10-725/36-725: Convex Optimization	**Fall 2015**

Lecture 15: October 26

Lecturer: Lecturer: Javier Pena *Scribes: Scribes: Taylor Pospisil*

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

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

15.1 Review of last time: Newton's Method

Recall that Newton's method has two uses: root-finding and optimization

The update for root-finding of system of equations $F(x)$ is

$$x^+ = x - F'(x)^{-1}F(x) \tag{15.1}$$

The update for minimizing a function $f(x)$ is

$$x^+ = x - \nabla^2 f(x)^{-1}\nabla f(x) \tag{15.2}$$

The benefit of Newton's method is assuming conditions such as strong-convexity, Lipschitz gradient and Hessian, then when initialized close enough to the minimum the convergence is quadratic.

We can remove the condition about initialization by using the damped Newton's method

$$x^+ = x - t\nabla^2 f(x)^{-1}\nabla f(x) \tag{15.3}$$

For details see the previous lecture notes.

15.2 Review from last time: Equality constrained Newton's Method

Suppose we wish to solve the following optimization problem

$$\begin{aligned} \text{Minimize} \quad & f(x) \\ \text{Subject to} \quad & Ax = b \end{aligned} \tag{15.4}$$

We want to maintain the equality constraints on each Newton step so we have update

$$x^+ = x + tv \tag{15.5}$$

where we have

$$v = \arg, \min_{A(x+z)=b} \left(f(x) + \nabla f(x)^T z + \frac{1}{2} z^T \nabla^2 f(x) z \right) \tag{15.6}$$

To solve this we note that we can write the KKT conditions as the following set of equations

$$\begin{bmatrix} \nabla f(x) + A^T y \\ Ax - b \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{15.7}$$

When we employ the root-finding formulation of Newton's method to solving the KKT conditions we derive the update where v is part of the solution of

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} v \\ w \end{bmatrix} = - \begin{bmatrix} \nabla f(x) \\ Ax - b \end{bmatrix} \tag{15.8}$$

Again, for details see previous notes.

15.3 Barrier Method

Consider the general optimization problem

$$\begin{aligned} \text{Minimize} \quad & f(x) \\ \text{Subject to} \quad & Ax = b \\ & h_i(x) \leq 0, i = 1, \ldots, m \end{aligned} \tag{15.9}$$

Using the usual replacement of constraints with indicator functions we can rewrite this as

$$\begin{aligned} \text{Minimize} \quad & f(x) + I_C(x) \\ \text{Subject to} \quad & Ax = b \end{aligned} \tag{15.10}$$

The main difficulty arises when x is near the boundary. If we knew which constraints were binding we can simply apply equality-constrained methods.

The following methods are considered interior-point methods because they avoid the boundary and thus find solutions in the interior of the feasible set. They achieve the boundary avoidance by approximating the boundary $I_C(x)$ with some smooth replacement $\phi(x)$. We thus solve the following optimization problem

$$\begin{aligned} \text{Minimize} \quad & t f(x) + \phi(x) \\ \text{Subject to} \quad & Ax = b \end{aligned} \tag{15.11}$$

where t determines the strength of the barrier. As $t \to 0$ the barrier becomes stronger forcing the solution away from the boundaries. If any of the constraints are violated we have $\phi(x) = \infty$. As $t \to \infty$ the barrier becomes weaker and the solution can more closely approach the boundary.

Note that this problem can be solved with equality-constrained Newton's method given suitable conditions on f and ϕ.

15.3.1 Logarithmic Barrier Function

A common choice for the barrier function when the inequality constraints h_i are convex and twice-differentiable; then the logarithmic barrier function is

$$\phi(x) = -\sum_{i=1}^{m} log(-h_i(x)) \tag{15.12}$$

This approximates the indicator function for the set

$$\{x : h_i(x) < 0, i = 1, \dots, m\} \tag{15.13}$$

From calculus we note that

$$\nabla\phi(x) = -\sum_{i=1}^{m} \frac{1}{h_i(x)} \nabla h_i(x) \tag{15.14}$$

and

$$\nabla^2\phi(x) = \sum_{i=1}^{m} \frac{1}{h_i(x)^2} \nabla h_i(x) \nabla h_i(x)^T - \sum_{i=1}^{m} \frac{1}{h_i(x)} \nabla^2 h_i(x) \tag{15.15}$$

.

15.3.2 Perturbed KKT Conditions

Consider the KKT conditions for the original problem

- Stationarity
$$\nabla f(x^*) + \sum_{i=1}^{m} u_i^* \nabla h_i(x^*) + A^T v^* = 0 \tag{15.16}$$
- Primal Feasibility
$$Ax^* = b \tag{15.17}$$
and
$$h_i(x^*) \le 0 \tag{15.18}$$
- Dual Feasibility
$$u_i^* \ge 0 \tag{15.19}$$
- Complementary Slackness
$$u_i^* h_i(x^*) = 0 \tag{15.20}$$

Now compare the KKT conditions for this new problem with the log barrier.

Letting $u_i^*(t) = -\frac{1}{h_i(x^*(t))}$ and $v^*(t) = \frac{w}{t}$

- Stationarity

$$\nabla f(x^*(t)) + \sum_{i=1}^{m} u_i^*(t)\nabla h_i(x^*(t)) + A^T v^*(t) = 0 \tag{15.21}$$

- Primal Feasibility

$$Ax^*(t) = b \tag{15.22}$$

and

$$h_i(x^*) \leq 0 \tag{15.23}$$

- Dual Feasibility

$$u_i^*(t) \geq 0 \tag{15.24}$$

- Complementary Slackness

$$u_i^*(t)h_i(x^*) = -\frac{1}{t} \tag{15.25}$$

This is the condition that's been perturbed. As $t \to \infty$ then this condition approaches the condition of the original problem.

15.3.3 Central Path

We call the set of the solutions $\{x^*(t) : t > 0\}$ the **central path**. Under regularity conditions this is a smooth path which converges as $t \to \infty$ to the solution of the original problem x^*.

15.3.4 Duality Gap

We can bound the gap between the central path solution $x^*(t)$ and x^* as follows:

By convexity of f

$$f(x^*(t)) - f(x^*) \leq \nabla f(x^*(t))^T(x^*(t) - x^*) \tag{15.26}$$

and by convexity of h_i

$$h_i(x^*(t)) - h_i(x^*) \leq \nabla h_i(x^*(t))^T(x^*(t) - x^*) \tag{15.27}$$

We combine these two equations and the KKT conditions to achieve

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

We then have

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

15.3.5 Barrier Method v0

Given the result above it's natural to assume we can pick $t = \frac{m}{\epsilon}$ and simply solve for $x^*(t)$ to achieve $f(x^*(t)) - f^* \leq \epsilon$. Unfortunately for large t the barrier problem is too difficult to solve. So beginning at the end of the central path will not work.

15.3.6 Barrier Method v1

An better alternative approach is to generate points $x^*(t)$ along the central path using the following algorithm.

1. Pick $t^{(0)} > 0$
2. Solve the barrier problem with $t = t^{(0)}$ to obtain $x^{(0)} = x^*(t^{(0)})$
3. While $\frac{m}{t} > \epsilon$
   (a) Pick $t^{(k+1)} > t^k$
   (b) Solve barrier problem with $t = t^{(k)}$ using Newton's method initialized at $x^{(k}$ to obtain $x^{(k+1)} = x^*(t^{(k+1)})$. This is called the **centering set** because it moves the point onto the central path.

There are some considerations

- How to choose $t^{(k)}$ a common choice is to pick $\mu > 0$ and set $t^{k+1} = \mu t^{(k)}$. If μ is too small then we have many outer iterations before t grows large enough. If μ is too large then we have many inner iterations before the Newton solves converge.
- How to choose $t^{(0)}$ if $t^{(0)}$ is too small then again we may have many outer iterations before t grows large enough. If $t^{(0)}$ is too large than the first Newton solve will require many iterations.

In practice the barrier method is robust to the choice of μ and $t^{(0)}$. It is important to note that the choice of both of these parameters is scale dependent.

15.3.7 Convergence Analysis

Given our choice of computing $t^{(k+1)}$ and the earlier duality gap result after k steps we have

$$f(x^{(k)}) - f^* \leq \frac{m}{\mu^k t^{(0)}} \tag{15.28}$$

Thus if we want accuracy ϵ we require

$$\frac{\log(\frac{m}{t^{(0)}\epsilon})}{\log(\mu)} + 1$$

outer iterations. Thus we have linear convergence even though Newton's method is quadratic. The slow-down is due to the fact that we must obtain solutions along the central path.

15.3.8 Barrier Method v2

The previous algorithm required exact solutions to the intermediate problems. However, the central path solutions are not the main goal; they're merely the "means to an end" of computing the solution to the original problem. Thus, if we can avoid computing exact solutions we can save computational work.

1. Pick $t^{(0)} > 0$
2. Approximately the barrier problem with $t = t^{(0)}$ to obtain $x^{(0)} \approx x^*(t^{(0)})$
3. While $\frac{m}{t} > \epsilon$
   (a) Pick $t^{(k+1)} > t^k$
   (b) Approximately solve barrier problem with $t = t^{(k)}$ using Newton's method initialized at $x^{(k}$ to obtain $x^{(k+1)} \approx x^*(t^{(k+1)})$.

The key issue is determining how to closely to approximate the solutions. Typically only a few Newton's method iterations are required.

15.3.9 Feasibility Methods

For the sections above we have assumed that we can find a strictly feasible point to initialize the first Newton's method solve. We can obtain such a point by solving another optimization problem.

$$\begin{aligned} \text{Minimize} s \\ \text{Subject to} Ax = b \\ h_i(x) \leq s, i = 1, \ldots, m \end{aligned} \tag{15.29}$$

If there is a strictly feasible point for the original problem, then this optimization problem will have a negative objective value and we can take the solution as our initialization point. Indeed, we need not solve this optimization problem exactly; we can stop once we find a single negative objective.

To solve this problem we simply run the interior point method on this far easier problem.

Alternatively we can solve the problem

$$\begin{aligned} \text{Minimize} 1^T s \\ \text{Subject to} Ax = b \\ h_i(x) \leq s_i, i = 1, \ldots, m \\ s \geq 0 \end{aligned} \tag{15.30}$$

In this case, if the solution isn't feasible then the nonzero entries of s identify the constraints which cannot be satisfied. However, this does not find a strictly feasible point.

15.3.10 Formalizing barrrier method

Suppose that ϕ is self-concordant and

$$\lambda(x)^2 = \nabla\phi(x)(\nabla^2\phi(x))^{-1}\nabla\phi(x) \leq \nu \tag{15.31}$$

Then solving the problem

$$\begin{array}{ll} \text{Minimize} & c^T x \\ \text{Subject to} & x \in \overline{D} \end{array} \tag{15.32}$$

with the barrier method

$$\text{Minimize} \quad c^T x + \phi(x) \tag{15.33}$$

Then letting $\phi_t(x) = tc^T x + \phi(x)$ and λ_t being the corresponding Newton decrement we have that for $t^+ > t$

$$\lambda_{t^+}(x) \leq \frac{t^+}{t}\lambda_t(x) + \left(\frac{t^+}{t} - 1\right)\sqrt{\nu} \tag{15.34}$$

and thus if $\lambda_t(x) \leq \frac{1}{9}$ and $\mu = \frac{t^+}{t} \leq 1 + \frac{1}{8\sqrt{v}}$ then $\lambda_{t^+}(x^+) \leq \frac{1}{9}$ for the update

$$x^+ = x - (\nabla^2\phi_{t^+}(x))^{-1}\nabla\phi_{t^+}(x) \tag{15.35}$$

This means in practice that if we have $\lambda_{t(0} \leq \frac{1}{9}$ and $\mu = 1 + \frac{1}{8\sqrt{\nu}}$ then one Newton iteration suffices to approximate each centering step.