Bootcamp: Interior Point Algorithms I

Yinyu Ye

Friday, September 01 2023

Simons Institute

LP Algorithms at a Glance

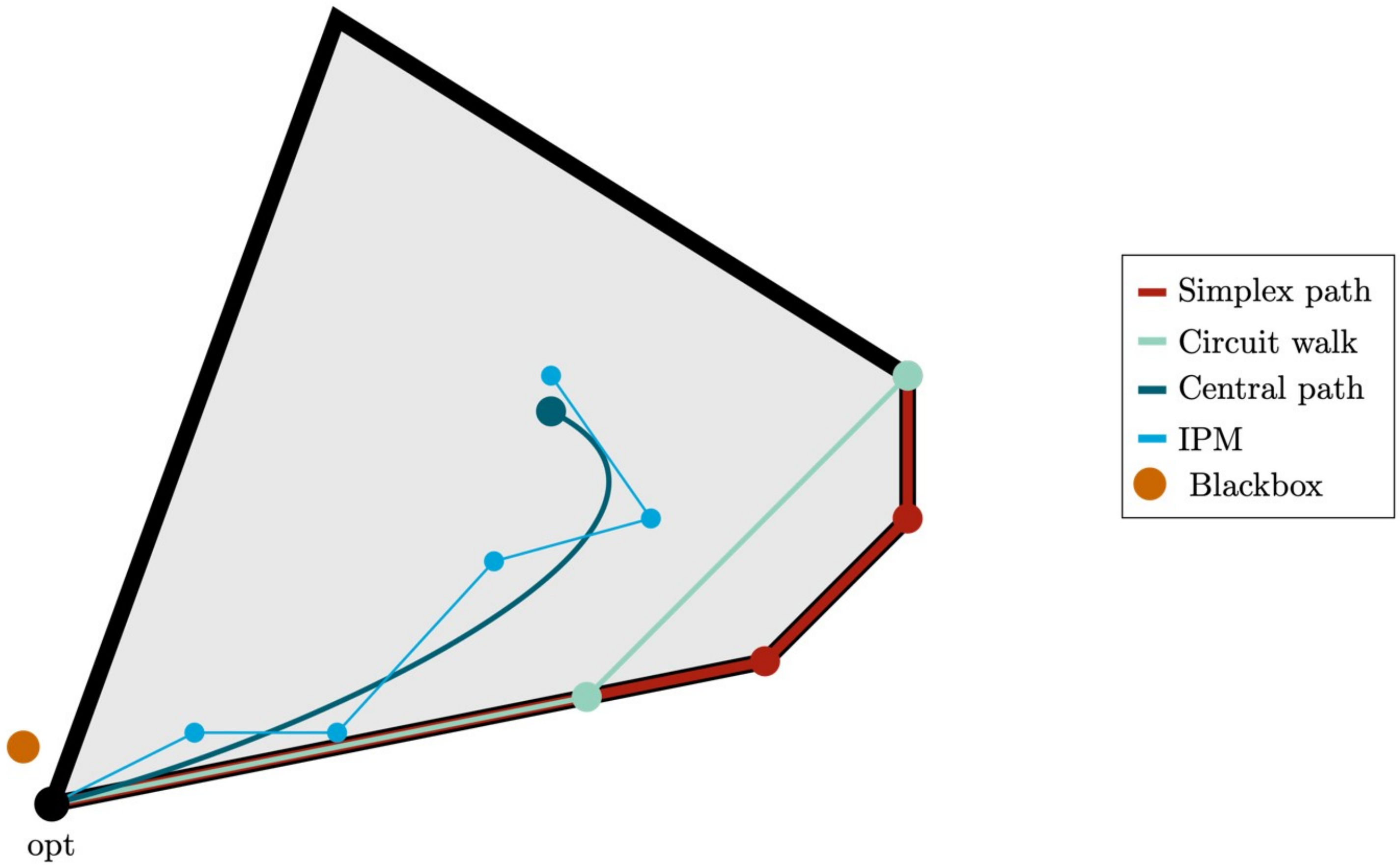

Figure 1: Slide from Daniel Dadush and Bento Natura 2023

Interior-Point Algorithms for LP

Consider linear program:

$$\min \quad \mathbf{c}^T\mathbf{x} \quad \text{s.t.} \quad A\mathbf{x} = \mathbf{b},\ \mathbf{x} \geq \mathbf{0}.$$

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

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

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.

Karmarkar 1984, Renegar 1986...

Logarithmic Barrier Functions for LP

Consider the logarithmic 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).

Much much earlier...

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

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

where $\mathbf{e}$ is the vector of all ones; or

$$\begin{aligned} \bar{X}\mathbf{s} &= \mathbf{e} \\ A\bar{\mathbf{x}} &= \mathbf{b} \\ -A^T\mathbf{y} - \mathbf{s} &= \mathbf{0} \\ \bar{\mathbf{x}} &> \mathbf{0} \end{aligned} \tag{1}$$

where X is the diagonal matrix generated from vector $\mathbf{x}$.

Sonnevend 1988, Bayer and Lagarias 1989, Megiddo 1989...

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

$$\begin{aligned} \bar{S}\mathbf{x} &= \mathbf{e} \\ A\mathbf{x} &= \mathbf{0} \\ -A^T\bar{\mathbf{y}} - \bar{\mathbf{s}} &= -\mathbf{c} \\ \bar{\mathbf{s}} &> \mathbf{0}. \end{aligned} \tag{2}$$

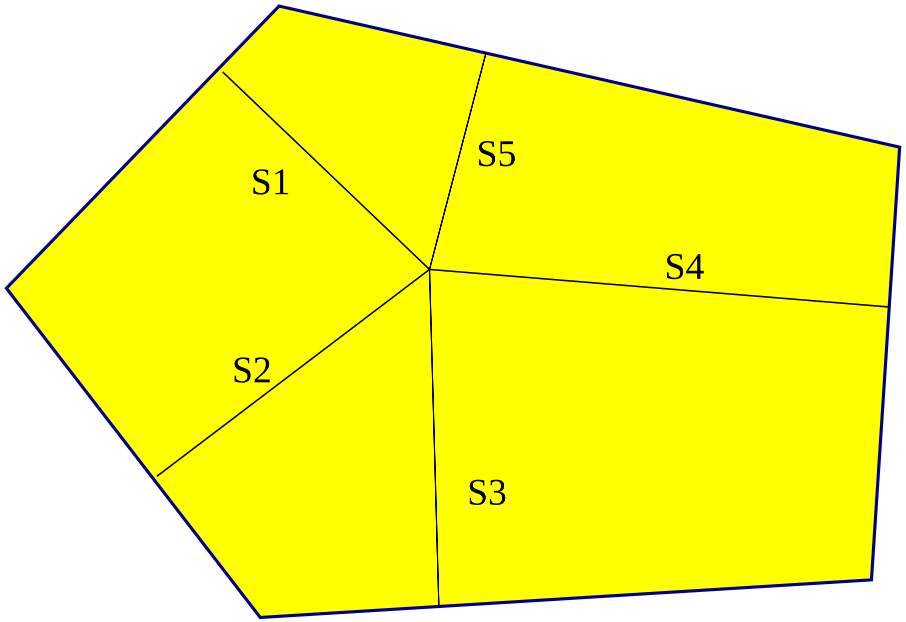

Figure 2: Analytic center maximizes the product of slacks.

Volumetric Center for the Dual Polytope

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

$$
\begin{array}{lll}
(DB) & \text{maximize} & -\log\det(AS^{-2}A^T) \\
& \text{s.t.} & (\mathbf{y}, \mathbf{s}) \in \operatorname{int} \mathcal{F}_d
\end{array}
$$

More details see Vaidya 1996, Lee-Sidford 13-'19, van den Brand, Lee, Liu, Saranurak, Sidford, Song, Wang 21, etc.

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 $c_1^+ = \mathbf{a}_1^T \bar{\mathbf{y}}$.

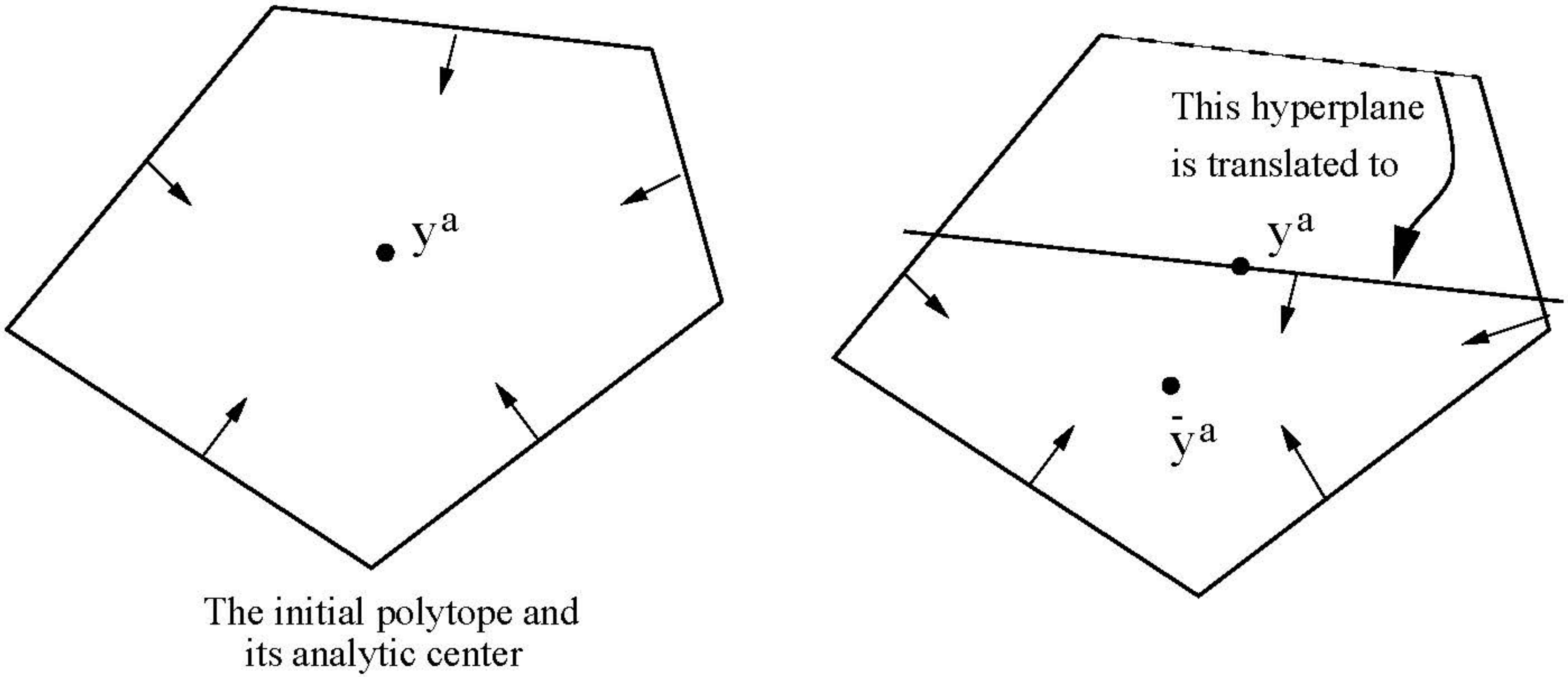

Figure 3: Translation of a hyperplane through the AC.

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 Cutting-Plane Method

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

Start with the initial polytope

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

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 $\bar{\mathbf{y}}^k \in \mathcal{F}$.

Fair Pareto Optimal Solutions of Multiple Objectives

Problem: Find a solution in the feasible set $\mathcal{F} := \{\mathbf{y} : \mathbf{a}_j^T\mathbf{y} \le c_j,\ j = 1, ..., n\}$ such that it is Pareto-Maximal for k objective function $\mathbf{b}_i^T\mathbf{y},\ i = 1, ..., k$

- Weighted-Sum Objective Maximization: $\sum_i w_i\mathbf{b}_i^T\mathbf{y}$
- Alternating Cuting-Plane: Cuting the AC alternatively, or simultaneously with fixed proportions?

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 Regularized by the Barrier Function

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

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

and also

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

where μ is called the barrier (weight) parameter.

They are again linearly constrained convex programs (LCCP).

Self-Duality of LPB and LDB

They share the same first-order KKT conditions:

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

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.

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.

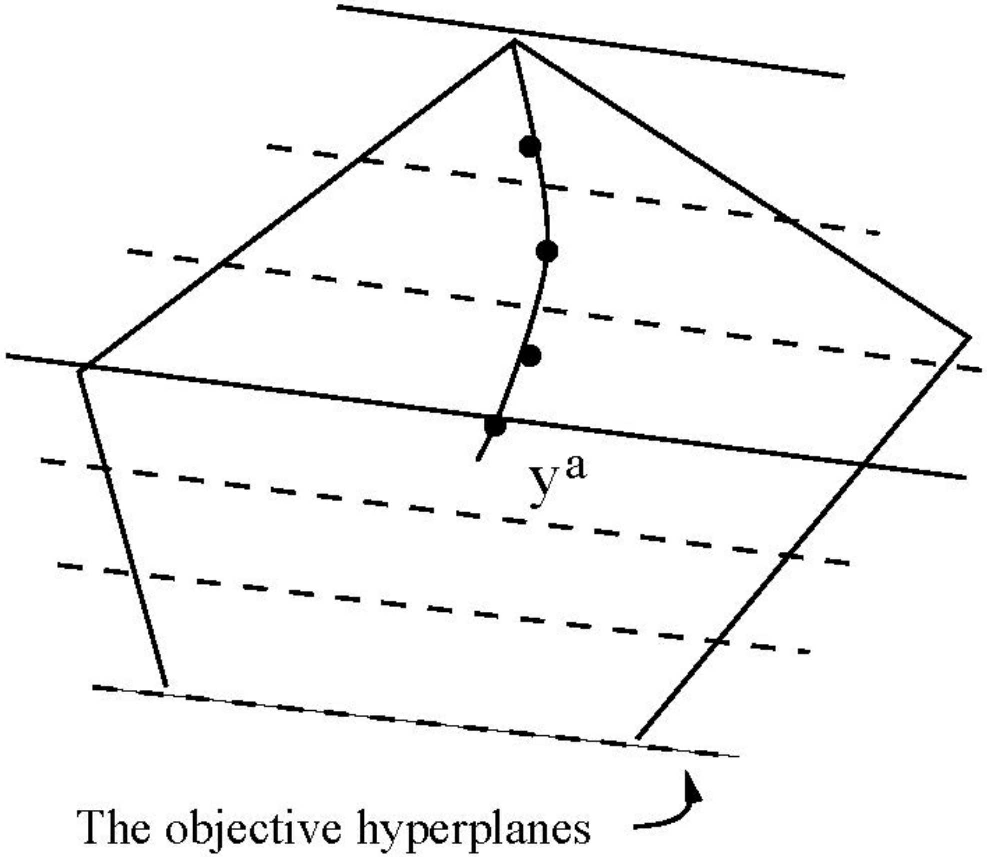

Figure 4: The central path of $\mathbf{y}(\mu)$ in a dual feasible region.

Central Path for Linear Programming Parametrized by μ

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

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

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^*) *are the analytic centers on the primal and dual optimal faces, respectively (Ǵuler and Y 1993).*

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).

The Path-Following Algorithms

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; see Renegar 1988, Gonzagar 1989, Kojima et al. 1989, Monteiro and Adler 1989,...

The Newton Method of Primal-Dual Path-Following

Given a pair $(\mathbf{x}^k, \mathbf{y}^k, \mathbf{s}^k) \in \operatorname{int} \mathcal{F}$ 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:

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

where $\gamma = (1 - \frac{1}{3\sqrt{n}})$. 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 one can prove

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

This leads to $\sqrt{n}$ iteration complexity.

Adaptive Path-Following Algorithms

Here we describe and analyze the Predictor-Corrector interior-point algorithm (Mizuno-Todd-Y 1993, Mehrotra 1993). Consider the neighborhood

$$\mathcal{N}_2(\eta) = \left\{ (\mathbf{x}, \mathbf{s}) \in \operatorname{int} \mathcal{F} : \|X\mathbf{s} - \mu\mathbf{e}\| \le \eta\mu \quad \text{where} \quad \mu = \frac{\mathbf{x}^T\mathbf{s}}{n} \right\} \text{for some } \eta \in (0,1).$$

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.

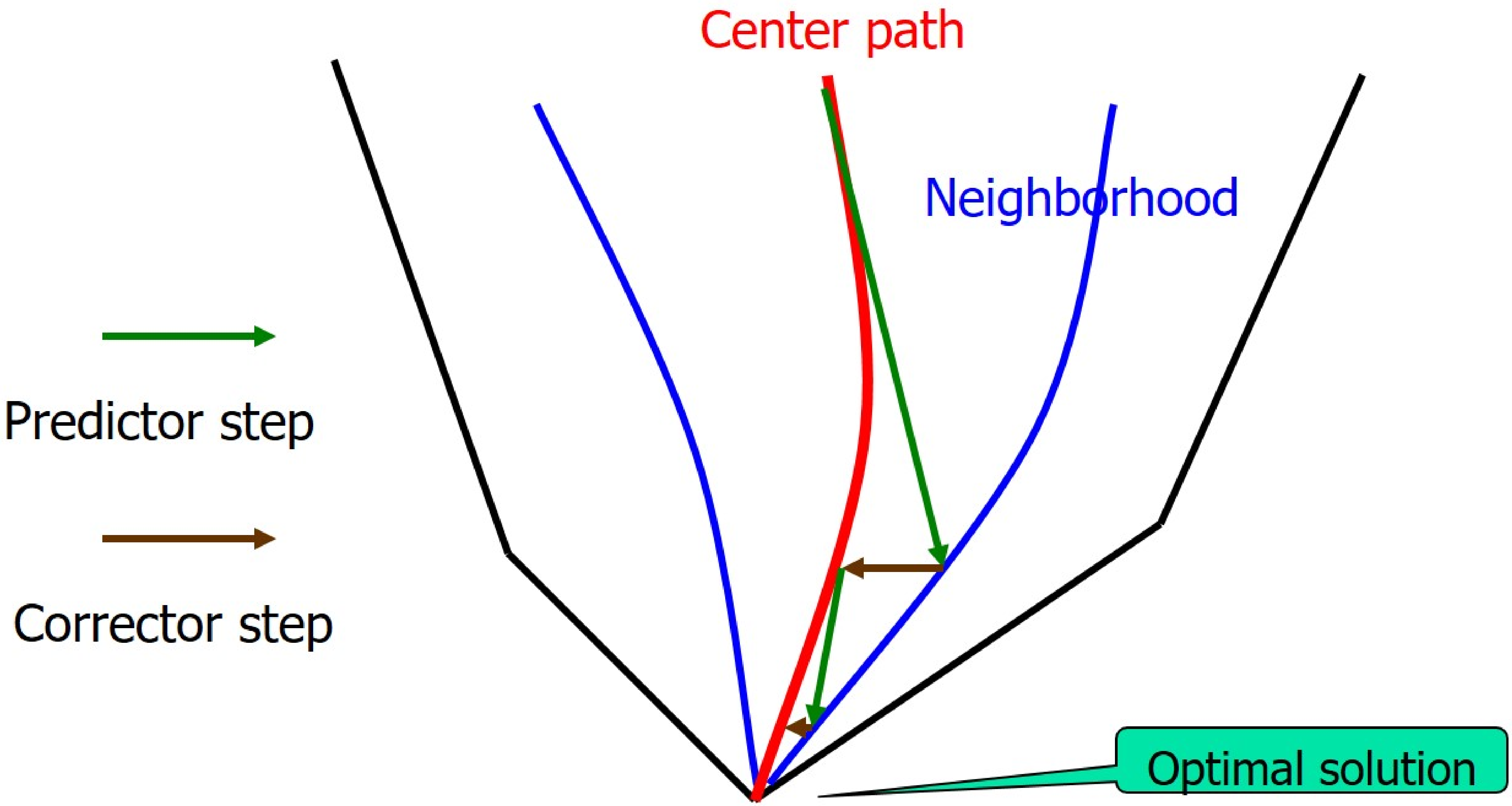

Figure 5: Illustration of the predictor-corrector algorithm.

Reduce the Analytical Volume Directly: Potential Reduction

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

For $\mathbf{x} \in \text{int}\, \mathcal{F}_p$, Karmarkar's primal potential function is defined by

$$\psi_n(\mathbf{x}) := n \log(\mathbf{c}^T \mathbf{x} - z^*) - \sum_{j=1}^{n} \log(x_j),$$

where z^* is the optimal objective value of the LP problem.

This leads to n iteration complexity.

Primal-Dual Potential Function for LP

Typically, a single merit-function driven algorithm is preferred since it can adaptively take large step sizes as long as the merit value is sufficiently reduced, comparing to check and balance of hyper-parameters/measures of the path-following type of algorithms.

For $\mathbf{x} \in \text{int}\, \mathcal{F}_p$ and $(\mathbf{y}, \mathbf{s}) \in \text{int}\, \mathcal{F}_d$, the joint Tanabe-Todd-Ye 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 $\rho \geq 0$ and it is fixed.

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

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

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

Choosing $\rho = \sqrt{n}$ leads to $\sqrt{n}$ iteration complexity.

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.

Allow to take longer step-size α!

Alternating Direction Method

Recall that for $\mathbf{x} \in \operatorname{int} \mathcal{F}_p$ and $(\mathbf{y}, \mathbf{s}) \in \operatorname{int} \mathcal{F}_d$, the joint primal-dual potential function is defined as

$$\begin{aligned} \psi_{n+\rho}(\mathbf{x}, \mathbf{s}) := & \quad (n+\rho) \log(\mathbf{x}^T \mathbf{s}) - \sum_{j=1}^n \log(x_j s_j) \\ = & \quad (n+\rho) \log(\mathbf{c}^T \mathbf{x} - \mathbf{b}^T \mathbf{y}) - \sum_{j=1}^n \log(x_j) - \sum_{j=1}^n \log(s_j). \end{aligned}$$

Alternate Updating of primal $\mathbf{x}$ and $(\mathbf{y}, \mathbf{s})$: at the kth step, fix $(\mathbf{y}^k, \mathbf{s}^k)$ and reduce the potential function by a constant via updating from $\mathbf{x}^k$ to $\mathbf{x}^{k+1}$ while keep $(\mathbf{y}^{k+1}, \mathbf{s}^{k+1}) = (\mathbf{y}^k, \mathbf{s}^k)$:

$$\psi_{n+\rho}(\mathbf{x}^{k+1}, \mathbf{s}^{k+1}) - \psi_{n+\rho}(\mathbf{x}^k, \mathbf{s}^k) \le -\delta.$$

Once can prove that, if by updating primal only one cannot reduce the potential function by a constant anymore, then one must be able to update the dual from $(\mathbf{y}^k, \mathbf{s}^k)$ to $(\mathbf{y}^{k+1}, \mathbf{s}^{k+1})$ (while keep $\mathbf{x}^{k+1} = \mathbf{x}^k$) and reduce the potential function by a constant; see Y 1989. The sample complexity result holds and it was the first one extended to solving SDP by Alizadeh 1992.

First-Order Potential Reduction

At the kth iteration, we compute the direction vectors $(\mathbf{d}_x, \mathbf{d}_y, \mathbf{d}_s)$ using the steepest descent direction:

$$\begin{array}{lll} \min & \nabla_x \phi(\mathbf{x}^k, \mathbf{s}^k)^T \mathbf{d}_x + \nabla_s \phi(\mathbf{x}^k, \mathbf{s}^k)^T \mathbf{d}_s & \\ \text{s.t.} & A\mathbf{d}_x & = \mathbf{0} \\ & A^T \mathbf{d}_y + \mathbf{d}_s & = \mathbf{0}. \end{array}$$

Thus,

$$\begin{aligned} \mathbf{d}_x &= -(I - A^T(AA^T)^{-1}A)\nabla_x \phi(\mathbf{x}^k, \mathbf{s}^k), \\ \mathbf{d}_y &= A\nabla_s \phi(\mathbf{x}^k, \mathbf{s}^k), \\ \mathbf{d}_s &= -A^T A \nabla_s \phi(\mathbf{x}^k, \mathbf{s}^k). \end{aligned}$$

First-Order Potential Reduction as a Presolver

- First-order method solves to 1e-02 accuracy and then switch to second-order
- An average solution reduction of 30%

Accuracy	1e-04	1e-06	1e-08	1e-10
First-order	7.5	798.0	>1200	>1200
Second-order	33.0	56.7	89.3	93.3
First + Second	5.4	12.1	14.1	15.2

Figure 6: Speed-Up on QAP-LP

Potential Reduction for General Linear Complementarity

Given $M \in R^{n\times n}$ and $\mathbf{q} \in R^n$, find $(\mathbf{x}, \mathbf{s})$

$$\mathbf{s} = M\mathbf{x} + \mathbf{q},\ \mathbf{x} \geq \mathbf{0},\ \mathbf{s} \geq \mathbf{0}, \quad \text{and} \quad \mathbf{x}^T\mathbf{s} = 0$$

- LP: M skew symmetric
- Convex QP: M symmetric and monotone: $(\mathbf{x}-\mathbf{y})^T M(\mathbf{x}-\mathbf{y}) \geq 0$
- Monotone LCP: $M + M^T$ is Positive Semidefinite. The PRA terminates in $O(\sqrt{n}\log(\epsilon^{-1}))$ iterations for above three problems.
- P-matrix: $0 < \gamma = \max_j \min \frac{\mathbf{x}^T M^T \mathbf{x}}{\|\mathbf{x}\|^2}$. The PRA terminates in $O(n^2 \max(|\lambda|/(\gamma n), 1)\log(\epsilon^{-1}))$ iterations, where $\lambda($ is the least eigenvalue of $(M+M^T)/2$; see Kojima et al. 1992
- General QP: M symmetric but non-convex: The PRA terminates in $O(n^2\epsilon^{-1}\log(\epsilon^{-1}) + n\log(n))$ iterations (no condition-numbers!) with a solution that is ϵ accurate on both the first and second order

optimality conditions; see Y 1998.

Initialization

- Combining the primal and dual into a single linear feasibility problem, then applying LP algorithms to find a feasible point of the problem. Theoretically, this approach can retain the currently best complexity result.
- The big M method, i.e., add one or more artificial column(s) and/or row(s) and a huge penalty parameter M to force solutions to become feasible during the algorithm.
- Phase I-then-Phase II method, i.e., first try to find a feasible point (and possibly one for the dual problem), and then start to look for an optimal solution if the problem is feasible and bounded.
- Combined Phase I-Phase II method, i.e., approach feasibility and optimality simultaneously. To our knowledge, the "best" complexity of this approach is $O(n \log(R/\epsilon))$.

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

$$
\text{(Solvable)} \quad
\begin{array}{rl}
A\mathbf{x} - \mathbf{b} & = \mathbf{0} \\
-A^T\mathbf{y} + \mathbf{c} & \geq \mathbf{0}, \\
\mathbf{b}^T\mathbf{y} - \mathbf{c}^T\mathbf{x} & = 0, \\
\mathbf{y} \text{ free}, \ \mathbf{x} & \geq \mathbf{0}
\end{array}
\quad \text{or} \quad
\text{(Infeasible)} \quad
\begin{array}{rl}
A\mathbf{x} & = \mathbf{0} \\
-A^T\mathbf{y} & \geq \mathbf{0}, \\
\mathbf{b}^T\mathbf{y} - \mathbf{c}^T\mathbf{x} & > 0, \\
\mathbf{y} \text{ free}, \ \mathbf{x} & \geq \mathbf{0}
\end{array}
$$

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

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

The Homogeneous System is Self-Dual

$$
\begin{array}{llll}
(HP) & A\mathbf{x} - \mathbf{b}\tau & = \mathbf{0}, \ (\mathbf{y}') \\
& -A^T\mathbf{y} + \mathbf{c}\tau & = \mathbf{s} \geq \mathbf{0}, \ (\mathbf{x}') \\
& \mathbf{b}^T\mathbf{y} - \mathbf{c}^T\mathbf{x} & = \kappa \geq 0, (\tau') \\
& \mathbf{y} \text{ free}, \ (\mathbf{x}; \tau) & \geq \mathbf{0}
\end{array}
\qquad
\begin{array}{llll}
(HD) & A\mathbf{x}' - \mathbf{b}\tau' & = \mathbf{0}, \\
& A^T\mathbf{y}' - \mathbf{c}\tau' & \leq \mathbf{0}, \\
& -\mathbf{b}^T\mathbf{y}' + \mathbf{c}^T\mathbf{x}' & \leq 0, \\
& \mathbf{y}' \text{ free}, \ (\mathbf{x}'; \tau') & \geq \mathbf{0}
\end{array}
$$

Theorem 3 *System (HP) is feasible (e.g. all zeros) and any feasible solution* $(\mathbf{y}, \mathbf{x}, \tau, \mathbf{s}, \kappa)$ *is self-complementary:*

$$\mathbf{x}^T\mathbf{s} + \tau\kappa = 0.$$

Furthermore, it has a strictly self-complementary feasible solution

$$\begin{pmatrix} \mathbf{x} + \mathbf{s} \\ \tau + \kappa \end{pmatrix} > \mathbf{0},$$

Start from any infeasible but interior-solution pair in the primal and dual cones, and apply IPM to solve the Phase I problem; see Y-Todd-Mizuno 1994.