Lecture 8

Intro to Algorithms: Line Search Methods

8.1 Lecture Objectives

- Articulate the reasons to study one-dimensional line search methods (even if we mostly care about higher-dimensional optimization problems)
- Understand simple algorithms based on evaluation of the function without the need for derivatives
- Understand Newton's algorithm, as well as its limitations
- Understand variations and approximations to Newton's method

8.2 Why Study One-Dimensional Line Search Methods?

Two major reasons:

- They provide a good one-dimensional introduction to multi-dimensional methods.
- They are often used inside multi-dimensional optimization algorithms, within each iteration once the descent direction is decided.

8.3 Taxonomy of Line Search Methods

Methods require more complicated iterations when each iteration includes the evaluation of higher order derivatives:

- Evaluate the function $f(x)$

- Evaluate the derivative $f'(x)$
- Evaluate the second derivative $f''(x)$

However, methods typically also converge faster as they use higher order derivatives at each iteration.

8.4 Golden Section Search

The golden section search method seeks the minimizer of a one-dimensional function $f(x)$ over an interval $[a_0, b_0]$. This methods works by narrowing down the interval progressively. This narrowing down is accomplished by evaluating the function at two intermediate points between a_0 and b_0. These intermediate points, a_1 and b_1, are picked such that $a_1 - a_0 = b_0 - b_1 = \rho(b_0 - a_0)$ for some $\rho < \frac{1}{2}$. Please see figure 8.1 for an illustration of this method. The function is then evaluated at a_1 and b_1, and the interval is narrowed down by the following rule:

- If $f(a_1) < f(b_1)$, then the minimizer is assumed to lie within $[a_0, b_1]$. Therefore, we set $b_0 = b_1$ and iterate again.
- If $f(a_1) \geq f(b_1)$, then the minimizer is assumed to lie within $[a_1, b_0]$. Therefore, we set $a_0 = a_1$ and iterate again.

This way, our interval will become progressively smaller as we seek the minimizer. After a suitable number of iterations (eg: when a_0 and b_0 are sufficiently close to each other), we can stop the iterations and return the value $(a_0 + b_0)/2$ as our estimated minimizer.

Now, the question is how to pick ρ. In practice, a clever choice of ρ is to satisfy the constraint:

$$\frac{1-\rho}{1} = \frac{\rho}{1-\rho} \tag{8.1}$$

which leads to $\rho \approx 0.382$. This relationship is also known as the golden section.

Question: Why is this choice of ρ common (and clever)? Think about the number of function evaluations needed at each iteration: can we get away with one single function evaluation instead of the expected two function evaluations?

Question: What is the size of our interval $[a_0, b_0]$ after K iterations?

Note that there are variants of the golden section search method that are designed to accelerate the convergence towards the minimizer. An example of such a variant is called Fibonacci search.

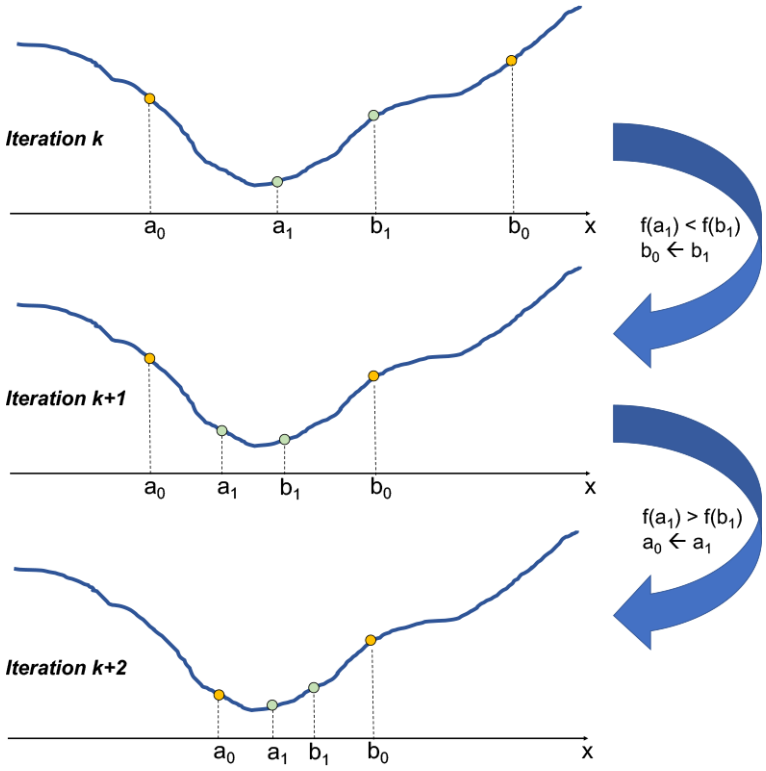

Figure 8.1: Pictorial depiction of the golden section search method.

8.5 Newton's Method

Newton's method is an iterative method based on approximating a function locally using a Taylor expansion of order two. This approximation has two desirable properties: i) it is a good (local) approximation to many smooth functions, and ii) it is easy to minimize. See figure 8.2 for an illustration of this method.

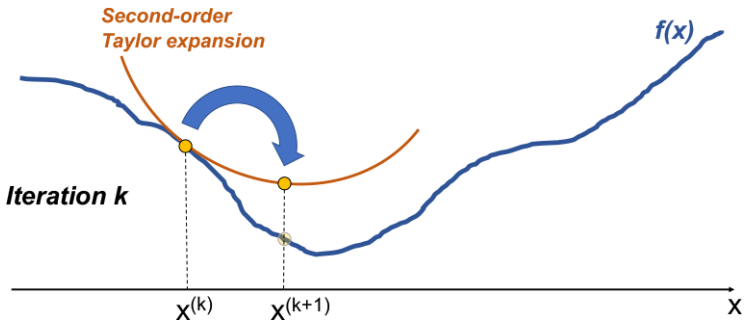

Figure 8.2: Pictorial depiction of the Newton line search method.

Suppose that at iteration k, we have an estimate $x^{(k)}$ for our minimizer of $f(x)$. Then,

we can expand $f(x)$ in the neighborhood of $x^{(k)}$ as follows:

$$f(x) \approx f(x^{(k)}) + f'(x^{(k)})(x - x^{(k)}) + \frac{1}{2} f''(x^{(k)})(x - x^{(k)})^2 \tag{8.2}$$

Then, we can update our estimate as follows:

$$x^{(k+1)} = x^{(k)} - \frac{f'(x^{(k)})}{f''(x^{(k)})} \tag{8.3}$$

Newton's method tends to work well if $f''(x) > 0$ everywhere. However, if at some iterations $f''(x^{(k)}) < 0$, Newton's method may fail to converge to a minimizer. Specifically, note that $-f'(x^{(k)})$ will point in a descent direction of $f(x)$. Therefore, if $f''(x^{(k)}) < 0$, then $-\frac{f'(x^{(k)})}{f''(x^{(k)})}$ will point in an ascent direction of $f(x)$.

8.5.1 Levenberg-Marquardt algorithm

So the question is, what can we do if $f''(x^{(k)}) < 0$? One alternative known as the Levenberg-Marquardt algorithm is to replace $f''(x^{(k)})$ in the iterations by $f''(x^{(k)}) + \mu_k$, for some $\mu_k \geq 0$:

$$x^{(k+1)} = x^{(k)} - \frac{f'(x^{(k)})}{f''(x^{(k)}) + \mu_k} \tag{8.4}$$

where we can pick μ_k to ensure descent at each iteration. The Levenberg-Marquardt algorithm will become important in its multi-dimensional version, where the first derivative is replaced by the gradient (vector), the second derivative is replaced by the Hessian (matrix), and μ_k is replaced by $\mu_k \mathbf{I}$ (where $\mathbf{I}$ is the identity matrix).

8.5.2 Main limitation

Despite their popularity, a limitation of both Newton's algorithm and the Levenberg-Marquardt variation is the need to calculate second derivatives, which are not always available or may be computationally expensive. An approximated variation of Newton's method is described next.

8.6 Secant Method

Newton's method requires, at each iteration, evaluation of the second derivative of the cost function $f(x)$. In cases where this is inconvenient or computationally challenging, we may approximate the second derivative as follows:

$$f''(x^{(k)}) \approx \frac{f'(x^{(k)}) - f'(x^{(k-1)})}{x^{(k)} - x^{(k-1)}} \tag{8.5}$$

which leads to the secant algorithm, where we update our estimate as follows:

$$x^{(k+1)} = x^{(k)} - \frac{x^{(k)} - x^{(k-1)}}{f'(x^{(k)}) - f'(x^{(k-1)})} f'(x^{(k)}) \tag{8.6}$$

Note that in the secant algorithm, each iteration relies on the previous two iterations in order to approximate the second derivative. Therefore, the secant algorithm requires two initial points. These can be two different initial guesses, or one initial guess followed by one iteration of a different algorithm (eg: one of the algorithms described above).