**Jim Lambers**
**MAT 419/519**
**Summer Session 2011-12**
**Lecture 10 Notes**

These notes correspond to Section 3.2 in the text.

## The Method of Steepest Descent

When it is not possible to find the minimium of a function analytically, and therefore must use an iterative method for obtaining an approximate solution, Newton's Method can be an effective method, but it can also be unreliable. Therefore, we now consider another approach.

Given a function $f : \mathbb{R}^n \to \mathbb{R}$ that is differentiable at $\mathbf{x}_0$, the *direction of steepest descent* is the vector $-\nabla f(\mathbf{x}_0)$. To see this, consider the function

$$\varphi(t) = f(\mathbf{x}_0 + t\mathbf{u}),$$

where $\mathbf{u}$ is a *unit* vector; that is, $\|\mathbf{u}\| = 1$. Then, by the Chain Rule,

$$\begin{aligned}
\varphi'(t) &= \frac{\partial f}{\partial x_1}\frac{\partial x_1}{\partial t} + \cdots + \frac{\partial f}{\partial x_n}\frac{\partial x_n}{\partial t} \\
&= \frac{\partial f}{\partial x_1}u_1 + \cdots + \frac{\partial f}{\partial x_n}u_n \\
&= \nabla f(\mathbf{x}_0 + t\mathbf{u}) \cdot \mathbf{u},
\end{aligned}$$

and therefore

$$\varphi'(0) = \nabla f(\mathbf{x}_0) \cdot \mathbf{u} = \|\nabla f(\mathbf{x}_0)\| \cos\theta,$$

where $\theta$ is the angle between $\nabla f(\mathbf{x}_0)$ and $\mathbf{u}$. It follows that $\varphi'(0)$ is minimized when $\theta = \pi$, which yields

$$\mathbf{u} = -\frac{\nabla f(\mathbf{x}_0)}{\|\nabla f(\mathbf{x}_0)\|}, \quad \varphi'(0) = -\|\nabla f(\mathbf{x}_0)\|.$$

We can therefore reduce the problem of minimizing a function of several variables to a single-variable minimization problem, by finding the minimum of $\varphi(t)$ for this choice of $\mathbf{u}$. That is, we find the value of $t$, for $t > 0$, that minimizes

$$\varphi_0(t) = f(\mathbf{x}_0 - t\nabla f(\mathbf{x}_0)).$$

After finding the minimizer $t_0$, we can set

$$\mathbf{x}_1 = \mathbf{x}_0 - t_0\nabla f(\mathbf{x}_0)$$

and continue the process, by searching from $\mathbf{x}_1$ in the direction of $-\nabla f(\mathbf{x}_1)$ to obtain $\mathbf{x}_2$ by minimizing $\varphi_1(t) = f(\mathbf{x}_1 - t\nabla f(\mathbf{x}_1))$, and so on.

This is the *Method of Steepest Descent*: given an initial guess $\mathbf{x}_0$, the method computes a sequence of iterates $\{\mathbf{x}_k\}$, where

$$\mathbf{x}_{k+1} = \mathbf{x}_k - t_k\nabla f(\mathbf{x}_k), \quad k = 0, 1, 2, \dots,$$

where $t_k > 0$ minimizes the function

$$\varphi_k(t) = f(\mathbf{x}_k - t\nabla f(\mathbf{x}_k)).$$

**Example** We apply the Method of Steepest Descent to the function

$$f(x, y) = 4x^2 - 4xy + 2y^2$$

with initial point $\mathbf{x}_0 = (2, 3)$. We first compute the steepest descent direction from

$$\nabla f(x, y) = (8x - 4y, 4y - 4x)$$

to obtain

$$\nabla f(\mathbf{x}_0) = \nabla f(2, 3) = (4, 4).$$

We then minimize the function

$$\varphi(t) = f((2, 3) - t(4, 4)) = f(2 - 4t, 3 - 4t)$$

by computing

$$\begin{aligned}
\varphi'(t) = -\nabla f(2 - 4t, 3 - 4t) \cdot (4, 4) & \\
&= -(8(2 - 4t) - 4(3 - 4t), 4(3 - 4t) - 4(2 - 4t)) \cdot (4, 4) \\
&= -(16 - 32t - 12 + 16t, 12 - 16t - 8 + 16t) \cdot (4, 4) \\
&= -(-16t + 4, 4) \cdot (4, 4) \\
&= 64t - 32.
\end{aligned}$$

This strictly convex function has a strict global minimum when $\varphi'(t) = 64t - 32$, or $t = 1/2$, as can be seen by noting that $\varphi''(t) = 64 > 0$. We therefore set

$$\mathbf{x}_1 = \mathbf{x}_0 - \frac{1}{2}\nabla f(\mathbf{x}_0) = (2, 3) - \frac{1}{2}(4, 4) = (0, 1).$$

Continuing the process, we have

$$\nabla f(\mathbf{x}_1) = \nabla f(0, 1) = (-4, 4),$$

and by defining

$$\varphi(t) = f((0,1) - t(-4,4)) = f(4t, 1-4t)$$

we obtain

$$\varphi'(t) = -(8(4t) - 4(1-4t), 4(1-4t) - 4(4t)) \cdot (-4,4) = -(48t-4, -32t+4) \cdot (-4,4) = 320t - 32.$$

We have $\varphi'(t) = 0$ when $t = 1/10$, and because $\varphi''(t) = 320$, this critical point is a strict global minimizer. We therefore set

$$\mathbf{x}_2 = \mathbf{x}_1 - \frac{1}{10}\nabla f(\mathbf{x}_1) = (0,1) - \frac{1}{10}(-4,4) = \left(\frac{2}{5}, \frac{3}{5}\right).$$

Repeating this process yields $\mathbf{x}_3 = (0, \frac{2}{10})$. We can see that the Method of Steepest Descent produces a sequence of iterates $\mathbf{x}_k$ that is converging to the strict global minimizer of $f(x,y)$ at $\mathbf{x}^* = (0,0)$. $\square$

The following theorems describe some important properties of the Method of Steepest Descent.

**Theorem** Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable on $\mathbb{R}^n$, and let $\mathbf{x}_0 \in D$. Let $t^* > 0$ be the minimizer of the function

$$\varphi(t) = f(\mathbf{x}_0 - t\nabla f(\mathbf{x}_0)), \quad t \geq 0$$

and let $\mathbf{x}_1 = \mathbf{x}_0 - t^*\nabla f(\mathbf{x}_0)$. Then

$$f(\mathbf{x}_1) < f(\mathbf{x}_0).$$

That is, the Method of Steepest Descent is guaranteed to make at least some progress toward a minimizer $\mathbf{x}^*$ during each iteration. This theorem can be proven by showing that $\varphi'(0) < 0$, which guarantees the existence of $\bar{t} > 0$ such that $\varphi(t) < \varphi(0)$.

**Theorem** Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable on $\mathbb{R}^n$, and let $\mathbf{x}_k$ and $\mathbf{x}_{k+1}$, for $k \geq 0$, be two consecutive iterates produced by the Method of Steepest Descent. Then the steepest descent directions from $\mathbf{x}_k$ and $\mathbf{x}_{k+1}$ are orthogonal; that is,

$$\nabla f(\mathbf{x}_k) \cdot \nabla f(\mathbf{x}_{k+1}) = 0.$$

This theorem can be proven by noting that $\mathbf{x}_{k+1}$ is obtained by finding a critical point $t^*$ of $\varphi(t) = f(\mathbf{x}_k - t\nabla f(\mathbf{x}_k))$, and therefore

$$\varphi'(t^*) = -\nabla f(\mathbf{x}_{k+1}) \cdot f(\mathbf{x}_k) = 0.$$

That is, the Method of Steepest Descent pursues completely independent search directions from one iteration to the next. However, in some cases this causes the method to "zig-zag" from the initial iterate $\mathbf{x}_0$ to the minimizer $\mathbf{x}^*$.

We have seen that Newton's Method can fail to converge to a solution if the initial iterate is not chosen wisely. For certain functions, however, the Method of Steepest Descent can be shown to be much more reliable.

**Theorem** Let $f : \mathbb{R}^n \to \mathbb{R}$ be a coercive function with continuous first partial derivatives on $\mathbb{R}^n$. Then, for any initial guess $\mathbf{x}_0$, the sequence of iterates produced by the Method of Steepest Descent from $\mathbf{x}_0$ contains a subsequence that converges to a critical point of $f$.

This result can be proved by applying the *Bolzano-Weierstrauss Theorem*, which states that any bounded sequence contains a convergent subsequence. The sequence $\{f(\mathbf{x}_k)\}_{k=0}^{\infty}$ is a decreasing sequence, as indicated by a previous theorem, and it is a bounded sequence, because $f(\mathbf{x})$ is continuous and coercive and therefore has a global minimum $f(\mathbf{x}^*)$. It follows that the sequence $\{\mathbf{x}_k\}$ is also bounded, for a coercive function cannot be bounded on an unbounded set.

By the Bolzano-Weierstrauss Theorem, $\{\mathbf{x}_k\}$ has a convergent subsequence $\{\mathbf{x}_{k_p}\}$, which can be shown to converge to a critical point of $f(\mathbf{x})$. Intuitively, as $\mathbf{x}_{k+1} = \mathbf{x}_k - t^*\nabla f(\mathbf{x}_k)$ for some $t^* > 0$, convergence of $\{\mathbf{x}_{k_p}\}$ implies that

$$0 = \lim_{p\to\infty} \mathbf{x}_{k_{p+1}} - \mathbf{x}_{k_p} = -\sum_{i=k_p}^{k_{p+1}-1} t_i^*\nabla f(\mathbf{x}_i), \quad t_i^* > 0,$$

which suggests the convergence of $\nabla f(\mathbf{x}_{k_p})$ to zero.

If $f(\mathbf{x})$ is also strictly convex, we obtain the following stronger result about the reliability of the Method of Steepest Descent.

**Theorem** Let $f : \mathbb{R}^n \to \mathbb{R}$ be a coercive, strictly convex function with continuous first partial derivatives on $\mathbb{R}^n$. Then, for any initial guess $\mathbf{x}_0$, the sequence of iterates produced by the Method of Steepest Descent from $\mathbf{x}_0$ converges to the unique global minimizer $\mathbf{x}^*$ of $f(\mathbf{x})$ on $\mathbb{R}^n$.

This theorem can be proved by noting that if the sequence $\{\mathbf{x}_k\}$ of steepest descent iterates does not converge to $\mathbf{x}^*$, then any subsequence that does not converge to $\mathbf{x}^*$ must contain a subsequence that converges to a critical point, by the previous theorem, but $f(\mathbf{x})$ has only one critical point, which is $\mathbf{x}^*$, which yields a contradiction.

## Exercises

1. Chapter 3, Exercise 8
2. Chapter 3, Exercise 11
3. Chapter 3, Exercise 12