10-725: Optimization **Fall 2013**

Lecture 17: October 22

Lecturer: Ryan Tibshirani/Barnabas Poczos *Scribe: Humphrey Hu*

This lecture discusses barrier methods for constrained optimization problems.

17.1 Barrier Method Formulation

The standard constrained minimization problem (P):

$$(P) : \min_{x \in S} f(x)$$

is modified to create a penalized unconstrained problem $(P(c))$:

$$(P(c)) : \min_{x} f(x) + \frac{1}{c} B(x)$$

The barrier penalty $B(x)$ must have the following three properties:

1. $B(x)$ must be continuous
2. $B(x) \geq 0 \forall x \in int(S)$
3. $\lim_{x \to \partial S} B(x) = \infty$

Example:

Consider the following constraint on x:

$$a \leq x \leq b$$

A potential barrier function is:

$$g_1(x) = x - b$$

$$g_2(x) = a - x$$

$$B(x) = -\left(\frac{1}{g_1(x)} + \frac{1}{g_2(x)} \right)$$

17.2 Log Barrier

For a system of constraints $g_i(x) \leq 0$ $i = 1, \ldots, m$, the log barrier penalty is defined as:

$$B(x) = \sum_{i=1}^{m} \log -g_i(x)$$

It can be verified that this penalty function possesses the aforementioned properties. Consider now a schedule of barrier weights c_i with the following ordering:

$$0 < c_i < c_2 < \ldots < c_k < c_{k+1} \to \infty$$

Let $\mu_k = \frac{1}{c_k}$. The solutions to the barrier problems are:

$$x_k = \arg\min_{x \in int(S)} f(x) + \mu_k B(x)$$

Define $r(\mu_k, x_k) = f(x_k) + \mu_k B(x_k)$. This method leads to the following lemmas:

Lemma 17.1 $r(\mu_k, x_k) \geq r(\mu_{k+1}, x_{k+1})$

Lemma 17.2 $B(x_k) \leq B(x_{k+1})$

Lemma 17.3 $f(x_k) \geq f(x_{k+1})$

Lemma 17.4 $f(x^*) \leq f(x_{k+1}) \leq f(x_k) \leq r(\mu_k, x_k)$

The proof for lemma 17.1 is as follows:

Proof:

$$\begin{aligned} r(\mu_k, x_k) &= f(x_k) + \mu_k B(x_k) \\ &\geq f(x_k) + \mu_{k+1} B(x_k) \text{ as } c_k < c_{k+1} \\ &\geq f(x_{k+1}) + \mu_{k+1} B(x_{k+1}) \text{ as } x_{k+1} \text{ is optimal for } \mu_{k+1} \\ &= r(c_{k+1}, x_{k+1}) \end{aligned}$$

■

The proof for lemma 17.2 is as follows:

Proof:

(1) $f(x_k) + \mu_k B(x_k) \leq f(x_{k+1}) + \mu_k B(x_{k+1})$ as x_k is optimal with c_k
(2) $f(x_{k+1}) + \mu_{k+1} B(x_{k+1}) \leq f(x_k) + \mu_{k+1} B(x_k)$ as x_k is optimal with c_{k+1}
Combining (1) and (2) $\Rightarrow$

$$\begin{aligned} \mu_k B(x_k) + \mu_{k+1} B(x_{k+1}) &\leq \mu_k B(x_{k+1}) + \mu_{k+1} B(x_k) \\ &\Rightarrow (\mu_k - \mu_{k+1}) B(x_k) \leq (\mu_k - \mu_{k+1}) B(x_{k+1}) \\ &\Rightarrow B(x_k) \leq B(x_{k+1}) \text{ as } \mu_{k+1} < \mu_k \end{aligned}$$

■

The proof for lemma 17.3 is as follows:

Proof:

$$\begin{aligned} f(x_k) + \mu_{k+1}B(x_k) &\geq f(x_{k+1}) + \mu_{k+1}B(x_{k+1}) \text{ as } x_{k+1} \text{ is optimal with } c_{k+1} \\ &\geq f(x_{k+1}) + \mu_{k+1}B(x_k) \text{ using Lemma 17.2} \\ &\Rightarrow f(x_k) \geq f(x_{k+1}) \end{aligned}$$

■

The proof for lemma 17.4 is as follows:

Proof:

$$\begin{aligned} f(x^*) &\leq f(x_k) \text{ as } x^* \text{ is optimal in } S \text{ and } x_k \in S \\ &\leq f(x_k) + \mu_k B(x_k) \\ &= r(\mu_k, x_k) \end{aligned}$$

■

17.3 Primal Log Barrier Method for LPs

Given a primal LP as:

$$\min_x c^T x \ s.t. \ Ax = b, \ x \geq 0, \ x \in \mathcal{R}^n$$

its corresponding barrier penalized problem $B(\mu)$ is:

$$B(\mu) : \min_x c^T x - \mu \sum_{j=1}^{n} \log(x_j) \ s.t. \ Ax = b$$

For simplicity we define a diagonal matrix $(D_x)_{ii} = x_i$. Let e be a $n \times 1$ vector of ones. The derivatives of the objective function are:

$$\begin{aligned} g(x) = \nabla f(x) = c - \mu D_x^{-1} e \\ G(x) = \nabla^2 f(x) = \mu D_x^{-1} \end{aligned}$$

Let $\bar{x}$ be a feasible solution to the penalty function to $B(\mu)$ and define $x^+ = \bar{x} + \Delta x$. We can approximate $f(x)$ with a quadratic (2nd order) Taylor expansion:

$$\begin{aligned} f(x) &= f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x}) + \frac{1}{2}(x - \bar{x})^T \nabla^2 f(x)(x - \bar{x}) \\ &= f(\bar{x}) + g(\bar{x})^T \Delta x + \frac{1}{2} \Delta x^T G(x) \Delta x \\ &= Q(\Delta x) \end{aligned}$$

We can now formulate the function as the following LP:

$$\min_{\Delta x} Q(\Delta x)$$

$$\begin{aligned} s.t.\ & A(\bar{x} + \Delta x) = b \\ & A(\bar{x}) = b \\ & \Rightarrow A\Delta x = 0 \end{aligned}$$

This LP has the following Lagrangian:

$$L(\Delta x, \pi_x) = g(\bar{x})^T \Delta x + \frac{1}{2}\Delta x^T G(\bar{x})\Delta x 0 \pi_x^T \left[A\Delta x\right]$$

$$\begin{aligned} \partial_{\Delta x} L &= g(\bar{x}) + G(\bar{x})\Delta x - (\pi_x^T A)^T = 0 \\ &= c - \mu D_x^{-1} e + \mu D_x^{-2} \\ \partial_{\pi_x} L &= -A\Delta x = 0 \\ & \begin{bmatrix} \mu D_x^{-2} & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} -\Delta x \\ \pi_x \end{bmatrix} = \begin{bmatrix} c - \mu D_x^{-1} e \\ 0 \end{bmatrix} \end{aligned}$$

This linear system can be used to solve for Δx and π_x. We can enforce the inequality and equality constraints with log penalty terms:

$$\min f(x) = c^T x - \mu \sum_{j=1}^{n} \log x_j,\ s.t.\ Ax = b$$

$$\min_{Ax \leq b} f(x) = c^T x - \mu \sum_{i=1}^{n} \log b_i - a_i^T x,\ \text{where } a_i \text{ is the } i\text{th row of } A$$

Iteratively solving this with the scheduled μs can be viewed as an interior point method, where $x(\mu^*)$ is the central path.

17.4 Primal-Dual Log Barrier Problem

17.4.1 Formulations

Symmetric Form I:

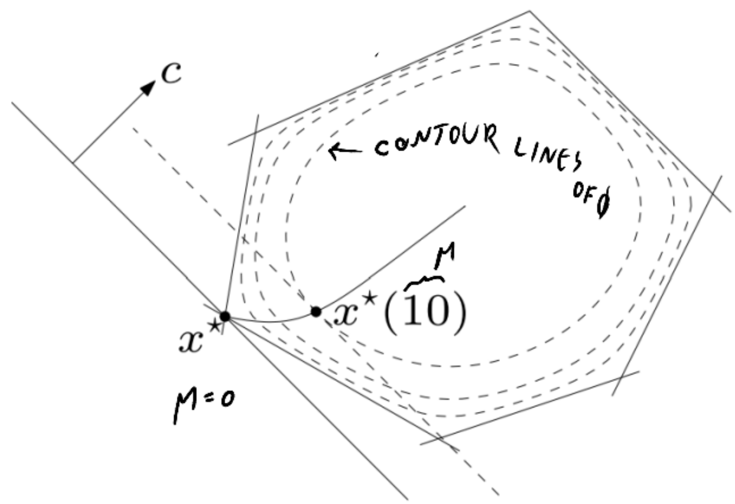

Figure 17.1: Interior method path

$$
\begin{aligned}
(P) \max_x c^T x \\
s.t.\ Ax \leq b \\
x \geq 0 \\
(D) \min_y y^T b \\
s.t.\ y^T A \geq c^T \\
y \geq 0
\end{aligned}
$$

Duality gap: $c^T x \leq y^T Ax \leq y^T b$

Symmetric Form II:

$$
\begin{aligned}
(P) \min_x c^T x \\
s.t.\ Ax \geq b \\
x \geq 0 \\
(D) \max_y y^T b \\
s.t.\ y^T A \leq c^T \\
y \geq 0
\end{aligned}
$$

Duality gap: $y^T b \leq c^T x$

Asymmetric Form I:

$$
\begin{aligned}
(P) \max_x c^T x \\
s.t.\ Ax \leq b \Rightarrow Ax + s = b \\
x \in \mathcal{R}^n \\
(D) \min_y y^T b \\
s.t.\ y^T A = c^T \\
y \geq 0
\end{aligned}
$$

Duality gap: $b^T y - c^T x = (Ax + s)^T y - y^T Ax = s^T y \geq 0$

Asymmetric Form II:

$$
\begin{aligned}
(P) \max_x c^T x \\
s.t.\ Ax = b \\
x \geq 0 \\
(D) \min_y y^T b \\
s.t.\ y^T A \leq c^T \Rightarrow y^T A + s^T = c^T \\
y \in \mathcal{R}^m
\end{aligned}
$$

Duality gap: $c^T x - b^T y = s^T x \geq 0$

17.4.2 KKT Conditions

Stationarity

$$
\begin{aligned}
\nabla_x f(x) + y^T \nabla_x (b - Ax) = 0 \\
(c - \mu D_x^{-1} e)^T + y^T(-A) = 0 \\
c - \mu D_x^{-1} e = A^T y
\end{aligned}
$$

Let $s = \mu D_x^{-1} e$. The stationary condition rewritten is:

$$
\begin{aligned}
\frac{1}{\mu} D_x s = e \\
\frac{1}{\mu} D_x D_s e = e
\end{aligned}
$$

The overall conditions are:

$$
\begin{aligned}
&Ax = b,\ x > 0 \\
&A^T y + s = c \\
&\frac{1}{\mu} D_x D_s e = e \\
&s \geq 0
\end{aligned}
$$

Lemma 17.5 *If (x, y, s) is a solution of the above KKT conditions, then the following are true:*

- *x is feasible for (P)*
- *(y, s) is feasible for (D)*
- *$x^T s = e^T D_x D_s e = \mu e^T e = \mu n$*

The first two KKT conditions are linear in x and s, but the third is not. We can use instead the following β-approximation:

$$\|\frac{1}{\mu} D_x s - e\| \leq \beta$$

Lemma 17.6 *If $(\bar{x}, \bar{y}, \bar{s})$ is a β-approximate solution of the approximated KKT conditions and $0 \leq \beta < 1$, we have the following bounds on the duality gap:*

$$n\mu(1-\beta) \leq c^T \bar{x} - b^T \bar{y} \leq n\mu(1+\beta)$$

In addition, $\bar{x}$ is feasible for (P) and $(\bar{y}, \bar{s})$ are feasible for (D).

The proof for 17.6 is as follows:

Proof:

Primal feasibility tells us that $\bar{x} \geq 0$. We need to show that $\bar{s} \geq 0$.

$$
\begin{aligned}
&\|\frac{1}{\mu} D_x s - e\| \leq \beta \\
&\Rightarrow -\beta \leq \frac{1}{\mu} x_j s_j - 1 \leq \beta \\
&\Rightarrow (1-\beta)\mu \leq x_j s_j \leq \mu(\beta + 1) \\
&\Rightarrow s_j > 0
\end{aligned}
$$

The duality gap is then:

$$
\begin{aligned}
n\mu(1-\beta) &\leq \sum_{j=1}^{n} x_j s_j \\
&= x^T s \\
&\leq n\mu(\beta + 1)
\end{aligned}
$$

■

We now solve the optimization using Newton's method. At the current iteration $\bar{x}, \bar{y}, \bar{s}$, from KKT we have:

$$\begin{aligned} &A\bar{x} = b,\ \bar{x} > 0 \\ &A^T\bar{y} + \bar{s} = c,\ \bar{s} > 0 \\ &\frac{1}{\mu} D_{\bar{x}} D_{\bar{s}} e - e = 0 \end{aligned}$$

The update step consists of $\bar{x} \to \bar{x} + \Delta x$, $\bar{y} \to \bar{y} + \Delta y$, and $\bar{s} \to \bar{s} + \Delta s$. This gives us the following KKT expressions:

$$\begin{aligned} &A(\bar{x} + \Delta x) = b \\ &A^T(\bar{y} + \Delta y) + (\bar{s} + \Delta s) = c \\ &(D_x + D_{\Delta x})(D_s + D_{\Delta s})e = \mu e \end{aligned}$$

Combining with the previous iterate conditions, we have:

$$\begin{aligned} &A\Delta x = 0 \\ &A^T\Delta y + \Delta s = 0 \\ &D_{\bar{x}} D_{\Delta s} e + D_{\Delta x} D_{\bar{s}} e = \mu e - D_{\bar{x}} D_{\bar{s}} e - D_{\Delta x} D_{\Delta s} e \\ &\Rightarrow D_{\bar{x}} \Delta s + D_{\bar{s}} \Delta x = \mu e - D_{\bar{x}} D_{\bar{s}} e \end{aligned}$$

where we have dropped the $D_{\Delta x} D_{\Delta s}$ term since it is second order and small compared to the other terms. These conditions give us a system of linear equations to solve for Δx, Δy, and Δs. The barrier coefficient is usually scaled at each iteration as $\mu_{k+1} = \alpha \mu_k$ or set to $\bar{x}^T \bar{s} \frac{n}{10}$, where n is the current duality gap.

17.4.3 Algorithm

The final primal dual algorithm is given below:

Step 0 Initialization
Start with a feasible point (x^0, y^0, s^0)

Step 1 Newton Step
Solve for Δx, Δy, and Δs using the following equations:
$A\Delta x = 0$
$A^T y + \Delta s = 0$
$D_{\bar{s}} \Delta x + D_{\bar{x}} \Delta s = \mu e - D_{\bar{x}} D_{\bar{s}} e$

Step 2 Update Step
$x^{k+1} = x^k + \Delta x$
$y^{k+1} = y^k + \Delta y$
$s^{k+1} = s^k + \Delta s$

Step 3 Check
$k = k + 1$
$\mu_{k+1} = \alpha \mu_k$ or $\frac{1}{10} \frac{\bar{x}^T \bar{s}}{n}$, where n is the current duality gap
Go back to Step 1 until convergence

Theorem 17.7 *Suppose that $(\bar{x}, \bar{y}, \bar{s})$ is a β-approximate solution of $P(\mu)$ for some $0 \leq \beta < \frac{1}{2}$. Let $(\Delta x, \Delta y, \Delta s)$ be the solution of the primal-dual newton system, and let $(x', y', s') = (\bar{x}, \bar{y}, \bar{s}) + (\Delta x, \Delta y, \Delta s)$. Then (x', y', s') is a $\frac{1+\beta}{(1-\beta)^2}\beta^2$-approximate solution of $P(\mu)$.*