Course Notes for EE227C (Spring 2018): Convex Optimization and Approximation

Instructor: Moritz Hardt

Email: `hardt+ee227c@berkeley.edu`

Graduate Instructor: Max Simchowitz

Email: `msimchow+ee227c@berkeley.edu`

April 24, 2018

25 Lecture 25: Interior Point Method

In last lecture, we have discussed about Newton's method. One advantage is its fast local convergence rate. However, there is no guarantee on global convergence for Newton's method. In this lecture, we'll present interior point method, which can be considered as an extension of Newton's method to ensure global convergence. First, the idea of interior point method will be introduced for general constrained optimization problems. Then the detailed algorithm and its convergence will be explained for linear programming.

25.1 Barrier Methods

We'll firstly introduce barrier methods which convert the inequality constraints into a function added upon the objective function of a optimization problem. Consider the following optimization problem:

$$
\begin{aligned}
\min_{x} \quad & f(x) \\
\text{s.t.} \quad & x \in \Omega, \\
& g_j(x) \leqslant 0, \ j = 1, 2, \cdots, r,
\end{aligned}
\tag{1}
$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $g_j : \mathbb{R}^n \to \mathbb{R}$ are given functions. f is continuous, and Ω is a closed set. For the rest of the lecture, we assume convex g_j and $\Omega = \mathbb{R}^n$. And we denote x^* as the optimal solution of the problem.

Definition 25.1 (Interior of the constraint region). The interior (relative to Ω) of the constraint region is defined as $S = \{x \in \Omega : g_j(x) < 0, \ j = 1, 2, \cdots, r\}$.

Assuming nonempty and convex S, we define a so-called barrier function $B(x)$ defined on S, such that $B(x)$ is continuous and $\lim_{g_j(x) \to 0_-} B(x) = \infty$. Two most common examples are logarithmic barrier function and inverse barrier function:

$$\text{Logarithmic:} \quad B(x) = -\sum_{j=1}^{r} \ln\{-g_j(x)\} \tag{2}$$

$$\text{inverse:} \quad B(x) = -\sum_{j=1}^{r} \frac{1}{g_j(x)}. \tag{3}$$

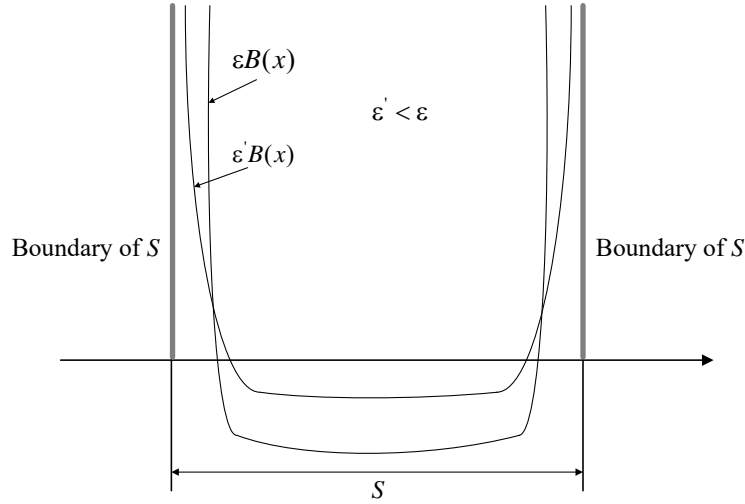Both of them are convex if all $g_j(x)$ are convex.

Figure 1: Form of a barrier term

Given $B(x)$, define a new cost function $f_\epsilon(x) = f(x) + \epsilon B(x)$, where ϵ is a positive real number. Then we can eliminate the inequality constraints in the original problem and obtain the following problem:

$$\begin{aligned} \min_{x} \quad & f_\epsilon(x) \\ \text{s.t.} \quad & x \in \Omega, \end{aligned} \tag{4}$$

The form of the barrier term $\epsilon B(x)$ is illustrated in Figure 1.

The barrier method is defined by introducing a sequence $\{\epsilon_t\}$ such that $0 < \epsilon_{t+1} < \epsilon_t$ for $t = 0, 1, 2, \dots$ and $\epsilon_t \to 0$. Then we find a sequence $\{x_t\}$ such that

$x_t \in \arg\min_{x \in S} f_{\epsilon_t}(x)$. Note that the barrier term $\epsilon_t B(x)$ goes to zero for all interior points $x \in S$ as $\epsilon_t \to 0$, allowing x_t to get increasingly closer to the boundary. Therefore, intuitively, x_t should approach x^* no matter x^* is in the interior or on the boundary of S. Its convergence is formalized in the following proposition.

Proposition 25.2. *Every limit point of a sequence $\{x_t\}$ generated by a barrier method is a global minimum of the original constrained problem.*

Proof. See Proposition 5.1.1 of [Ber16]. ∎

As $\Omega = \mathbb{R}^n$, Newton's method can be applied with properly selected stepsize to ensure that all iterates lies in S. Concretely, an initial interior point can be obtained for some large ϵ_0. Then in each iteration, we can use x_t as an initialization to find x_{t+1} by Newton's method. When ϵ_0 is large, it is easier to find a interior point. And since x_t is close to x_{t+1}, it is likely that x_t is in the local convergence region for Newton's method. Therefore, intuitively, global convergence of Newton's method can be enabled by using barrier methods.

25.2 Linear Programming

We'll now adopt the logarithmic barrier method to solve the linear programming (LP) problem defined as follows:

$$
\text{LP}: \quad \begin{aligned} &\min_{x} && c^T x \\ &\text{s.t.} && Ax \geqslant b \end{aligned}
\tag{5}
$$

where $A \in \mathbb{R}^{m \times n}$ with $m \geqslant n$ and rank(A)=n. Denote x^* as the optimal point.

First we write out the augmented cost function by the logarithmic barrier method, i.e.,

$$
f_\epsilon(x) = c^T x - \epsilon \sum_{j=1}^{m} \ln\left(A_j^T x - b\right).
\tag{6}
$$

where A_j^T is the j-th row of A. Define $x_\epsilon^* = \operatorname{argmin}_x f_\epsilon(x)$.

Fact: x_ϵ^* exists and is unique for any $\epsilon > 0$.

Proof. We can easily check that $f_\epsilon(x)$ is convex (as a sum of two convex functions). Therefore, the minimizer x_ϵ^* must exist and is unique.

Remark: For the convexity of f_ϵ, we can also check the second-order derivative, which is positive definite as shown in (8) later. ∎
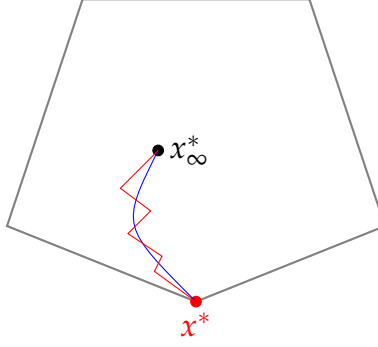
Figure 2: The central path

25.2.1 Central Path

The central path of the LP problem in 5 is depicted by the set of $\{x_\epsilon^* | \epsilon > 0\}$, as shown in Fig. 2.

Our goal is to design an algorithm that will approximately follow the central path. Assume that we already get a "good" enough initial point, then at every step, we apply one step of Newton's method. To guarantee that the algorithm converges, we need to answer the following two questions:

- Under what conditions that the single-step Newton method works?

- How to carefully update ϵ?

25.2.2 Newton Decrement

To apply the Newton's method, first we need to find out the first-order and second-order derivatives of f_ϵ. Note that

$$\nabla f_\epsilon(x) \;=\; c - \epsilon \sum_{j=1}^{m} \frac{A_j}{A_j^T x - b} \triangleq c - \epsilon A^T S^{-1} \mathbb{1} \tag{7}$$

$$\nabla^2 f_\epsilon(x) \;=\; \epsilon A^T S^{-2} A = \epsilon \sum_{j=1}^{m} \frac{A_j A_j^T}{S_j^2} \tag{8}$$

where $\mathbb{1} = [1, 1, \cdots, 1]^T \in \mathbb{R}^{m \times 1}$, $S = diag\{A_1^T x - b, A_2^T x - b, \cdots, A_m^T x - b\}$.

Then the Newton's update can be applied:

$$\bar{x} = x - [\nabla^2 f_\epsilon(x)]^{-1} \nabla f_\epsilon(x) = x - [\epsilon A^T S^{-2} A]^{-1} \left(c - \epsilon A^T S^{-1} \mathbb{1} \right) \tag{9}$$

Recall that the Newton's method finds the solution by making the first-order condition zero. To measure how much the Newton update will decrease the first-order approximation, we introduce the concept of Newton decrement.

4

Define the Newton decrement as

$$q^2(x, \epsilon) = \nabla f_\epsilon^T(x)[\nabla^2 f_\epsilon(x)]^{-1} \nabla f_\epsilon(x) \tag{10}$$

or $q(x, \epsilon) = [\nabla^2 f_\epsilon(x)]^{-1/2} \nabla f_\epsilon(x)$.

Note that the Newton decrement also relates to the difference between $f_\epsilon(x)$ and the minimum of its second-order approximation:

$$f_\epsilon(x) - \min_{\bar{x}} \left(f_\epsilon(x) + \nabla f_\epsilon(x)^T(\bar{x} - x) + (\bar{x} - x)^T \nabla^2 f_\epsilon(x)(\bar{x} - x) \right)$$

$$= f_\epsilon(x) - \left(f_\epsilon(x) - \frac{1}{2} \nabla f_\epsilon^T(x)[\nabla^2 f_\epsilon(x)]^{-1} \nabla f_\epsilon(x) \right)$$

$$= \frac{1}{2} \nabla f_\epsilon^T(x)[\nabla^2 f_\epsilon(x)]^{-1} \nabla f_\epsilon(x) \triangleq \frac{1}{2} q^2(x, \epsilon). \tag{11}$$

We now will use the Newton decrement to answer the above two questions: conditions for single-step Newton to work and rules for updating ϵ.

25.2.3 Propositions, Algorithms and Convergence Theorem

Proposition 25.3. *Assume $Ax > b$ and $\|q(x, \epsilon)\| < 1$, then we have*

$$c^T x - c^T x^* \leqslant 2\epsilon n. \tag{12}$$

In particular, if we maintain that x_t is interior point satisfying $Ax_t > b$, and $q(x_t, \epsilon_t) < 1$, then $c^T x_t$ converges to $c^T x^*$ as ϵ_t goes to 0, i.e., x_t converges to global optimum. However, the condition $q(x_t, \epsilon_t) < 1$ is not trivial.

Proposition 25.4. *If $Ax > b$, and $q(x, \epsilon) < 1$, then the pure Newton iterate step $\bar{x}$ satisfies,*

$$\|q(\bar{x}, \epsilon)\| \leqslant \|q(x, \epsilon)\|^2 \tag{13}$$

It ensures that $\|q(\bar{x}, \epsilon)\| < 1$ given $q(x, \epsilon) < 1$ and x is interior point. But we also want that $\|q(\bar{x}, \bar{\epsilon})\| < 1$ for some $\bar{\epsilon} < \epsilon$.

Proposition 25.5. *Assume $q(x, \epsilon) \leqslant \frac{1}{2}$, interior point $Ax > b$, put*

$$\bar{\epsilon} = (1 - \frac{1}{6\sqrt{n}})\epsilon, \tag{14}$$

then we have

$$\|q(\bar{x}, \bar{\epsilon})\| \leqslant \frac{1}{2} \tag{15}$$

These propositions indicates the following update rule,

$$x_{t+1} = x_t - \nabla^T f_{\epsilon_t}(x)^{-1} \nabla f_{\epsilon_t}(x_t) \tag{16}$$

$$\epsilon_t = (1 - \frac{1}{6\sqrt{n}})\epsilon \tag{17}$$

Theorem 25.6. *Suppose (x_0, ϵ_0) satisfies $Ax_0 > b$ and $\|q(x_0, \epsilon_0)\| \leqslant \frac{1}{2}$, then the algorithm converges in $\mathcal{O}(\sqrt{n} \log(n/\eta))$ iterations to η error, i.e., we have $c^T x_t \leqslant c^T x^* + \eta$ after $\mathcal{O}(\sqrt{n} \log(n/\eta))$ iterations.*

Proof. As Newton step maintains x_{t+1} in the interior, by using the three propositions above, we have

$$
\begin{aligned}
c^T x_t &\leqslant c^T x^* + 2\epsilon_t n \\
&= c^T x^* + 2(1 - \frac{1}{6\sqrt{n}})^t \epsilon_0 \\
&\leqslant c^T x^* + 2\exp(-\frac{t}{6\sqrt{n}})\epsilon_0
\end{aligned}
\tag{18}
$$

Therefore, to have a error of η, $t \geqslant \frac{6\sqrt{n}}{\epsilon_0} \log \frac{2n}{\eta}$. We can then conclude that the algorithm converges in $\mathcal{O}(\sqrt{n} \log(n/\eta))$ iterations to η error. $\blacksquare$

The algorithm stated above is the so-called short-step method. Although theoretical convergence rate is guaranteed, the combination of small decrease in ϵ and a single Newton step is slow in practice. Instead, a more practical method is the so-called long-step method, where ϵ is reduced in faster rate and several Newton steps are taken per iteration.

References

[Ber16] D.P. Bertsekas. *Nonlinear Programming*. Athena scientific optimization and computation series. Athena Scientific, 2016.