ECE 490: Introduction to Optimization **Fall 2018**

Lecture 24

Interior Point Method for Linear Programming

Lecturer: Bin Hu, Date:12/4/2018

We have introduced the idea of barrier methods (or interior point methods). To demonstrate their applications, we will talk about the related implementations for linear programming in this lecture.

24.1 Short-step Implementation

Consider a linear programming problem in the following form:

$$\begin{aligned} \text{minimize} \quad & c^\mathsf{T} x \\ \text{subject to} \quad & Ax \geq b \end{aligned} \tag{24.1}$$

where $A \in \mathbb{R}^{m\times n}$ with $m \geq n$ and $\text{rank}(A) = n$. In the above problem, we only have inequality constraints. Denote the j-th row of A as A_j^T. We can rewrite $Ax \geq b$ as m constraints: $A_j^\mathsf{T} x \geq b_j$ for $j = 1, \ldots, m$. Therefore, we can formulate the logarithmic barrier function as

$$f_\varepsilon(x) = c^\mathsf{T} x - \varepsilon \sum_{j=1}^{m} \ln(A_j^\mathsf{T} x - b)$$

It is straightforward to verify $\nabla^2 f_\varepsilon(x) = \varepsilon \sum_{j=1}^m \frac{A_j A_j^\mathsf{T}}{(A_j^\mathsf{T} x - b_j)^2} > 0$ under the assumption $\text{rank}(A) = n$. Therefore, for every $\varepsilon > 0$, we know f_ε has a unique solution and can be optimized using standard unconstrained optimization methods. However, when ε gets smaller, the function f_ε becomes more ill-conditioned and the optimization becomes harder.

Recall that the basic version of interior point methods iterates as $x_k = \arg\min f_{\varepsilon_k}(x)$. At the step $k+1$, one updates $\varepsilon_{k+1} = \rho \varepsilon_k$ where $\rho < 1$, and then calculate $x_{k+1} = \arg\min f_{\varepsilon_{k+1}}(x)$ by applying Newton's method with an initialization point at x_k.

It turns out that we do not need to solve $\arg\min f_{\varepsilon_{k+1}}(x)$ exactly. At every step $k+1$, we only need to perform one step for Newton update as $x_{k+1} = x_k - (\nabla^2 f_{\varepsilon_{k+1}}(x_k))^{-1} \nabla f_{\varepsilon_{k+1}}(x_k)$. This is the so-called short-step implementation of the interior point method. The gradient of $f_\varepsilon(x)$ can be calculated as $\nabla f_\varepsilon(x) = c - \varepsilon \sum_{j=1}^m \frac{A_j}{A_j^\mathsf{T} x - b_j}$. Therefore, the short-step method updates as

$$x_{k+1} = x_k - \left(\sum_{j=1}^{m} \frac{A_j A_j^\mathsf{T}}{(A_j^\mathsf{T} x - b_j)^2} \right)^{-1} \left(\frac{c}{\varepsilon_{k+1}} - \sum_{j=1}^{m} \frac{A_j}{A_j^\mathsf{T} x - b_j} \right)$$

where $\varepsilon_{k+1} = \rho\varepsilon_k$ with some $\rho < 1$. If the short-step method is initialized with reasonably good (x_0, ε_0) and the constant ρ is well chosen, then the short-step method is guaranteed to converge at a linear rate. Typically the theoretical guarantees state that if $\rho = 1 - \frac{1}{\gamma\sqrt{n}}$ and $\nabla f_{\varepsilon_0}(x_0)^\mathsf{T}[\nabla^2 f_{\varepsilon_0}(x_0)]^{-1}\nabla f_{\varepsilon_0}(x_0) \leq \beta$ for some γ and β, then the short-step method converges as a linear rate $(1 - \frac{M}{\sqrt{n}})$ where M depends on γ and β. Many results in this form exist in the literature. For example, one can choose $\gamma = 6$ and $\beta = \frac{1}{2}$.

Remarks on the initial conditions. The initial condition is important. The condition $\nabla f_{\varepsilon_0}(x_0)^\mathsf{T}[\nabla^2 f_{\varepsilon_0}(x_0)]^{-1}\nabla f_{\varepsilon_0}(x_0) \leq \beta$ just states that x_0 is not that far from the optimal point of $f_{\varepsilon_0}(x)$. Therefore, typically one has to solve the first optimization problem $\min f_{\varepsilon_0}(x)$ with some accuracy and then the rest of the iterations will work fine. To make things simple, in the homework, you are asked to start from the initial point $x_0 = 0$. This is a strictly feasible point if $b < 0$. However, this point may not be that close to the solution for $\min f_{\varepsilon_0}(x)$. Hence the short-step method is not guaranteed to work with such naive initializations. Increasing ρ sometimes helps stabilizing the short-step method if x_0 is not super far from the solution of $\min f_{\varepsilon_0}(x)$. Tuning ρ can definitely affect the algorithm performance.

24.2 Primal-Dual Implementation

Now we consider the linear programming in the following form.

$$\begin{aligned} \text{minimize} \quad & c^\mathsf{T}x \\ \text{subject to} \quad & Ax = b, \\ & x \geq 0 \end{aligned} \tag{24.2}$$

Recall the dual problem for the above LP is

$$\begin{aligned} \text{maximize} \quad & -b^\mathsf{T}\lambda \\ \text{subject to} \quad & c + A^\mathsf{T}\lambda - \mu = 0 \\ & \mu \geq 0 \end{aligned} \tag{24.3}$$

When solving (24.2), one tries to find the points satisfying the following KKT conditions:

$$c + A^\mathsf{T}\lambda - \mu = 0 \tag{24.4}$$
$$Ax = b \tag{24.5}$$
$$x \geq 0 \tag{24.6}$$
$$\mu \geq 0 \tag{24.7}$$
$$\mu(j)x(j) = 0, \ \forall j \tag{24.8}$$

where $\mu(j)$ and $x(j)$ denote the j-th entry of μ and x, respectively. Notice (24.4) and (24.5) are linear equations and can be efficiently solved. The two inequality conditions (24.6) and

(24.7) are also relatively easy to handle and let's ignore them for now. We will come back to this in the end of this note. The real difficult condition is the complementary slackness condition (24.8), which is bilinear in $x(j)$ and $\mu(j)$ for all j. Solving this bilinear equations is less efficient.

The primal-dual implementation of the interior point method aims at finding points satisfying the KKT conditions by making use of the primal and dual formulations simultaneously. Suppose we have initial points (x_0, λ_0, μ_0) satisfying (24.4) and (24.5). Clearly the complementary slackness condition (24.8) is violated since typically one has $\mu_0(j)x_0(j) \neq 0$. The idea of the primal-dual method is that we find (x_k, λ_k, μ_k) that satisfy (24.4), (24.5), and $\mu_k(j)x_k(j) = \frac{\sigma}{n}\mu_{k-1}^{\mathsf{T}}x_{k-1}$ where $0 < \sigma < 1$. Clearly $\mu_k^{\mathsf{T}}x_k = \sigma\mu_{k-1}^{\mathsf{T}}x_{k-1}$. Then as $k \to \infty$, we have $\mu_k(j)x_k(j) \to 0$. Then we will end up with points satisfying (24.4), (24.5) and (24.8) simultaneously.

Suppose (x_k, λ_k, μ_k) satisfy (24.4) and (24.5). How can we find points $(x_{k+1}, \lambda_{k+1}, \mu_{k+1})$ satisfying (24.4), (24.5), and $\mu_{k+1}(j)x_{k+1}(j) = \frac{\sigma}{n}\mu_k^{\mathsf{T}}x_k$? Denote $x_{k+1} = x_k + \Delta x$, $\lambda_{k+1} = \lambda_k + \Delta\lambda$, and $\mu_{k+1} = \mu_k + \Delta\mu$. Suppose (x_k, λ_k, μ_k) satisfy (24.4) and (24.5). Now if we have $A^{\mathsf{T}}\Delta\lambda - \Delta\mu = 0$ and $A\Delta x = 0$, we guarantee that $(x_{k+1}, \lambda_{k+1}, \mu_{k+1})$ satisfying (24.4) and (24.5). However, the condition $\mu_{k+1}(j)x_{k+1}(j) = \frac{\sigma}{n}\mu_k^{\mathsf{T}}x_k$ seems hard to enforce. Specifically, we need

$$\Delta\mu(j)x_k(j) + \mu_k(j)\Delta x(j) + \mu_k(j)x_k(j) + \Delta\mu(j)\Delta x(j) = \frac{\sigma}{n}\mu_k^{\mathsf{T}}x_k$$

Given x_k and μ_k, the above equation is bilinear in $\Delta\mu$ and Δx, causing trouble for computation. However, if we restrict $\Delta\mu$ and Δx to be not too large, then the higher order product $\Delta\mu(j)\Delta x(j)$ can be ignored and we have a linear equation in $\Delta\mu$ and Δx. Hence we can solve Δx, $\Delta\mu$, and $\Delta\lambda$ using the following linear equations

$$\begin{aligned} A^{\mathsf{T}}\Delta\lambda - \Delta\mu &= 0 \\ A\Delta x &= 0 \\ \Delta\mu(j)x_k(j) + \mu_k(j)\Delta x(j) &= \frac{\sigma}{n}\mu_k^{\mathsf{T}}x_k - \mu_k(j)x_k(j) \end{aligned}$$

Therefore, the idea for the primal-dual method is that we find a sequence of points that always satisfy simple linear conditions (24.4) and (24.5) in the KKT conditions and come closer and closer to satisfy the more complicated bilinear complementary slackness condition (24.8).

The final issue is that we also need $x \geq 0$ and $\mu \geq 0$. A straightforward way to handle this is that we initialize the update at points that strictly satisfy $x_0 > 0$ and $\mu_0 > 0$. Then we choose α_0 small enough such that $x_0 + \alpha\Delta x > 0$ and $\mu_0 + \alpha\Delta\mu > 0$. Basically at every step we can choose a stepsize α_k to make sure that $x_k > 0$ and $\mu_k > 0$ are still enforced. By doing this, we restrict our updates to be interior points of the inequality constraints (24.6) and (24.7). There are also other ways to handle these two inequality constraints. We skip the details here.