10-725/36-725: Convex Optimization **Fall 2016**

Lecture 15: October 26

Lecturer: Ryan Tibshirani *Scribes: Zhiqian Qiao, Weikun Zhen*

Note: *LaTeX template courtesy of UC Berkeley EECS dept.*

15.1 Review

15.1.1 Newton's Method

Two different ways using Newton's method: root finding and optimization.

For root finding, the goal is to solve a function $F(x) = 0$, which is typically nonlinear. The update rule of Newton's method in this case is

$$x^+ = x - F'(x)^{-1}F(x)$$

For optimization, the goal is to minimize function $f(x)$ and the update rule is

$$x^+ = x - \nabla^2 f(x)^{-1}\nabla f(x)$$

It can be thought of finding a root for the optimality condition: $\nabla f(x) = 0$ or minimizing a quadratic approximation of $f(x)$.

If we assume f is strongly convex and both is gradient ∇f and Hessian $\nabla^2 f$ are Lipschitz, then when x is close enough to x^*, Newton's method guarantees quadratic convergence.

Note that in practical, damped Newton's method is more popular for global convergence, which has update rule as

$$x^+ = x - t\nabla^2 f(x)^{-1}\nabla f(x)$$

15.1.2 Newton's Method with linear constraints

When we are given a minimization problem with linearly equality constraint:

$$\begin{aligned} \min \quad & f(x) \\ \text{subject to} \quad & Ax = b \end{aligned}$$

and would like to use Newton's method to solve it, some changes need to be made for the update rule in order to maintain the equality constraints in each step. Specifically, we will update

$$x^+ = x + tv$$

where

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} v \\ w \end{bmatrix} = - \begin{bmatrix} \nabla f(x) \\ Ax - b \end{bmatrix}$$

which is the Newton's update to find the root for the KKT conditions of the linearly constrained problem. And the root finding problem is given below:

$$\begin{bmatrix} \nabla f(x) + A^T y \\ Ax - b \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Details can be found in the notes of last lecture.

15.2 Barrier Method

15.2.1 Motivation

Special case:

$$\begin{aligned} \min_x \quad & c^T x \\ \text{subject to} \quad & Dx \leq e \end{aligned}$$

Solving this problem with Newton's method will encounter the problem of figuring out the binding and non-binding in the constraints, namely which inequality becomes equality. This can be understood as the solution lies on the boundary of the feasible set C which is defined as

$$C := \{x : h_i(x) \leq 0, i = 1, \cdots, n\}$$

This is the major problem for this kind of problem and it's also the goal to introduce the barrier method. The idea of barrier is relax the inequality constraints and introduce a barrier function that intuitively pushes the solution away from the boundary of feasible set. Formally, consider the convex optimization problem

$$\begin{aligned} \min_x \quad & f(x) \\ \text{subject to} \quad & Ax = b \\ & h_i(x) \leq 0, i = 1, \cdots, n \end{aligned}$$

It is equivalent to

$$\begin{aligned} \min_x \quad & f(x) + I_C(x) \\ \text{subject to} \quad & Ax = b \end{aligned}$$

Then the $I_C(x)$ is further substituted with a barrier function which avoid the boundary of C. In this way, we work around the binding and non-binding headache.

15.2.2 Log Barrier Function

First of all the definition of log barrier function for set C, assuming non-empty, is as follows

$$\phi(x) = -\sum_{i=1}^{m} \log(-h_i(x))$$

Then the original problem is approximated using the log barrier function:

$$\begin{aligned} \min_x \quad & f(x) + \frac{1}{t}\phi(x) \\ \text{subject to} \quad & Ax = b \end{aligned}$$

where $t > 0$. It is also equivalent to

$$\begin{aligned} \min_x \quad & tf(x) + \phi(x) \\ \text{subject to} \quad & Ax = b \end{aligned}$$

Note that when $h_i(x) \to 0$, $\phi(x) \to \infty$. This guarantees that we are always trying to push the value of x away from the boundary to minimize the objective function. Now the objective function is smooth, we can apply Newton's method to solve it.

Another observation is that the value of t determines how far away we are from the optimal value of original problem. If t is small, then the second term $\phi(x)$ dominates and the optimal solution will be far away from the boundary. If t is large, the first term $f(x)$ dominates, and the solution will be very close to the boundary. More detail will be provided in the section of duality gap about the bound of how far we are from the optimal value.

15.2.3 Central Path

Since barrier is an approximation of the original problem and the parameter t affects how far the optimal solution with the barrier function to the optimal solution of the original objective function. So different t leads to different solutions, and the set of solutions is called central path. Specifically, it is the set $\{x^*(t) : t > 0\}$.

Under suitable condition, the set is a smooth path and as $t \to \infty$, we'll have $x^*(t) \to x^*$.

15.2.4 Perturbed KKT

Consider the KKT conditions of the original problem:

$$\begin{aligned} \text{stationary:} \quad & \nabla f(x^*) + \sum_{i=1}^m u_i^* \nabla h(x^*) + A^T v^* = 0 \\ \text{primal feasibility:} \quad & Ax^* = b, h_i(x^*) \le 0 \\ \text{dual feasibility:} \quad & u_i^* \ge 0 \\ \text{complementary slackness:} \quad & h_i(x^*) u_i^* = 0 \end{aligned}$$

And consider the KKT conditions of the barrier problem:

$$\begin{aligned} \text{stationary:} \quad & t\nabla f(x^*(t)) + \sum_{i=1}^m \frac{1}{h_i(x^*(t))} \nabla h(x^*) + A^T w^* = 0 \\ \text{primal feasibility:} \quad & Ax^*(t) = b, h_i(x^*(t)) < 0 \\ \text{dual feasibility:} \quad & \\ \text{complementary slackness:} \quad & \end{aligned}$$

Now we let

$$u_i(t) := -\frac{1}{th_i(x^*(t))}, \quad v := \frac{1}{t}w$$

we have another formulation of KKT conditions for the barrier problem:

$$\begin{aligned}
\text{stationary:} \quad & \nabla f(x^*(t)) + \sum_{i=1}^{m} u_i(t)\nabla h(x^*) + A^T v^* = 0 \\
\text{primal feasibility:} \quad & Ax^*(t) = b, h_i(x^*(t)) < 0 \\
\text{dual feasibility:} \quad & u_i(t) > 0 \\
\text{complementary slackness:} \quad & u_i(t)h_i(x^*(t)) = -\frac{1}{t}
\end{aligned}$$

Compared with the original KKT condition, the KKT condition for the barrier problem is tweaked and the complementary condition is replaced with $u_i(t)h_i(x^*(t)) = -\frac{1}{t}$.

15.2.5 Duality Gap

In this section, we'll show the gap between central path $x^*(t)$ and x^*.

By convexity we have

$$f(x^*(t)) - f(x^*) \leq \nabla f(x^*(t))^T(x^*(t) - x^*)$$

and

$$h_i(x^*(t)) - h_i(x^*) \leq \nabla h_i(x^*(t))^T(x^*(t) - x^*), i = 1, \cdots, m$$

Then we have

$$\begin{aligned}
& f(x^*(t)) - f(x^*) + \sum_{i=1}^{m} u_i(t)(h_i(x^*(t)) - h_i(x^*)) \\
\leq & \left(\nabla f(x^*(t)) + \sum_{i=1}^{m} u_i(t)\nabla h_i(x^*(t))\right)^T (x^*(t) - x^*) \\
= & (-A^T v)^T(x^*(t) - x^*) \\
= & 0
\end{aligned}$$

Knowing that, we have

$$\begin{aligned}
f(x^*(t)) - f(x^*) &= -\sum_{i=1}^{m} u_i(t)(h_i(x^*(t)) - h_i(x^*)) \\
&= -\sum_{i=1}^{m} u_i(t)h_i(x^*(t)) + \sum_{i=1}^{m} u_i(t)h_i(x^*) \\
&= \frac{m}{t} + \sum_{i=1}^{m} u_i(t)h_i(x^*) \\
&\leq \frac{m}{t}
\end{aligned}$$

This is going to be a useful stopping criterion to guarantee a required precision.

15.2.6 Barrier Method

15.2.6.1 Barrier method v.0

Naively, since we know $x^*(t) \to x^*$ as $t \to \infty$ and $f(x^*(t)) - f(x^*) \leq \frac{m}{t}$, we can just set $t = \frac{m}{\epsilon}$, where $\epsilon > 0$ is the error precision we want to achieve. This is too difficult to solve since it aims to find a point near the end of the central path. Thus it may cause numerical issues and it is also way too slow to converge. We want a method generating points along the central path.

15.2.6.2 Barrier method v.1

A better way to solve the barrier problem is increase the value of t sequentially. Algorithm is described as follows:

1. Pick $t^{(0)} > 0$ and $k := 0$

2. Solve the barrier problem for $t = t^{(0)}$ to produce $x^{(0)} = x^*(t)$ (see more in feasibility methods)

3. While $\frac{m}{t} > \epsilon$

   Pick $t^{(k+1)} = \mu t^{(k)}$ for $\mu > 1$

   Solve the barrier problem at $t = t^{(k+1)}$, using Newton's method initialized at $x^{(k)}$, to produce $x^{(k)} = x^*(t)$.

Note that the step solving the barrier problem is called a centering step.

There are some considerations of the choice of μ and $t^{(0)}$:

If μ is too small, then many outer iterations might be needed; if μ is too big, then Newtons method (each centering step) might take many iterations to converge. If $t^{(0)}$ is too small, then many outer iterations might be needed; if $t^{(0)}$ is too big, then the first Newtons solve (first centering step) might require many iterations to compute $x^{(0)}$.

15.2.6.3 Barrier method v.2

The previous method requires 'solve' the barrier problem, which means the $x^{(k)}$ is exactly on the central path. In practice, we don't need to solve the barrier problem at each centering step, an approximate solution is enough to lead to the final solution for the original problem.

1. Pick $t^{(0)} > 0$ and $k := 0$

2. Solve the barrier problem for $t = t^{(0)}$ to produce $x^{(0)} \approx x^*(t)$ (see more in feasibility methods)

3. While $\frac{m}{t} > \epsilon$

   Pick $t^{(k+1)} = \mu t^{(k)}$ for $\mu > 1$

   Solve the barrier problem at $t = t^{(k+1)}$, using Newton's method initialized at $x^{(k)}$, to produce $x^{(k)} \approx x^*(t)$.

15.2.7 Convergence Analysis

The barrier method after k centering iterations will satisfy

$$f(x^{(k)}) - f^* \leq \frac{m}{\mu^k t^{(0)}}$$

which means linear convergence rate.

Recall that for linear convergence: $k = \log(\frac{1}{\epsilon})$; for quadratic convergence: $k = \log\log(\frac{1}{\epsilon})$. So quadratic convergence is exponentially faster than linear convergence.

15.2.8 Feasibility Methods

In the barrier method section, we are assuming that a strictly feasible point for the first centering step is given. To find such a strictly feasible point, we solving following linear programming problem:

$$\begin{aligned} \min_{x,s} \quad & s \\ \text{subject to} \quad & h_i(x) \leq s \\ & Ax = b \end{aligned}$$

Note that we only need to find a strictly negative s and the corresponding x would be a strictly feasible point for the barrier problem.

An alternative is to solve the following problem:

$$\begin{aligned} \min_{x,s} \quad & 1^T s \\ \text{subject to} \quad & h_i(x) \leq s_i \\ & Ax = b \end{aligned}$$

Compared with the previous approach, each inequality has its own invisibility variable s_i. One advantage is that, the non-zero entries of s will tell us which of the constraints is not feasible.

15.2.9 Formal Barrier Methods

Suppose ϕ is a self-condordant barrier with parameter ν, which means

$$\lambda(x)^2 = \nabla\phi(x)(\nabla^2\phi(x))^{-1}\nabla\phi(x) \leq \nu$$

Now approximate the problem

$$\begin{aligned} \min_{x} \quad & c^T x \\ \text{subject to} \quad & x \in \bar{D} \end{aligned}$$

with

$$\min_{x} \quad tc^T x + \phi(x)$$

For convenience let $\phi_t(x) := tc^T x + \phi(x)$ and let $\lambda_t(x)$ denote the corresponding Newton decrement. Then we have a key observation saying that for $t^+ > t$

$$\lambda_{t^+}(x) \leq \frac{t^+}{t}\lambda_t(x) + \left(\frac{t^+}{t} - 1\right)\sqrt{\nu}$$

Then according to the theorem that if $\lambda_t(x) \leq \frac{1}{9}$ and $\frac{t^+}{t} \leq 1 + \frac{1}{s\sqrt{\nu}}$, then $\lambda_{t^+}(x) \leq \frac{1}{9}$, if we start with $x^{(0)}, t^{(0)}$, such that $\lambda_{t^{(0)}}(x^{(0)}) \leq \frac{1}{9}$ nad choose $\mu = 1 + \frac{1}{s\sqrt{\nu}}$ in the barrier method, one Newton iteration suffices at each centering step.