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

Lecture 15: Log Barrier Method

Lecturer: Ryan Tibshirani *Scribes: Pradeep Dasigi, Mohammad Gowayyed*

15.1 Introduction

Previously, we looked at Newton's method for minimizing twice differentiable convex functions with equality constraints. One of the limitations of this method is that we cannot deal with inequality constraints. The barrier method is a way to address this issue. Formally, given the following problem,

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

assuming that f, h_i are all convex and twice differentiable functions, all with domain $\mathbb{R}^n$, the log barrier is defined as

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

It can be seen that the domain of the log barrier is the set of strictly feasible points, $\{x : h_i(x) < 0, i = 1 \dots m\}$. Note that the equality constraints are ignored for the rest of this chapter, because those can be solved using the Newton's method directly.

15.1.1 Approximation of Indicator Functions

The idea behind the definition of the log barrier is that it is a smooth approximation of the indicator functions. More precisely, given that our problem with inequality constraints is to minimize

$$f(x) + \sum_{i=1}^{m} I_{\{h_i(x) \leq 0\}}(x)$$

we approximate it as follows

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

As t approaches ∞, the approximation becomes closer to the indicator function, as shown in Figure 15.1. Also, for any value of t, if any of the constraints is violated, the value of the barrier approaches infinity

15.1.2 Log Barrier Calculus

The gradient and hessian of the log-barrier, defined as follows will be useful in the Newton's method with log-barrier

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

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

15.2 Central Path

Now that we have defined the log barrier, we can rewrite our problem as

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

The central path of this problem is defined as $x^*(t)$, that the solution of the problem as a function of t. The necessary and sufficient conditions for $x^*(t)$ to be optimal, given by the KKT conditions are

$$\begin{aligned} Ax^*(t) = b, h_i(x^*(t)) < 0, i = 1 \ldots m \quad & \text{Primal feasibility} \\ t\nabla f(x^*(t)) - \sum_{i=1}^{m} \frac{1}{h_i(x)} \nabla h_i(x) + A^\intercal w \quad & \text{Stationarity} \end{aligned}$$

w is the dual variable used for defining the Lagrangian with the equality constraint.

15.2.1 Example

Consider the barrier problem for the linear program

$$\min \quad tc^\intercal x - \sum_{i=1}^{m} log(e_i - d_i^\intercal x)$$

where the barrier function corresponds to a polyhedral constraint $Dx \leq e$. Figure 15.2 shows the contours of the log barrier function at different values of t as dashed lines. It can be seen that the gradient of the log barrier contour at of the points on the central path is parallel to $-c$. This can also be seen from applying the stationarity condition.

$$0 = tc - \sum_{i=1}^{m} \frac{1}{e_i - d_i^\intercal x^*(t)} d_i$$

As we move along the central path, the contour gets closer to the actual polyhedron. Because of this geometry, log barrier method is called an interior point method, since we start an interior point, and move along the central path to get to the solution.

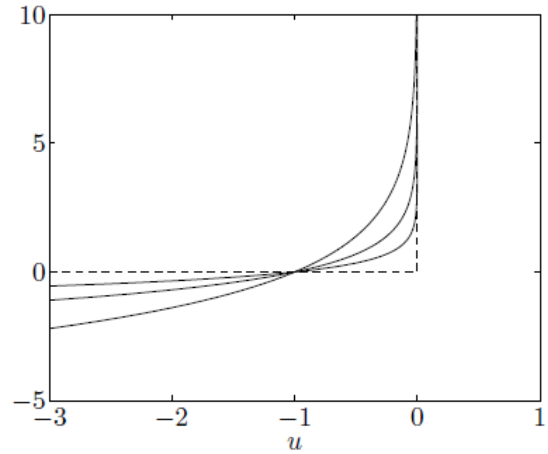

Figure 15.1: As t approaches ∞, the approximation becomes closer to the indicator function.

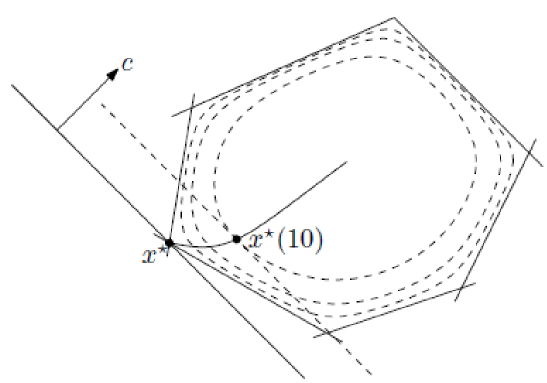

Figure 15.2: The contours of the log barrier function at different values of t as dashed lines.

15.2.2 Dual Points from Central Path

Given $x^*(t)$ and corresponding w, we can define the feasible dual points for the original problem (not the barrier problem).

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

We can prove that these are indeed dual feasible. Firstly, $u_i^*(t) > 0$, since $h_i(x^*(t)) < 0$ for all i. If we plug in the definitions of $u^*(t)$ and $v^*(t)$ into the stationarity condition for the central path, and divide the whole expression by t, we get

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

which shows that $(u^*(t), v^*(t))$ lies in the domain of $g(u, v)$.

We can now compute the minimum of the Lagrangian as follows

$$g(u^*(t), v^*(t)) = f(x^*(t)) + \sum_{i=1}^{m} u_i^*(t)h_i(x^*(t)) + v^*(t)^\intercal(Ax^*(t) - b)$$

From the definition of $u^*(t)$, the second term in the above expression becomes $-\frac{m}{t}$ and from primal feasibility, the third term becomes 0. Hence, we have,

$$g(u^*(t), v^*(t)) = f(x^*(t)) - \frac{m}{t}$$

That is, $\frac{m}{t}$ is the duality gap, and

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

This proves that as $t \to \infty$, we approach the solution. Also, the above condition can be used as a stopping criterion in our algorithm.

15.2.3 Perturbed KKT Conditions

We can think of the central path solution, $x^*(t)$ and the corresponding dual solution $(u^*(t), v^*(t))$ as solving the following conditions

$$\begin{aligned}
f(x^*(t)) + \sum_{i=1}^{m} u_i(x^*(t))\nabla h_i(x) + A^\intercal v^*(t) = 0 \quad & \text{Stationarity} \\
u_i^*(t)h_i^*(x^*) = \frac{1}{t}, i = 1 \dots m \quad & \text{Perturbed complementary slackness} \\
h_i^*(x^*) \leq 0, i = 1 \dots m \quad Ax^*(t) = b \quad & \text{Primal feasibility} \\
u_i(x^*(t)) \geq 0, i = 1 \dots m \quad & \text{Dual feasibility}
\end{aligned}$$

It can be seen that it is just the condition for complementary slackness that got *perturbed*, since the original condition would have been $u_i^*(t)h_i^*(x^*) = 0$. Hence, the solution above can be interpreted as solving something very close to KKT conditions. Also, as $t \to \infty$, the condition above approaches the original condition.

15.3 Algorithm for Barrier Method

Let us now try to define an algorithm for solving the barrier problem

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

Given a desired level of accuracy ϵ, we may want to solve this by setting $t = \frac{m}{\epsilon}$, and directly using the Newton's method. However, the required t would be huge, and it might make the algorithm numerically unstable. This is the reason why we need interior point methods, and we traverse the central path by increasing the values of t, using Newton's method at each step. This idea of solving by traversing over the central path is very similar to using warm starts, or more generally to the idea of traversing the solution path.

Formally, the following is the barrier method algorithm.

- Start with a value $t = t^{(0)} > 0$, and solve the barrier problem using Newton's method to get $x^{(0)} = x^*(t)$
- For barrier parameter $\mu > 1$, and $k = 1, 2, 3, \ldots$, repeat:
  - Solve barrier problem at $t = t^{(k)}$ using Newton's method initialized at $x^{(k-1)}$ to produce $x^{(k)} = x^*(t)$
  - Stop if $\frac{m}{t} \leq \epsilon$
  - Else, update $t^{(k+1)} = \mu t$

The first step in the outer loop is called the centering step, since it brings $x^{(k)}$ onto the central path.

15.3.1 Practical Considerations

For the above algorithm, we have to choose $t^{(0)}$ and μ. If μ is too small, many outer iterations might be needed, and if it is too big, each Newton's method step might need many iterations. Hence, we need to choose the value from a good range. The conditions for choosing $t^{(0)}$ are similar. If it is too small, many outer iterations are needed, and if it is too big, many iterations of Newton's method are needed for the centering step.

15.4 Convergence analysis of barrier method

Assuming that we can solve the centring steps exactly, then the following theorem immediately holds. If we cannot, then there would be other more complex bounds.

Theorem 15.1 *The barrier method after k centering steps satisfies:* $f(x^{(k)}) - f^* \leq \frac{m}{\mu^k t^{(0)}})$

where, $f(x^{(k)})$ is the objective value, f^* is the optimal objective, μ is the factor by which we multiply t every step, and m is the number of constraints.

So, in order to reach a desired accuracy level of ϵ, the theorem requires: $\frac{log(m/(t^{(0)}\epsilon)}{log\mu} + 1$ centring steps. Which is considered a linear convergence rate in terms of outer iterations that are required for the barrier method. There are more other fine-grained bounds, but this is the most basic one.

The previous bound is only for the outer iterations, but one might ask, how many total Newton iterations are required? And the answer, under sufficiently general conditions, is still in the order of $log(1/\epsilon)$ inner iterations to get an ϵ accurate solution.

There is also a much more tight bound that can be made under self-concordance, which we will talk about when we talk about the interior point method. In this bound, you can get rid of some of the parameters in this bound. It is not of this form but it gives you a more precise answer about the number of newton steps required.

Figure 15.3 shows the progress of the barrier method for a linear program with a growing number of constraints. We are trying to see how the convergence of barrier method depend on number of constraints. From the bound we can see that it should depend logarithmically on m, but what about number of newton iterations? We can see that increasing the number of constraints from 50 to 500 only required additional 10 newton steps, which is almost logarithmic, as we mentioned earlier.

15.5 Feasibility methods

Now, let's see how can we get the initial centering step. Remember that the algorithm requires that we have an initial strictly feasible point. Because if the point is not strictly feasible, the criteria is infinite. We get a feasible point by solving another problem:

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

We can think of s as the maximum violation of feasibility. If we can solve this problem so that the solution s is negative, we are good and the solution is strictly feasible. If s is positive, it is not good.

So, in total we run two interior point method, one to get a strictly feasible point to start with in the second one that actually optimize the original function. The first problem is much easier.

References

- S. Boyd and L. Vandenberghe (2004), "Convex optimization", Chapter 11.
- A. Nemirovski (2004), "Interior point polynomial time methods in convex programming", Chapter 4.
- J. Nocedal and S. Wright (2006), "Numerical optimization", Chapters 14 and 19.

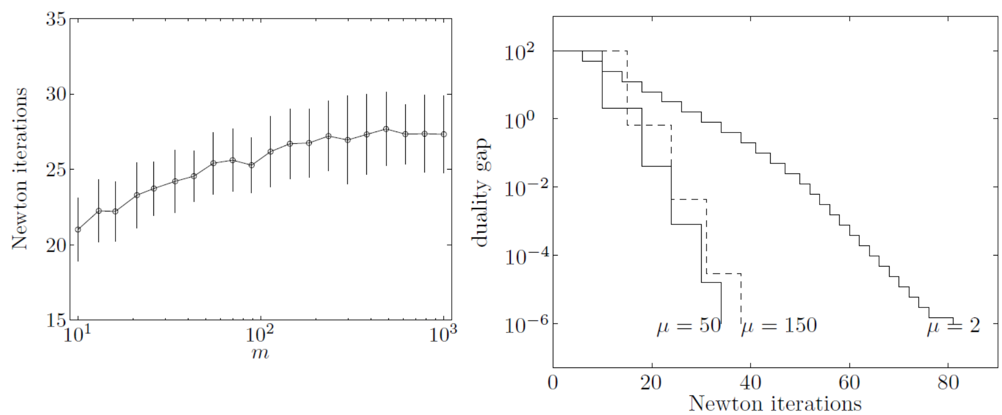

Figure 15.3: The progress of the barrier method for a linear program with a growing number of constraints.