, the analytic volume of $\mathcal{F}^+$

$$AV(\mathcal{F}^+) = \prod_{j=1}^{n}(c_j^+ - \mathbf{a}_j^T\bar{\mathbf{y}}^+) = (\mathbf{a}_1^T\bar{\mathbf{y}} - \mathbf{a}_1^T\bar{\mathbf{y}}^+)\prod_{j=2}^{n}(c_j - \mathbf{a}_j^T\bar{\mathbf{y}}^+).$$

We have the following volume reduction theorem:

Theorem 1

$$\frac{AV(\mathcal{F}^+)}{AV(\mathcal{F})} \le \exp(-1).$$

Proof

Since $\bar{\mathbf{y}}$ is the analytic center of $\mathcal{F}$, there exists $\bar{\mathbf{x}} > \mathbf{0}$ such that

$$\bar{X}\bar{\mathbf{s}} = \bar{X}(\mathbf{c} - A^T\bar{\mathbf{y}}) = \mathbf{e} \quad \text{and} \quad A\bar{\mathbf{x}} = \mathbf{0}.$$

Thus,

$$\bar{\mathbf{s}} = (\mathbf{c} - A^T\bar{\mathbf{y}}) = \bar{X}^{-1}\mathbf{e} \quad \text{and} \quad \mathbf{c}^T\bar{\mathbf{x}} = (\mathbf{c} - A^T\bar{\mathbf{y}})^T\bar{\mathbf{x}} = \mathbf{e}^T\mathbf{e} = n.$$

We have

$$\begin{aligned} \mathbf{e}^T\bar{X}\bar{\mathbf{s}}^+ &= \mathbf{e}^T\bar{X}(\mathbf{c}^+ - A^T\bar{\mathbf{y}}^+) = \mathbf{e}^T\bar{X}\mathbf{c}^+ \\ &= \mathbf{c}^T\bar{\mathbf{x}} - \bar{x}_1(c_1 - \mathbf{a}_1^T\bar{\mathbf{y}}) = n - 1. \end{aligned}$$

$$
\begin{aligned}
\frac{AV(\mathcal{F}^+)}{AV(\mathcal{F})} &= \prod_{j=1}^{n} \frac{\bar{s}_j^+}{\bar{s}_j} \\
&= \prod_{j=1}^{n} \bar{x}_j \bar{s}_j^+ \\
&\leq \left(\frac{1}{n} \sum_{j=1}^{n} \bar{x}_j \bar{s}_j^+ \right)^n \\
&= \left(\frac{1}{n} \mathbf{e}^T \bar{X} \bar{\mathbf{s}}^+ \right)^n \\
&= \left(\frac{n-1}{n} \right)^n \leq \exp(-1).
\end{aligned}
$$

Analytic Volume of Polytope and Multiple Cutting Planes

Now suppose we translate $k(< n)$ hyperplanes, say $1, 2, ..., k$, moved to cut the analytic center $\bar{\mathbf{y}}$ of $\mathcal{F}$, that is,

$$\mathcal{F}^+ := \{\mathbf{y} : \ \mathbf{a}_j^T \mathbf{y} \leq c_j^+, \ j = 1, ..., n\},$$

where $c_j^+ = c_j$ for $j = k+1, ..., n$ and $c_j^+ = \mathbf{a}_j^T \bar{\mathbf{y}}$ for $j = 1, ..., k$.

Corollary 1

$$\frac{AV(\mathcal{F}^+)}{AV(\mathcal{F})} \leq \exp(-k).$$

The Analytic Center Method Cutting-Plane Method

Problem: Find a solution in the feasible set $\mathcal{F} := \{\mathbf{y} : \ \mathbf{a}_j^T\mathbf{y} \le c_j, \ j = 1, ..., n\}$.

Start with the initial polytope

$$\mathcal{F}^0 := \{\mathbf{y} : \ \mathbf{a}_j^T\mathbf{y} \le c_j^0 := c_j + R, \ j = 1, ..., n\}$$

where R is sufficiently large such that $\bar{\mathbf{y}}^0 = \mathbf{0}$ is an (approximate) analytic center of $\mathcal{F}^0$.

Check if the (approximate) analytic center $\bar{\mathbf{y}}^k$ of $\mathcal{F}^k$ is in $\mathcal{F}$ or not. If not, define a new polytope $\mathcal{F}^{k+1}$ by translating one or multiple violated constraint hyperplanes through $\bar{\mathbf{y}}^k$ as defined earlier, and compute an approximate analytic center$\bar{\mathbf{y}}^{k+1}$ of $\mathcal{F}^{k+1}$.

Continue this step till $\bar{\mathbf{y}}^k \in \mathcal{F}$.

Trajectory of Analytic Centers: Central Path for LP

Now consider the problem

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

Assume that the feasible region is bounded, and the analytic center of the region is $\mathbf{y}^0$.

Start with a polytope

$$\mathcal{F}(R) := \{\mathbf{y} : \; A^T\mathbf{y} \le \mathbf{c}, \overbrace{\mathbf{b}^T\mathbf{y} \ge R, \cdots, \mathbf{b}^T\mathbf{y} \ge R}^{k \text{ times}}\}$$

where R is so low such that $\mathbf{y}^0$ is also an (approximate) analytic center of $\mathcal{F}(R)$.

Define a family of polytopes $\mathcal{F}(R)$ by continuously increasing R toward the maximal value and consider its analytic center $\mathbf{y}(R)$: it forms a path of analytic centers from $\mathbf{y}^0$ toward the optimal solution set.

Better Parameterization: LP with Barrier Function

An equivalent algebraic representation of the path is to consider the LP problem with the weighted barrier function

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

and also

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

where μ is called the barrier (weight) parameter.

They are again linearly constrained convex programs (LCCP).

Common Optimality Conditions for both LPB and LDB

They share the same first-order KKT conditions:

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

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.

Denote by $(\mathbf{x}(\mu), \mathbf{y}(\mu), \mathbf{s}(\mu))$ the (unique) solution satisfying the conditions. As μ decreases to zero, $\mathbf{x}(\mu)$ form a path in the primal feasible region and $\mathbf{y}(\mu)$ form a path in the dual feasible region to-warding optimality respectively.

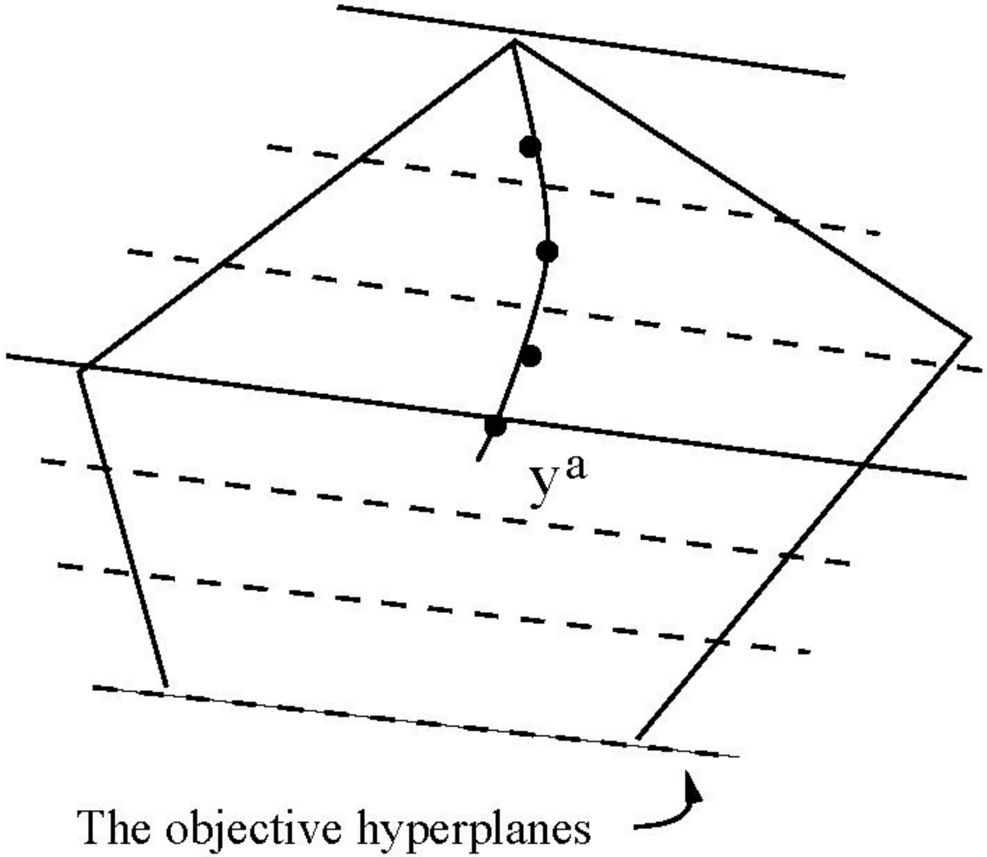

Figure 3: The central path of $\mathbf{y}(\mu)$ in a dual feasible region.

Central Path for Linear Programming

The path

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

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

Theorem 2 *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 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 $0 < \mu' < \mu$,*

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

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

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

or

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

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 Primal-Dual 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 at most $1 - \frac{1}{3\sqrt{n}}$, wheere k is the iteration count.

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

The Newton Method of Path-Following

Given a pair $(\mathbf{x}^k, \mathbf{y}^k, \mathbf{s}^k) \in \text{int}\,\mathcal{F}$ closely to the central path, that is,

$$\|X^k S^k \mathbf{e} - \mu^k \mathbf{e}\| \le \eta \mu^k$$

for a small positive constant η, we compute direction vectors $\mathbf{d}_x$, $\mathbf{d}_y$ and $\mathbf{d}_s$ from the system equations:

$$\begin{aligned} S^k \mathbf{d}_x + X^k \mathbf{d}_s &= (1 - \tfrac{1}{3\sqrt{n}})\mu^k \mathbf{e} - X^k S^k \mathbf{e} = \tfrac{-1}{3\sqrt{n}}\mu^k \mathbf{e} + (\mu^k \mathbf{e} - X^k S^k \mathbf{e}), \\ A\mathbf{d}_x &= \mathbf{0}, \\ -A^T \mathbf{d}_y - \mathbf{d}_s &= \mathbf{0}. \end{aligned} \tag{3}$$

(Note that $\mathbf{d}_x^T \mathbf{d}_s = -\mathbf{d}_x^T A^T \mathbf{d}_y = 0$ here.) Then we update

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \mathbf{d}_x > \mathbf{0},\ \mathbf{y}^{k+1} = \mathbf{y}^k + \mathbf{d}_y,\ \mathbf{s}^{k+1} = \mathbf{s}^k + \mathbf{d}_s > \mathbf{0}.$$

Then we can prove

$$\|X^{k+1} S^{k+1} \mathbf{e} - \mu^{k+1} \mathbf{e}\| \le \eta(1 - \frac{1}{3\sqrt{n}})\mu^k = \eta \mu^{k+1}.$$