Advanced Mathematical Programming IE417

Lecture 14

Dr. Ted Ralphs

Reading for This Lecture

- Sections 8.1-8.5

One-dimensional Line Search

- One-dimensional line search is the fundamental subproblem for many non-linear algorithms.
- Given a function f, a current location x, and a direction d, we want to solve the following problem

$$\begin{aligned} &\min f(x+\lambda d) \\ &\text{s.t. } a \leq \lambda \leq b \end{aligned}$$

- Recall the typical iterative algorithm discussed in Chapter 7.

Line Search Methods

- Exact Methods
  - Solve the line search problem analytically.
  - Take the derivative with respect to λ and set it to zero.
- Iterative Methods
  - Methods using function evaluations.
  - Methods using derivatives.
  - Generally guaranteed to converge for pseudoconvex functions.

The Interval of Uncertainty

- The *interval of uncertainty* is the interval within which the optimal solution has to lie.
- Most derivative-free line search methods are based on iteratively reducing the interval of uncertainty.

Theorem 1. *Let $\Theta : \mathbb{R} \to \mathbb{R}$ be strictly quasiconvex over the interval $[a, b]$. Let $\lambda, \mu \in [a, b]$ be such that $\lambda < \mu$.*

- *If $\Theta(\lambda) > \Theta(\mu)$, then $\Theta(z) \geq \Theta(\mu)$ for all $z \in [a, \lambda)$.*
- *If $\Theta(\lambda) \leq \Theta(\mu)$, then $\Theta(z) \geq \Theta(\lambda)$ for all $z \in (\mu, b]$.*

Derivative-free Line Search

- The previous theorem shows that we can reduce the interval of uncertainty through function evaluations.
- There are a number of line search methods based on this idea.
  - Uniform search
  - Dichotomous search
  - Golden section
  - Fibonacci search
- These methods differ essentially in how they choose the points at which to evaluate the function.

Golden Section

- At each iteration, the new interval of uncertainty $[a_{k+1}, b_{k+1}]$ is given by either $[\lambda_k, b_k]$ or $[a_k, \mu_k]$.
- Idea: Make the length of these two intervals equal.
- Idea: Either take $\lambda_{k+1} = \mu_k$ or $\mu_{k+1} = \lambda_k$ at each iteration so that only one function evaluation is required.
- This results in a set of equations which can be solved to determine the unique iterates.
- The method reduces the interval in each iteration by .618. . . , the so-called "golden ratio."

Using Derivative Information

- Notice that without derivatives, we needed two function iterations to reduce the interval of uncertainty.
- In essence, we estimate the sign of the derivative using the difference of the two function values.
- With derivative information, we get better information.
- For convex functions, the sign of the derivative tells us in which direction the optimal solution must lie.
- Bisection Method: Evaluate the derivative at the midpoint of the interval of uncertainty.

Newton's Method

- Uses a quadratic approximation to the function Θ.
  $q(\lambda) = \Theta(\lambda_k) + \Theta'(\lambda_k)(\lambda - \lambda_k) + \frac{1}{2}\Theta''(\lambda_k)(\lambda - \lambda_k)^2$
- The next iterate is taken to be the point at which the derivative is 0.
  $\Rightarrow \Theta'(\lambda_k) + \Theta''(\lambda_k)(\lambda_{k+1} - \lambda_k) = 0$
  $\Rightarrow \lambda_{k+1} = \lambda_k - \Theta'(\lambda_k)/\Theta''(\lambda_k)$

Convergence of Newton's Method

- With previous methods, convergence was guaranteed by consistently reducing the interval of uncertainty.
- Newton's method, however, does not always converge.
- We cannot apply Theorem 7.2.3 because there is no descent function available.
- If the starting point is "close enough," then we can get convergence (see text).
- There is a quadratic fit line search method with global convergence in the text.

Convergence of Line Search Methods

Theorem 2. *Let* $f : \mathbb{R}^n \to \mathbb{R}$ *and* L*, a closed interval in* $\mathbb{R}$ *be given. Consider the following line search map*

$$M(x, d) = \{x + \lambda d : \lambda \in L \text{ and } f(y) \leq f(x + \lambda d) \; \forall \lambda \in L$$

If f *is continuous at* x *and* $d \neq 0$*, then* M *is closed at* (x, d)*.*

- Hence, as long as the direction vector d is nonzero outside the solution set Ω, Theorem 7.2.3 applies and we have convergence.

Multi-dimensional Search

- Uses line search in multiple directions.
- Cyclic Coordinate Method: At each iteration, do n successive line searches, one along each of the coordinate axes.
- If f is differentiable, convergence to a point with a zero gradient is guaranteed by Theorem 7.3.5, as long as
  - The minimum of f along any line is unique.
  - The sequence of points generated is bounded.
- If f is not differentiable, the method can stall.

Method of Hooke and Jeeves

- This method attempts to overcome the problems with coordinate search by injecting an acceleration step.
- Method
  - From x_k, perform coordinate search to derive x_{k+1}.
  - Do a line search in the direction $x_{k+1} - x_k$.
  - Iterate.
- If f is differentiable, convergence can be established using Theorem 7.3.4, under mild assumptions.

Rosenbrock's Method

- Instead of using acceleration steps, this method readjusts the set of orthogonal directions at each step.
- The readjustment is done using the Gram-Schmidt procedure.
- This method has the same convergence properties as Hooke and Jeeves, but may converge faster.
- We will look at rate of convergence later.

Using Discrete Steps

- Because of the expense of performing line search, these methods can be modified to use discrete steps.
- The step size is pre-determined, but is adjusted depending on the outcome.
- In general, the step length is made smaller as the algorithm progresses.
- Once the step length reaches a threshold value, the algorithm stops.