Bootcamp: Interior Point Algorithms I

Yinyu Ye

Friday, September 01 2023
Simons Institute
Figure 1: Slide from Daniel Dadush and Bento Natura 2023
Consider linear program:

$$\begin{align*}
\min & \quad c^T x \\
\text{s.t.} & \quad Ax = b, \quad x \geq 0.
\end{align*}$$

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

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

Let $z^*$ denote the optimal value and

$$\mathcal{F} = \mathcal{F}_p \times \mathcal{F}_d.$$ 

We are interested in finding an $\epsilon$-approximate solution for the LP problem:

$$c^T x - b^T y \leq \epsilon.$$ 

For simplicity, we assume that an interior-point pair $(x^0, y^0, s^0)$ is known, and we will use it as our initial point pair.

Karmarkar 1984, Renegar 1986...
Consider the logarithmic barrier function optimization

\[(PB) \quad \text{minimize} \quad -\sum_{j=1}^{n} \log x_j \]
\[\text{s.t.} \quad x \in \text{int } F_p\]

and

\[(DB) \quad \text{maximize} \quad \sum_{j=1}^{n} \log s_j \]
\[\text{s.t.} \quad (y, s) \in \text{int } F_d\]

They are linearly constrained convex programs (LCCP).

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

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

where $e$ is the vector of all ones; or

$$\bar{X}s = e$$

$$A\bar{x} = b$$

$$-A^T y - s = 0$$

$$\bar{x} > 0$$

where $X$ is the diagonal matrix generated from vector $x$.

Sonnevend 1988, Bayer and Lagarias 1989, Megiddo 1989...
The maximizer \((\bar{y}, \bar{s})\) of (DB) is called the **analytic center** of polytope \(\mathcal{F}_d\), and we have

\[
\begin{align*}
\bar{S}x &= e \\
Ax &= 0 \\
-A^T\bar{y} - \bar{s} &= -c \\
\bar{s} &> 0.
\end{align*}
\] (2)

Figure 2: Analytic center maximizes the product of slacks.
Volumetric Center for the Dual Polytope

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

\[
(DB) \quad \text{maximize} \quad -\log \det(AS^{-2}A^T) \\
\text{s.t.} \quad (y, s) \in \text{int} \mathcal{F}_d
\]

More details see Vaidya 1996, Lee-Sidford 13-’19, van den Brand, Lee, Liu, Saranurak, Sidford, Song, Wang 21, etc.
The analytic center of polytope $\mathcal{F}_d$ is an analytic function of input data $A, c$.

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

Consider $\Omega' = \{ y \in \mathbb{R} : \underbrace{-y \leq 0, \cdots, -y \leq 0}_{\text{n times}}, y \leq 1 \}$, 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 $x = ((n + 1)/n, \cdots, (n + 1)/n, (n + 1))^T$. 
Analytic Volume of Polytope and Cutting Plane

\[ AV(\mathcal{F}_d) := \prod_{j=1}^{n} \bar{s}_j = \prod_{j=1}^{n} (c_j - a_j^T \bar{y}) \]

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 \( a_1^T y \leq c_1 \) to \( a_1^T y \leq a_1^T \bar{y} \); i.e., the first inequality is parallelly moved and it now cuts through \( \bar{y} \) and divides \( \mathcal{F} \) into two bodies.

Analytically, \( c_1 \) is replaced by \( a_1^T \bar{y} \) and the rest of data are unchanged. Let

\[ \mathcal{F}^+ := \{ y : a_j^T y \leq c_j^+, \ j = 1, \ldots, n \}, \]

where \( c_j^+ = c_j \) for \( j = 2, \ldots, n \) and \( c_1^+ = a_1^T \bar{y} \).
Figure 3: Translation of a hyperplane through the AC.

The initial polytope and its analytic center

This hyperplane is translated to $\tilde{y}^a$
Let $\bar{y}^+$ be the analytic center of $\mathcal{F}^+$. Then, the analytic volume of $\mathcal{F}^+$

$$AV(\mathcal{F}^+) = \prod_{j=1}^{n} (c_j^+ - a_j^T \bar{y}^+) = (a_1^T \bar{y} - a_1^T \bar{y}^+) \prod_{j=2}^{n} (c_j - a_j^T \bar{y}^+).$$

We have the following volume reduction theorem:

**Theorem 1**

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

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

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

Thus,

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

We have

\[
e^T\bar{X}\bar{s}^+ = e^T\bar{X}(c^+ - A^T\bar{y}^+) = e^T\bar{X}c^+
\]

\[
= c^T\bar{x} - \bar{x}_1(c_1 - a_1^T\bar{y}) = n - 1.
\]
\[
\frac{AV(\mathcal{F}^+)}{AV(\mathcal{F})} = \prod_{j=1}^{n} \frac{s_j^+}{s_j}
\]

\[
= \prod_{j=1}^{n} \bar{x}_j s_j^+
\]

\[
\leq \left( \frac{1}{n} \sum_{j=1}^{n} \bar{x}_j s_j^+ \right)^n
\]

\[
= \left( \frac{1}{n} e^T \bar{X} s^+ \right)^n
\]

\[
= \left( \frac{n-1}{n} \right)^n \leq \exp(-1).
\]
Now suppose we translate $k(< n)$ hyperplanes, say $1, 2, \ldots, k$, moved to cut the analytic center $\bar{y}$ of $\mathcal{F}$, that is,

$$\mathcal{F}^+ := \{y : a_j^T y \leq c_j^+, j = 1, \ldots, n\},$$

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

**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 \( \mathcal{F} := \{ \mathbf{y} : \mathbf{a}_j^T \mathbf{y} \leq c_j, \ j = 1, \ldots, n \} \).

Start with the initial polytope

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

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 \( \mathcal{F} := \{ y : \mathbf{a}_j^T y \leq c_j, \ j = 1, \ldots, n \} \) such that it is Pareto-Maximal for \( k \) objective function \( \mathbf{b}_i^T y, \ i = 1, \ldots, k \)

- Weighted-Sum Objective Maximization: \( \sum_i w_i \mathbf{b}_i^T y \)

- Alternating Cutting-Plane: Cutting the AC alternatively, or simultaneously with fixed proportions?
Now consider the problem

$$\begin{align*}
\text{maximize} & \quad b^T y \\
\text{s.t.} & \quad A^T y \leq c.
\end{align*}$$

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

Start with a polytope

$$\mathcal{F}(R) := \left\{ y : A^T y \leq c, \underbrace{b^T y \geq R, \ldots, b^T y \geq R}_{k \text{ times}} \right\}$$

where $R$ is so low such that $y^0$ 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 $y(R)$: it forms a path of analytic centers from $y^0$ toward the optimal solution set.
An equivalent algebraic representation of the path is to consider the LP problem with the weighted barrier function

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

and also

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

where \(\mu\) is called the barrier (weight) parameter.

They are again linearly constrained convex programs (LCCP).
They share the same first-order KKT conditions:

\[
Xs = \mu e \\
Ax = b \\
-A^Ty - s = -c;
\]

where we have

\[
\mu = \frac{x^Ts}{n} = \frac{c^Tx - b^Ty}{n},
\]

so that it’s the average of complementarity or duality gap.

Denote by \((x(\mu), y(\mu), s(\mu))\) the (unique) solution satisfying the conditions. As \(\mu\) decreases to zero, \(x(\mu)\) form a path in the primal feasible region and \(y(\mu)\) form a path in the dual feasible region to-warding optimality respectively.
Figure 4: The central path of $y(\mu)$ in a dual feasible region.
Central Path for Linear Programming Parametrized by $\mu$

**Theorem 2** Let $(x(\mu), y(\mu), s(\mu))$ be on the central path of an linear program in standard form.

i) The central path point $(x(\mu), s(\mu))$ is bounded for $0 < \mu \leq \mu^0$ and any given $0 < \mu^0 < \infty$.

ii) For $0 < \mu' < \mu$,
\[ c^T x(\mu') < c^T x(\mu) \quad \text{and} \quad b^T y(\mu') > b^T y(\mu) \]

if both primal and dual have nontrivial optimal solutions.

iii) $(x(\mu), s(\mu))$ converges to an optimal solution pair for (LP) and (LD). Moreover, the limit point $x(0)_{P^*} > 0$ and the limit point $s(0)_{Z^*} > 0$, where $(P^*, Z^*)$ are the analytic centers on the primal and dual optimal faces, respectively (G"uler and Y 1993).
Proof of (i)

\[ (x(\mu^0) - x(\mu))^T (s(\mu^0) - s(\mu)) = 0, \]

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

\[ \sum_{j}^{n} (s(\mu^0)_j x(\mu)_j + x(\mu^0)_j s(\mu)_j) = n(\mu^0 + \mu) \leq 2n\mu^0, \]

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, \(x(\mu)\) and \(s(\mu)\) are bounded, which proves (i).
In general, one can start from an (approximate) central path point \( x(\mu^0), (y(\mu^0), s(\mu^0)) \), or \( (x(\mu^0), y(\mu^0), s(\mu^0)) \) where \( \mu^0 \) is sufficiently large.

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

\( \mu \) might be reduced at each stage by a specific factor, giving \( \mu^{k+1} = \gamma \mu^k \) where \( \gamma \) is at most \( 1 - \frac{1}{3\sqrt{n}} \), where \( k \) 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,...
Given a pair \((x^k, y^k, s^k) \in \text{int } \mathcal{F}\) closely to the central path, that is,

\[
\|X^k S^k e - \mu^k e\| \leq \eta \mu^k
\]

for a small positive constant \(\eta\), we compute direction vectors \(d_x, d_y\) and \(d_s\) from the system equations:

\[
\begin{align*}
S^k d_x + X^k d_s &= \gamma \mu^k e - X^k S^k e, \\
A d_x &= 0, \\
-A^T d_y - d_s &= 0.
\end{align*}
\] (3)

where \(\gamma = (1 - \frac{1}{3\sqrt{n}})\). Then we update

\[
x^{k+1} = x^k + d_x > 0, \quad y^{k+1} = y^k + d_y, \quad s^{k+1} = s^k + d_s > 0.
\]

Then one can prove

\[
\|X^{k+1} S^{k+1} e - \mu^{k+1} e\| \leq \eta (1 - \frac{1}{3\sqrt{n}}) \mu^k = \eta \mu^{k+1}.
\]

This leads to \(\sqrt{n}\) iteration complexity.
Here we describe and analyze the Predictor-Corrector interior-point algorithm (Mizuno-Todd-Y 1993, Mehrotra 1993). Consider the neighborhood

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

for some $\eta \in (0, 1)$.

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

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

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

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

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

$$(x^{k+1}, s^{k+1}) = (x' + d'_x, s' + d'_s).$$

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

For \( x \in \text{int} \ \mathcal{F}_p \), Karmarkar’s primal potential function is defined by

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

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

This leads to \( n \) iteration complexity.
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 $x \in \text{int } F_p$ and $(y, s) \in \text{int } F_d$, the joint Tanabe-Todd-Ye primal-dual potential function is defined by

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

where $\rho \geq 0$ and it is fixed.

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

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

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

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

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

While \((x^k)^T s^k \geq \epsilon\) do

1. Set \((x, s) = (x^k, s^k)\) and \(\gamma = n/(n + \rho)\) and compute \((d_x, d_y, d_s)\) from (3).

2. Let \(x^{k+1} = x^k + \bar{\alpha}d_x, y^{k+1} = y^k + \bar{\alpha}d_y,\) and \(s^{k+1} = s^k + \bar{\alpha}d_s\) where

\[
\bar{\alpha} = \arg \min_{\alpha \geq 0} \psi_{n+\rho}(x^k + \alpha d_x, s^k + \alpha d_s).
\]

3. Let \(k := k + 1\) and return to Step 1.

Allow to take longer step-size \(\alpha\)!
Recall that 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 as

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

**Alternate Updating** of primal $\mathbf{x}$ and $(\mathbf{y}, \mathbf{s})$: at the $k$th 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 $(\mathbf{y}^{k+1}, \mathbf{s}^{k+1}) = (\mathbf{y}^k, \mathbf{s}^k)$:

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

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 $k$th iteration, we compute the direction vectors $(d_x, d_y, d_s)$ using the steepest descent direction:

$$
\begin{align*}
\min & \quad \nabla_x \phi(x^k, s^k)^T d_x + \nabla_s \phi(x^k, s^k)^T d_s \\
\text{s.t.} & \\
& A d_x = 0 \\
& A^T d_y + d_s = 0.
\end{align*}
$$

Thus,

$$
\begin{align*}
 d_x &= -(I - A^T(AA^T)^{-1}A) \nabla_x \phi(x^k, s^k), \\
 d_y &= A \nabla_s \phi(x^k, s^k), \\
 d_s &= -A^T A \nabla_s \phi(x^k, s^k).
\end{align*}
$$
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%

<table>
<thead>
<tr>
<th>Accuracy</th>
<th>1e-04</th>
<th>1e-06</th>
<th>1e-08</th>
<th>1e-10</th>
</tr>
</thead>
<tbody>
<tr>
<td>First-order</td>
<td>7.5</td>
<td>798.0</td>
<td>&gt;1200</td>
<td>&gt;1200</td>
</tr>
<tr>
<td>Second-order</td>
<td>33.0</td>
<td>56.7</td>
<td>89.3</td>
<td>93.3</td>
</tr>
<tr>
<td>First + Second</td>
<td>5.4</td>
<td>12.1</td>
<td>14.1</td>
<td>15.2</td>
</tr>
</tbody>
</table>

Figure 6: Speed-Up on QAP-LP
Potential Reduction for General Linear Complementarity

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

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

- **LP:** $M$ skew symmetric
- **Convex QP:** $M$ symmetric and monotone: $(x - y)^T M (x - y) \geq 0$
- **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:** $0 < \gamma = \max_j \min \frac{x^T M^T x}{\|x\|^2}$. The PRA terminates in $O(n^2 \max(\|x\|/(\gamma n), 1) \log(\epsilon^{-1}))$ iterations, where $\lambda(\cdot)$ 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 $\epsilon$ accurate on both the first and second order
optimality conditions; see Y 1998.
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 $O(n \log(\frac{R}{\varepsilon}))$. 
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.
A pair of LP has two alternatives

(Solvable) \[ Ax - b = 0 \]
\[ -A^T y + c \geq 0, \]
\[ b^T y - c^T x = 0, \]
\[ y \text{ free, } x \geq 0 \]

(Infeasible) \[ Ax = 0 \]
\[ -A^T y \geq 0, \]
\[ b^T y - c^T x > 0, \]
\[ y \text{ free, } x \geq 0 \]
An Integrated Homogeneous System

The two alternative systems can be homogenized as one:

\[(HP) \quad Ax - b\tau = 0\]
\[-A^T y + c\tau = s \geq 0,\]
\[b^T y - c^T x = \kappa \geq 0,\]
\[y \text{ free, } (x; \tau) \geq 0\]

where the two alternatives are

(Solvable) : \((\tau > 0, \kappa = 0)\) or (Infeasible) : \((\tau = 0, \kappa > 0)\)
The Homogeneous System is Self-Dual

\[
(HP) \quad Ax - b\tau = 0, \quad (y')
\]
\[
(HD) \quad Ax' - b\tau' = 0,
\]
\[
-A^T y + c\tau = s \geq 0, \quad (x')
\]
\[
A^T y' - c\tau' \leq 0,
\]
\[
b^T y - c^T x = \kappa \geq 0, \quad (\tau')
\]
\[
-b^T y' + c^T x' \leq 0,
\]
\[
y \text{ free, } (x; \tau) \geq 0
\]
\[
y' \text{ free, } (x'; \tau') \geq 0
\]

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

\[
x^T s + \tau \kappa = 0.
\]

Furthermore, it has a strictly self-complementary feasible solution

\[
\begin{pmatrix}
  x + s \\
  \tau + \kappa 
\end{pmatrix} > 0,
\]

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.