6.854J / 18.415J Advanced Algorithms
Fall 2008

Lecture 7

Lecturer: Michel X. Goemans *Scribe: Nick Hanssens and Nick Matsakis*

1 Properties of Barrier Problem

Last lecture, we used the strict convexity of the logarithmic barrier function to show that $\mathbf{BP}(\mu)$ has at most one solution. Now we will prove that a solution exists. We assume throughout and without loss of generality that the rank of $A = m$ where A is $m \times n$.

Theorem 1 *Assume that* $\exists\, \hat{x} > 0 : Ax = b$ *(i.e.* $\mathbf{BP}(\mu)$ *is feasible), and* $\exists\, \hat{s} > 0, \exists\, y : A^T\hat{y} + \hat{s} = c$. *Then* $\mathbf{BP}(\mu)$ *is finite and has a unique solution.*

Proof of Theorem 1:
Take any $x > 0$ such that $Ax = b$. We have:

$$\begin{aligned}
c^T x - \mu \sum_j \ln x_j &= (\hat{s}^T + \hat{y}^T A)x - \mu \sum_j \ln x_j \\
&= \hat{s}^T x + \hat{y}^T Ax - \mu \sum_j \ln x_j \\
&= \hat{y}^T b + \sum_j (\hat{s}_j \cdot x_j - \mu \ln x_j), \quad \text{(this sum cannot be arbitrarily negative)} \\
&\geq \hat{y}^T b + \sum_j \min_x (\hat{s}_j \cdot x + \mu \ln x),
\end{aligned}$$

implying that the objective function is lower bounded by a constant (this follows from the fact that for $\mu > 0$, $\hat{s}_j \cdot x + \mu \ln x$ tends to $+\infty$ as x goes to 0 or to $+\infty$). Therefore the infimum of $\mathbf{BP}(\mu)$ is finite for every $\mu > 0$. To show that the infimum is attained (that there exists an optimum solution), it is sufficient to notice that the argument above also leads to upper and lower bounds on x_j in order to have a value below the one for $\hat{x}$, which means that we can restrict our attention to a compact set; this implies that the infimum is attained. Finally, we have shown last time that if an optimum solution exists then it is unique. □

For any $\mu > 0$, the unique solution to $\mathbf{BP}(\mu)$ is called the μ-center.

2 Karush, Kuhn, and Tucker (KKT) Conditions

Remember the optimality conditions from last lecture. The solution x is optimum for $\mathbf{BP}(\mu)$ if $\exists\, y$ such that

$$\begin{aligned}
Ax &= b \\
x &> 0 \\
c - \mu X^{-1} e &= A^T y,
\end{aligned}$$

where

$$X = \begin{pmatrix} x_1 & 0 & \cdots & 0 \\ 0 & x_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & x_n \end{pmatrix}.$$

By setting s to be $\mu X^{-1}e$, these conditions can be re-written as $\exists\ y, s$ such that

$$\begin{aligned} Ax &= b, \quad x > 0, && (1) \\ A^T y + s &= x, \quad s > 0, && (2) \\ x_j \cdot s_j &= \mu, \quad \forall\ j. && (3) \end{aligned}$$

2.1 Definition of Algorithm

To find the μ-center, we need to solve (1)–(3). However, the constraints (3) are quadratic, making this hard to solve[1]. Instead, in order to find (or approximate) the μ-center, we use an iterative method based on Netwon's method. We assume we have a solution that satisfies (1) and (2), but not necessarily (3). We will then linearize equations (3) around our values of x and s, and solve the corresponding linear system. This gives us new values for x and s and we proceed. We will show that if we start "close enough" from the μ-center then after this update step we will be even closer, and this iterative process will converge to the μ-center of $\mathbf{BP}(\mu)$.

2.2 Update Derivation

Replacing x, y and s with

$$\begin{aligned} x &\leftarrow x + \Delta x \\ y &\leftarrow y + \Delta y \\ s &\leftarrow s + \Delta s \end{aligned}$$

and ignoring $\Delta x \cdot \Delta s$ in (3), we arrive at

$$\begin{aligned} A\Delta x &= 0, && (4) \\ A^T \Delta y + \Delta s &= 0, && (5) \\ x_j \cdot s_j + \Delta x_j \cdot s_j + x_j \cdot \Delta s_j &= \mu. && (6) \end{aligned}$$

We claim this system has the unique solution,

$$\begin{aligned} \Delta y &= (AXS^{-1}A^T)^{-1}(b - \mu AS^{-1}e)) && (7) \\ \Delta s &= -A^T \Delta y && (8) \\ \Delta x &= \mu S^{-1}e - x - XS^{-1}\Delta s, && (9) \end{aligned}$$

where

$$S = \begin{pmatrix} s_1 & 0 & \cdots & 0 \\ 0 & s_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & s_n \end{pmatrix}.$$

[1] if such a system was easy to solve then this would give a simple algorithm for linear programming by setting μ to 0 and replacing the strict inequalities by inequalities.

Indeed, (6) implies that $\Delta x_j + x_j s_j^{-1} \Delta s_j = \mu s_j^{-1} - x_j$, or in vector notation,

$$\Delta x + XS^{-1}\Delta s = \mu S^{-1}e - x. \tag{10}$$

Premultiplying by A, using (4) and the fact that $Ax = b$, we get

$$AXS^{-1}\Delta s = \mu AS^{-1}e.$$

Observe that this is not a square system of equations (but we have m equations in n unknowns). Substituting Δs by $-A^T\Delta y$ (because of (5)), we get

$$-(AXS^{-1}A^T)\Delta y = \mu AS^{-1}e - b.$$

But $AXS^{-1}A^T$ is an $m \times m$ matrix of rank m since A has rank m and X and S^{-1} are diagonal matrices with positive diagonal elements. Thus $AXS^{-1}A^T$ is invertible and we derive (7). (5) then immediately implies (8), and (10) implies (9).

At each step, then, replace x and s with the values $x + \Delta x$ and $s + \Delta s$ (y can always be derived from x and s). We will show that this iteration will converge to the μ-center of $\mathbf{BP}(\mu)$.

3 Definitions and Properties

3.1 Proximity Measure

Let $\sigma(x, s, \mu) = ||v||$ be the proximity measure where $v_j = \frac{x_j \cdot s_j}{\mu} - 1$. Note that this will be zero at the μ-center. We will show that $||v||$ decreases with each iteration.

3.2 *ds* and *dx*

As $(x, s, \mu) \to (x + \Delta x, s + \Delta s, \mu)$, our proximity vector v becomes w where:

$$\begin{aligned} v_j = \frac{x_j \cdot s_j}{\mu} - 1 \quad \to \quad w_j &= \frac{(x_j + \Delta x_j)(s_j + \Delta s_j)}{\mu} - 1 \\ &= \frac{\mu + \Delta x_j \cdot \Delta s_j}{\mu} - 1 \\ &= \frac{\Delta x_j \cdot \Delta s_j}{\mu} \quad \text{which we are hoping will be small.} \end{aligned}$$

For the analysis, it will be useful to rescale the x-space and the s-space so that the the current iterates x and s are equal, but in a way that $x_j s_j$ remains constant. For this, we will rescale component j of any vector in the x-space by $\sqrt{\frac{s_j}{x_j}}$ and component j of any vector in the s space by $\sqrt{\frac{x_j}{s_j}}$. Rescaling Δx and Δs, we express w_j as $w_j = dx_j \cdot ds_j$ where $dx_j = \left(\frac{\Delta x_j}{\sqrt{\mu}} \cdot \sqrt{\frac{s_j}{x_j}}\right)$ and $ds_j = \left(\frac{\Delta s_j}{\sqrt{\mu}} \cdot \sqrt{\frac{x_j}{s_j}}\right)$.

3.3 Properties

Property 1 $\Delta x \perp \Delta s$.

Proof of Property 1: This is true because Δx is in the nullspace of A while Δs is in the columnspace of A^T. Indeed, premultiplying (5) by Δx^T and using (4), we get that $\Delta x^T \Delta s = 0$. □

Observe that although $x + \Delta x$ and $s + \Delta s$ do not necessarily satisfy (and will not) that $(x_j + \Delta x_j)(s_j + \Delta s_j) = \mu$, on average they do since the duality gap $(x + \Delta x)^T(s + \Delta s) = \sum_j (x_j s_j + \Delta x_j s_j + x_j \Delta s_j + \Delta x_j \Delta s_j) = n\mu$ by the above property and (6).

Property 2 $dx \perp ds$.

Proof of Property 2: $dx^T ds = \sum_j dx_j \cdot ds_j = \frac{1}{\mu}\sum_j \Delta x_j \cdot \Delta s_j = 0$, using the definition of dx_j, ds_j and Property 1. □

Property 3

$$dx_j + dx_j = \sqrt{\frac{\mu}{x_j \cdot s_j}} - \sqrt{\frac{x_j \cdot s_j}{\mu}} = -v_j \sqrt{\frac{\mu}{x_j \cdot s_j}}.$$

Proof of Property 3:

$$\begin{aligned} dx_j + ds_j &= \frac{\Delta x_j \cdot s_j + \Delta s_j \cdot x_j}{\sqrt{\mu x_j \cdot s_j}} \\ &= \frac{\mu - x_j \cdot s_j}{\sqrt{\mu x_j \cdot s_j}} \\ &= \sqrt{\frac{\mu}{x_j \cdot s_j}} - \sqrt{\frac{x_j \cdot s_j}{\mu}} \\ &= \left(1 - \frac{x_j \cdot s_j}{\mu}\right) \cdot \sqrt{\frac{\mu}{x_j \cdot s_j}} \\ &= -v_j \sqrt{\frac{\mu}{x_j \cdot s_j}} \end{aligned}$$

□

4 Theorems 2 and 3

Theorem 2 *If $\sigma(x, s, \mu) < 1$, then $x + \Delta x > 0$ and $s + \Delta s > 0$.*

Theorem 3 *If $\sigma(x, s, \mu) < \sigma < 1$, then $\sigma(x + \Delta x, s + \Delta s, \mu) \leq \frac{\sigma^2}{2 \cdot (1-\sigma)}$.*

These two theorems guarantee that, provided $\sigma(x, s, \mu)$ is sufficiently small, repeated updates $x + \Delta x, s + \Delta s$ given in the algorithm will converge to the μ-center. Theorem 2 was not proved in this lecture. The proof for theorem 3 is provided below. Theorem 3 shows that the convergence is quadratic (provided we are close enough).

Proof of Theorem 3: We have that $\sigma^2(x + \Delta x, s + \Delta s, \mu) = \|w\|^2 = \sum_j w_j^2 = \sum_j dx_j^2 \cdot ds_j^2$. Using the fact that $4 \cdot a \cdot b \leq (a + b)^2$ and Property 2,

$$\begin{aligned} \sigma^2(x + \Delta x, s + \Delta s, \mu) &\leq \frac{1}{4} \sum_j \left(dx_j^2 + ds_j^2\right)^2 \\ &= \frac{1}{4} \left[\sum_j \left(dx_j^2 + ds_j^2\right) + 2 dx_j \cdot ds_j \right]^2 \\ &\leq \frac{1}{4} \left[\sum_j (dx_j + ds_j)^2 \right]^2. \end{aligned}$$

Using property 3,

$$\sigma^2(x + \Delta x, s + \Delta s, \mu) \leq \frac{1}{4} \left[\sum_j v_j^2 \cdot \frac{\mu}{x_j \cdot s_j} \right]^2.$$

Taking the square root, we get

$$\begin{aligned}\sigma(x+\Delta x, s+\Delta s, \mu) &\leq \frac{1}{2}\sum_j v_j^2 \cdot \frac{\mu}{x_j \cdot s_j} \\ &\leq \frac{1}{2}\left(\sum_j v_j^2\right) \cdot \left(\max_j \frac{\mu}{x_j \cdot s_j}\right) \qquad (11)\end{aligned}$$

Now, consider these two terms. The first, $\sum_j v_j^2$, is equal to σ^2 by definition. The second, $\max_j \mu/(x_j \cdot s_j)$, is at most $1/(1-\sigma)$ by the following argument:

$$\begin{aligned}\|v\| < \sigma &\Rightarrow |v_j| < \sigma, \quad \forall j \\ &\Rightarrow \left|\frac{x_j \cdot s_j}{\mu} - 1\right| < \sigma \\ &\Rightarrow 1 - \sigma < \frac{x_j \cdot s_j}{\mu} < 1 + \sigma \\ &\Rightarrow \frac{\mu}{x_j \cdot s_j} < \frac{1}{1-\sigma}, \quad \forall j\end{aligned}$$

since $1-\sigma$ is strictly positive under the conditions of the theorem. Using these upper bounds in (11), we can conclude the proof by stating,

$$\sigma(x+\Delta x, s+\Delta s, \mu) \leq \frac{\sigma^2}{2 \cdot (1-\sigma)}.$$

□

Corollary 4 *If $\sigma < \frac{2}{3}$, then $\frac{\sigma^2}{2\cdot(1-\sigma)} < \sigma < \frac{2}{3}$.*

This corollary gives us a necessary initial bound on σ to guarantee convergence.

5 Theorem 5

Theorems 2 and 3 show that the updates Δx, Δs given in the algorithm will converge to the μ-center. However, instead of making the Newton updates to converge to the μ-center for a fixed value of μ, we take one step to get closer to the μ-center (σ becomes now $\sigma^2/(2(1-\sigma))$) and then decrease μ (since our goal is let μ tend to 0) in such a way that our new iterate is within the original value of σ but with respect to the updated μ. Theorem 5 shows how the proximity measure changes as we change μ.

Theorem 5 *Suppose $x^T s = n\mu$ and $\sigma = \sigma(x,s,\mu)$, then $\sigma(x,s,\mu(1-\theta)) = \frac{1}{1-\theta}\sqrt{\sigma^2 + \theta^2 \cdot n}$.*

Observe that our new iterate $x+\Delta x$ and $s+\Delta s$ satisfy the condition that $(x+\Delta x)^T(s+\Delta s) = n\mu$, and hence Theorem 5 can be applied to see how the proximity measure changes when we modify μ after a Newton iterate.
Proof of Theorem 5: Let v' be the vector v after having changed μ, i.e. $v'_j = \frac{x_j \cdot s_j}{\mu\cdot(1-\theta)} - 1$. We have that

$$v'_j = \frac{x_j \cdot s_j}{\mu \cdot (1-\theta)} - 1$$

$$
\begin{aligned}
&= \frac{1}{1-\theta} \cdot \left(\frac{x_j \cdot s_j}{\mu} - 1 \right) + \frac{\theta}{1-\theta} \\
&= \frac{1}{1-\theta} \cdot v_j + \frac{\theta}{1-\theta}.
\end{aligned}
$$

Thus $v' = \frac{1}{1-\theta} v + \frac{\theta}{1-\theta} e$.

Our assumption that $x^T s = n\mu$ can be translated into $v^T e = 0$ since $v^T e = \sum_j v_j = \sum_j \left(\frac{x_j s_j}{\mu} - 1 \right) = \frac{x^T s}{\mu} - n$. Therefore, we get that

$$
\|v'\|^2 = \frac{1}{(1-\theta)^2} \cdot \|v\|^2 + \frac{\theta^2}{(1-\theta)^2} \cdot \|e\|^2 = \frac{1}{(1-\theta)^2} \cdot \|v\|^2 + \frac{\theta^2}{(1-\theta)^2} n.
$$

Taking square roots, we get the desired result. □