Interior Point Methods
Chapter 8

University of Chicago
Booth School of Business
Kipp Martin

November 29, 2016
Files

primalInterior.m
Outline

Motivation

Optimality Conditions

Barrier Function Approach

Primal Path Following

Other Issues
Motivation

Most algorithms work by moving from point to point until a set of optimality conditions are satisfied.

**Generic Algorithm:**

**Initialization:** Start with a point that satisfies a subset of the optimality conditions.

**Iterative Step:** Move to a better point.

**Termination:** Stop when you have satisfied (to numerical tolerances) all the optimality conditions.
Start with a linear program (LP) in standard form:

\[
\begin{align*}
\min & \quad c^\top x \\
Ax & = b \\
x & \geq 0
\end{align*}
\]

Assume we have a feasible \( \bar{x} \), i.e. \( A\bar{x} = b \) and \( \bar{x} \geq 0 \).

A necessary condition for \( \bar{x} \) to be optimal is that we cannot find a direction to move that improves the objective function value.
Optimality Conditions

How do we formally characterize: *we cannot find a direction to move from \( \bar{x} \) that improves the objective function value?*

We cannot find a \( \Delta x \) such that

1. \( c^\top \Delta x < 0 \)
2. \( A\Delta x = 0 \)
3. \( \bar{x} + \Delta x \geq 0 \)

That is, **there cannot be a solution to:**

\[
- c^\top \Delta x > 0 \\
A\Delta x = 0 \\
\Delta x \geq -\bar{x}
\]
Optimality Conditions

How do we characterize no solution (I will unmercifully torture any student that cannot do this) to the following system?

\[-c^\top \Delta x > 0 \tag{1}\]
\[A \Delta x = 0 \tag{2}\]
\[\Delta x \geq -\bar{x} \tag{3}\]

**Characterization 1:** \((\bar{u}_0 > 0, \bar{w} \geq 0)\)

\[-\bar{u}_0 c^\top + \bar{u}^\top A + \bar{w} = 0 \tag{4}\]
\[\bar{w}^\top \bar{x} \geq 0 \tag{5}\]

**Characterization 2:** \((\bar{u}_0 = 0, \bar{w} \geq 0)\)

\[\bar{u}^\top A + \bar{w} = 0 \tag{6}\]
\[\bar{w}^\top \bar{x} > 0 \tag{7}\]
Optimality Conditions

Characterization 1:

\[-\bar{u}_0 c^\top + \bar{u}^\top A + \bar{w} = 0 \quad (8)\]
\[-\bar{w}^\top \bar{x} \geq 0 \quad (9)\]

Observations:

1. We can assume \( \bar{u}_0 = 1 \) (why?) and condition (8) becomes

\[-c^\top + \bar{u}^\top A + \bar{w} = 0 \quad (10)\]

2. Condition (9) is equivalent to

\[\bar{w}^\top \bar{x} = 0 \quad (11)\]

Why?
Optimality Conditions

Characterization 1:

\[ \overline{u}^\top A \leq c^\top \quad \text{(from (10) } \overline{w} \geq 0) \quad (12) \]

\[ (c^\top - \overline{u}^\top A)\overline{x} = 0 \quad \text{(substitute } \overline{w} = c^\top - \overline{u}^\top A \text{ into (11))} \quad (13) \]

\[ A\overline{x} = b, \quad \overline{x} \geq 0 \quad \text{primal feasibility} \quad (14) \]

Characterization 2 cannot happen! Why?
Optimality Conditions

Now extend the problem to have a nonlinear, but differentiable, objective function.

\[
\begin{align*}
\min & \quad f(x) \\
Ax & = b \\
x & \geq 0
\end{align*}
\]

What does it mean for \( \bar{x} \) to be a local optimum?

It means the exactly the same thing as in the linear case.

*We cannot find a direction to move from \( \bar{x} \) that improves the objective function value?*
Optimality Conditions

How do we formally characterize: we cannot find a direction to move from $\bar{x}$ that improves the objective function value?

We cannot find a $\Delta x$ such that

1. $(\nabla f(\bar{x}))^\top \Delta x < 0$
2. $A\Delta x = 0$
3. $\bar{x} + \Delta x \geq 0$

That is, there cannot be a solution to:

$$\begin{align*}
-(\nabla f(\bar{x}))^\top \Delta x &> 0 \\
A\Delta x &= 0 \\
\Delta x &\geq -\bar{x}
\end{align*}$$
Optimality Conditions

Then, again using, a Farkas’ variant:

\[-(\nabla f(x))^T + \bar{u}^T A + \bar{w} = 0\]  \hspace{1cm} (15)

\[-\bar{w}^T x \geq 0\]  \hspace{1cm} (16)

These are called the Karush-Kuhn-Tucker conditions.

There is an interesting bit of University of Chicago history here.
Barrier Function Approach

\[
\begin{align*}
&\text{min } c^\top x \\
&\text{s.t. } Ax = b \\
&\quad x \geq 0
\end{align*}
\]

\[(LP)\]

\[
\begin{align*}
&\text{min } c^\top x - \mu \sum_{j=1}^{n} \ln(x_j) \\
&\text{s.t. } Ax = b \\
&\quad x > 0
\end{align*}
\]

\[(LP_\mu)\]
Barrier Function Approach

That is, if \( x \) is an optimal solution to \((LP_\mu)\) there cannot be a solution to the system

\[
(c - \mu X^{-1} e)^T \Delta x < 0 \\
A \Delta x = 0
\]  \hspace{1cm} (17) \hspace{1cm} (18)

What does it mean for \((17)-(18)\) to be infeasible? It means I can eliminate the \( \Delta x \) variables and get \( 0 < 0 \).

Projecting out \( \Delta x \) means there is a solution to:

\[
(c - \mu X^{-1} e)u_0 - A^T u = 0, \quad u_0 > 0
\]  \hspace{1cm} (19)
Barrier Function Approach

If a solution $x$ is an optimal solution to $(LP_\mu)$ there is a corresponding dual solution $u$ such that the primal-dual pair $(x, u)$ satisfies the system ($u_0 = 1$)

$$c - \mu X^{-1}e - A^\top u = 0 \quad (20)$$
$$Ax = b \quad (21)$$
$$x > 0. \quad (22)$$

Let $w = \mu X^{-1}e$ and the optimality conditions are

$$A^\top u + w = c \quad (23)$$
$$Ax = b \quad (24)$$
$$x > 0 \quad (25)$$
$$w = \mu X^{-1}e. \quad (26)$$
Lemma 1: If problem \((LP)\) has a feasible solution then the set of optimal solutions to \((LP)\) is bounded and not empty if and only if there is a solution \((\overline{u}, \overline{w})\) with \(\overline{w} > 0\) to the dual problem \((DLP)\).

Lemma 2: If there exists a feasible solution \(\overline{x} > 0\) to the primal problem \((LP)\), then for any \(\mu > 0\), problem \((LP_\mu)\) has an optimal solution if and only if the set of optimal solutions to \((LP)\) is bounded and not empty.
Barrier Function Approach

Barrier Assumptions

1. The set \( \{ x \in \mathbb{R}^n | Ax = b, \ x > 0 \} \) is not empty.

2. The set \( \{(u, w) \in \mathbb{R}^m \times \mathbb{R}^n | A^T u + w = c, \ w > 0 \} \) is not empty.

3. The constraint matrix \( A \) has rank \( m \).

By Assumptions 1 and 2, and Lemma 1, it follows that set of optimal solutions to \((LP)\) is bounded and not empty.

Then by Lemma 2, for any \( \mu > 0 \), problem \((LP_{\mu})\) has an optimal solution.
Barrier Function Approach

**Proposition 1:** Given Barrier Assumptions 1-3, and $\mu > 0$, there is a unique solution $(x(\mu), u(\mu), w(\mu))$ to the Karush-Kuhn-Tucker conditions (23)-(26) and $x(\mu)$ is the optimal solution to problem $(LP_\mu)$.

**Proposition 2:** Given Barrier Assumptions 1-3, $(x(\mu), u(\mu), w(\mu))$ converges to an optimal primal-dual solution as $\mu \to 0$. 
Primal Path Following

The Karush-Kuhn-Tucker conditions for \((LP_\mu)\) are

\[
A^\top u + w = c
\]

\[
Ax - b = 0
\]

\[
w - \mu X^{-1}e = 0
\]

\[
x > 0.
\]

We have nonlinear term. What to do?
Primal Path Following

Replace the nonlinear term

\[
f(x_j, w_j) = w_j - \frac{\mu_k}{x_j} = 0
\]

with the first order approximation about the point \((x_j^k, w_j^k)\) which is

\[
f(x_j, w_j) \approx f(x_j^k, w_j^k) + \nabla f(x_j^k, w_j^k) \begin{bmatrix} x_j - x_j^k \\ w_j - w_j^k \end{bmatrix}
\]

\[
= (w_j^k - \frac{\mu_k}{x_j^k}) + [\frac{\mu_k}{(x_j^k)^2}, 1] \begin{bmatrix} x_j - x_j^k \\ w_j - w_j^k \end{bmatrix}.
\]
Primal Path Following

Let $\Delta x = x - x^k$, $\Delta u = u - u^k$ and $\Delta w = w - w^k$. We want to solve the system:

$$A^\top \Delta u + I \Delta w = 0$$

$$A \Delta x = 0$$

$$w^k - \mu_k X_k^{-1} e = -\mu_k X_k^{-2} \Delta x - \Delta w.$$  

The variables are $\Delta u$, $\Delta x$, and $\Delta w$. We solve the system for these variables. This is a linear system!
Primal Path Following

If we can solve the system

\[
A^\top \Delta u + l \Delta w = 0
\]
\[
A \Delta x = 0
\]
\[
-\mu_k X_k^{-2} \Delta x - \Delta w = w^k - \mu_k X_k^{-1} e
\]

then given iterate \((x^k, u^k, w^k)\), and solution \((\Delta x, \Delta u \text{ and } \Delta w)\), for sufficiently small \(\alpha_k\), we calculate

\[
(x^{k+1}, u^{k+1}, w^{k+1}) = (x^k, u^k, w^k) + \alpha_k (\Delta x, \Delta u, \Delta w)
\]
Primal Path Following

Take

$$w^k - \mu_k X_k^{-1} e = -\mu_k X_k^{-2} \Delta x - \Delta w.$$  

and rewrite as

$$X_k W_k e - \mu_k e = -\mu_k X_k^{-1} \Delta x - X_k \Delta w$$

Now multiply by $AX_k$

$$AX_k (X_k W_k e - \mu_k e) = AX_k (-\mu_k X_k^{-1} \Delta x - X_k \Delta w)$$
Primal Path Following

\[ AX_k(X_k W_k e - \mu_k e) = AX_k(-\mu_k X_k^{-1} \Delta x - X_k \Delta w) \]

\[ AX_k(X_k W_k e - \mu_k e) = (-\mu_k AX_k X_k^{-1} \Delta x - AX_k X_k \Delta w) \]
\[ = (-\mu_k A \Delta x - AX_k^2 \Delta w) \]

But

\[ A^\top \Delta u + I \Delta w = 0 \]
\[ A \Delta x = 0 \]

and we have

\[ AX_k(X_k W_k e - \mu_k e) = AX_k^2 A^\top \Delta u \]
Primal Path Following

\[ AX_k(X_k W_k e - \mu_k e) = AX_k^2 A^\top \Delta u \]

Solve for \( \Delta u \)

\[ \Delta u = (AX_k^2 A^\top)^{-1}(AX_k)(X_k W_k e - \mu_k e) \]

We had

\[ X_k W_k e - \mu_k e = -\mu_k X_k^{-1}\Delta x - X_k \Delta w \]

Using \( A^\top \Delta u + I \Delta w = 0 \) gives

\[ X_k W_k e - \mu_k e = -\mu_k X_k^{-1}\Delta x + X_k A^\top \Delta u \]
Primal Path Following

\[ \Delta x = -\frac{1}{\mu_k} X_k (X_k W_k e - \mu_k e) + \frac{1}{\mu_k} X_k (X_k A^\top) \Delta u \]

\[ = -\frac{1}{\mu_k} X_k \left( I - (X_k A^\top)(AX_k^2 A^\top)^{-1}(AX_k) \right) (X_k W_k e - \mu_k e) \]

\[ = -\frac{1}{\mu_k} X_k \left( I - (X_k A^\top)(AX_k^2 A^\top)^{-1}(AX_k) \right) (X_k c - \mu_k e) \]

Finally using \( A^\top \Delta u + I \Delta w = 0 \) we have

\[ \Delta w = -A^\top \Delta u. \]  

(27)
Primal Path Following

We calculate an $\alpha_k$ in both primal and dual space. The $\alpha_k$ is selected so that $w^{k+1}, x^{k+1} > 0$. It is therefore possible to select both a “primal” $\alpha^k_P$ so that $x^{k+1} = x^k + \alpha^k_P \Delta x$ remains positive and a dual $\alpha^k_D$ so that $w^{k+1} = w^k + \alpha^k_D \Delta w$ remains positive. This is done by performing the two ratio tests

$$
\alpha^k_P = \alpha \left( \min_i \{x_i^k / (-\Delta x_i) | \Delta x_i < 0 \} \right),
$$

$$
\alpha^k_D = \alpha \left( \min_i \{w_i^k / (-\Delta w_i) | \Delta w_i < 0 \} \right)
$$

where $\alpha \in (0, 1)$.
Primal Path Following

Step 1: (Initialization) $k \leftarrow 0$, $x^0$, $w^0 > 0$ and $u^0$ such that $Ax^0 = b$, $A^\top u^0 + w^0 = c$, $\alpha, \theta \in (0, 1)$ and $\epsilon, \mu_0 > 0$.

Step 2: Find the directions in which to move:

$$\Delta u \leftarrow (AX_k^2 A^\top)^{-1}(AX_k)(X_k W_k e - \mu_k e)$$

$$\Delta x \leftarrow -\frac{1}{\mu_k} X_k \left( I - (X_k A^\top)(AX_k^2 A^\top)^{-1}(AX_k) \right)(X_k W_k e - \mu_k e)$$

$$\Delta w \leftarrow -A^\top \Delta w$$
Primal Path Following

Step 3: (Calculate New Solution)

\[ x^{k+1} \leftarrow x^k + \alpha^k_P \Delta x \]
\[ u^{k+1} \leftarrow u^k + \alpha^k_D \Delta u \]
\[ w^{k+1} \leftarrow w^k + \alpha^k_D \Delta w \]

where \( \alpha^k_P \) and \( \alpha^k_D \) are calculated as shown earlier.

Step 4: (Termination Test) If \( c^\top x^k - b^\top u^k \geq \epsilon \), update \( \mu_k \leftarrow (1 - \theta)\mu_k \), \( k \leftarrow k + 1 \) and return to Step 2; otherwise, stop.
Primal Path Following

Optimal Solution
Primal Path Following

Where is the work?

What is it in Simplex?
Primal Path Following

Answer:

\[(AX^2_iA^\top)^{-1}\]

What kind of matrix are we inverting?

1. Symmetric

2. Positive definite

We find the **Cholesky decomposition**.
Dual Path Following

\[
\begin{align*}
\max \quad & b^\top u + \mu \sum_{j=1}^{n} \ln(w_j) \\
\text{s.t.} \quad & A^\top u + w = c \\
& w > 0.
\end{align*}
\]

\[
\begin{align*}
A^\top u + w &= c \\
Ax &= b \\
\mu W^{-1} e - x &= 0 \\
w &> 0.
\end{align*}
\]
Primal-Dual Path Following

\[ W X e = \mu e \]

\[ A x = b \]

\[ A^T u + w = c. \]

What is the interpretation?