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

Lecture 23

Methods for General Constrained Optimization Problems

Lecturer: Bin Hu, Date:11/29/2018

In this lecture, we talk about several standard optimization methods for the following constrained problem

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & h_i(x) = 0,\ i = 1, \ldots, m \\ & g_j(x) \leq 0,\ j = 1, \ldots, l \end{aligned} \tag{23.1}$$

Specifically, we will talk about penalty and barrier methods.

23.1 Revisit Augmented Lagrangian Method

Suppose we are considering a constrained optimization problem with only equality constraints.

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & h(x) = 0 \end{aligned} \tag{23.2}$$

Recall that we have talked about the method of multipliers in the previous lectures. We can form the augmented Lagrangian $L_\rho(x, \lambda) = f(x) + \lambda^\mathsf{T} h(x) + \frac{\rho}{2}\|h(x)\|^2$ and then apply the method of multipliers.

$$\begin{aligned} x_{k+1} &= \arg\min_x L_\rho(x, \lambda_k) \\ \lambda_{k+1} &= \lambda_k + \rho h(x_{k+1}) \end{aligned}$$

The key issue here is how to select ρ. Since ρ is the stepsize for the update of λ_k, one has to choose it carefully. If ρ is too small, the convergence of λ_k becomes slow. If ρ is too large, the problem becomes ill-conditioned, and the update for λ_k can diverge. Typically we need to vary ρ as k increases. The initial value ρ_0 should not be too large to cause ill-conditioning at the first step. Then ρ_k should increase at a reasonable rate to help the convergence of λ_k. One practical scheme is to choose $\rho_{k+1} = \beta\rho_k$ where $\beta > 1$ is some fixed constant.

We can convert (23.1) to the following optimization problem with only equality constraints.

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & h_i(x) = 0,\ i = 1, \ldots, m \\ & g_j(x) + s_j^2 = 0,\ j = 1, \ldots, l \end{aligned} \tag{23.3}$$

In the above problem, both x and s_j together form the decision variables.

Then any method designed for (23.2) can be directly tailored for (23.1). Clearly we can apply the method of multipliers to the above problem. Notice both x and z are decision variables. But the decision variables z_j is only involved in the constraint and does not show up in the objective function. When applying the method of multipliers, we can first minimize the augmented Lagrangian with respect to z and then minimize it with respect to x.

23.2 Penalty Methods

The method of multipliers can be viewed as a special case of general penalty methods. When applying the method of multipliers, one does not enforce $h(x_k) = 0$ for all k. Therefore, the constraint $h(x) = 0$ is violated during the optimization process. The hope is that eventually $h(x_k)$ will converge to 0 as k increases. The term $\|h(x)\|^2$ in the augmented Lagrangian can be viewed as a penalty for violation of the equality constraint $h(x) = 0$. The larger $h(x)$ is, the more penalty is added. The penalty methods do not require the iterates to be strictly feasible and this can be beneficial for some problems. Sometimes the Lagrangian multipliers are not involved in the penalty methods. One just minimizes $f(x) + \frac{\rho_k}{2}\|h(x)\|^2$ directly. There are also other types of penalty functions that are used in the penalty methods. We skip the details here. The key message is that penalty methods do not require the iterates to be strictly feasible points for the original problem but do penalize the violation of constraints by adding a penalty term into the objective function.

23.3 Barrier Methods

Different from the penalty methods, barrier methods do enforce the iterations to be strictly feasible points. Hence barrier methods are also called interior point methods. Instead of converting the problem into a form of (23.3), barrier methods handle the inequality constraints directly by using the barrier functions.

For simplicity, consider the case where only inequality constraints are involved.

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & g_j(x) \leq 0, \ j = 1, \ldots, l \end{aligned} \tag{23.4}$$

The barrier method replaces the inequality constraints with barrier functions that are added to the objective functions. The barrier method involves solving a sequence of optimization problems that become harder and harder but approximate the original problem (23.4) better and better. Specifically, at every k, we solve the unconstrained minimization

$$\text{minimize}\, f(x) + \varepsilon_k B(x) \tag{23.5}$$

where ε_k is a monotone decreasing sequence converging to 0. The function $B(x)$ is the barrier function. Recall that the feasible set is $\{x : g_j(x) \leq 0,\ j = 1, \ldots, l\}$. We define the interior to be $\{x : g_j(x) < 0, \ j = 1, \ldots, l\}$. A key property for the barrier function is that $B(x)$ goes to∞ as x approaches the boundary of the set $\{x : g_j(x) \leq 0, \ j = 1, \ldots, l\}$ from its interior. A typical example for the barrier function is the logarithmic barrier function, i.e. $B(x) = -\sum_{j=1}^{l} \ln(-g_j(x))$. As ε_k approaches 0, $\varepsilon_k B(x)$ behaves as the indicator function for the set $\{x : g_j(x) \leq 0, \ j = 1, \ldots, l\}$. Suppose ε_k is sufficiently small. Then for the points near the boundary of the feasible set, $\varepsilon_k B(x)$ is very large. But for most points in the feasible set, we have $\varepsilon_k B(x) \approx 0$.

Suppose $x_k = \arg\min f(x) + \varepsilon_k B(x)$. Since $B(x)$ is not defined outside the feasible set, we know x_k is a feasible point. Hence the barrier method enforces x_k to be strictly feasible points during the optimization process. One can show that x_k converges to the solution for (23.4) as $\varepsilon_k \to 0$.

A natural question is why we need to solve the optimization in a sequential way? Why don't we start with some ε_0 that is extremely small? The answer is that the problem (23.5) is well-conditioned when ε_k is relatively large. Hence we can start with some relatively large ε_k and the problem (23.1) is relatively simple. As we decrease ε_k, the unconstrained optimization problem (23.5) approximates the original problem (23.4) better and better but becomes more and more ill-conditioned and difficult to solve. The hope is that the solutions for relatively larger ε_k provide some good initialization points when we attempt to solve the problem with smaller ε_k later. Specifically, given $\varepsilon_{k+1} < \varepsilon_k$, we know $f(x) + \varepsilon_{k+1} B(x)$ is more difficult to optimize compared with $f(x) + \varepsilon_k B(x)$. However, $\arg\min\{f(x) + \varepsilon_{k+1} B(x)\}$ may not be that far from $\{\arg\min f(x) + \varepsilon_k B(x)\}$. Therefore, we apply Newton's method to optimize $f(x) + \varepsilon_{k+1} B(x)$ with an initialization at $\arg\min\{f(x) + \varepsilon_k B(x)\}$, it is reasonable to expect that we will be able to quickly get an optimal point for $f(x) + \varepsilon_{k+1} B(x)$ under many situations.

The barrier method (or interior point method) works well for several important convex optimization problems including linear programming and moderate-size semidefinite programming. However, when you have a general non-convex constrained optimization problem, it is not a trivial task to decide which methods you want to apply. One can try to relax the given problems as linear programming or SDP (if possible) and then apply interior point methods (or other available solvers) to solve the relaxations and obtain upper bounds for the original problem. Another option is of course trying multiplier methods (or penalty methods) directly.