Jim Lambers
MAT 280
Spring Semester 2009-10
Lecture 7 Notes

These notes correspond to Section 11.6 in Stewart and Section 2.6 in Marsden and Tromba.

Directional Derivatives and the Gradient Vector

Previously, we defined the gradient as the vector of all of the first partial derivatives of a scalar-valued function of several variables. Now, we will learn about how to use the gradient to measure the rate of change of the function with respect to a change of its variables in *any* direction, as opposed to a change in a single variable. This is extremely useful in applications in which the minimum or maxmium value of a function is sought. We will also learn how the gradient can be used to easily describe tangent planes to level surfaces, thus providing an alternative to implicit differentiation or the Chain Rule.

The Gradient Vector

Let $f : D \subseteq \mathbb{R}^n \to \mathbb{R}$ be a scalar-valued function of n variables $x_1, x_2, \ldots, x_n$. Recall that the vector of its first partial derivatives,

$$\nabla f = \begin{bmatrix} f_{x_1} & f_{x_2} & \cdots & f_{x_n} \end{bmatrix},$$

is called the *gradient* of f.

Example Let $f(x, y, z) = e^{-(x^2+y^2)} \cos z$. Then

$$\nabla f = \begin{bmatrix} -2xe^{-(x^2+y^2)} \cos z & -2ye^{-(x^2+y^2)} \cos z & -e^{-(x^2+y^2)} \sin z \end{bmatrix}.$$

Therefore, at the point $(x_0, y_0, z_0) = (1, 2, \pi/3)$, the gradient is the vector

$$\nabla f(x_0, y_0, z_0) = \begin{bmatrix} f_x(1, 2, \pi/3) & f_y(1, 2, \pi/3) & f_z(1, 2, \pi/3) \end{bmatrix} = \left\langle -e^{-5}, -2e^{-5}, -\frac{\sqrt{3}}{2} e^{-5} \right\rangle.$$

□

It should be noted that various differentiation rules from single-variable calculus have direct generalizations to the gradient. Let u and v be differentiable functions defined on $\mathbb{R}^n$. Then, we have:

- *Linearity:*

$$\nabla(au + bv) = a\nabla u + b\nabla v$$

  where a and b are constants

- *Product Rule:*
$$\nabla(uv) = u\nabla v + v\nabla u$$
- *Quotient Rule:*
$$\nabla\left(\frac{u}{v}\right) = \frac{v\nabla u - u\nabla v}{v^2}$$
- *Power Rule:*
$$\nabla u^n = nu^{n-1}\nabla u$$

Directional Derivatives

The components of the gradient vector ∇f represent the instantaneous rates of change of the function f with respect to any *one* of its independent variables. However, in many applications, it is useful to know how f changes as its variables change along *any* path from a given point. To that end, given $f : D \subseteq \mathbb{R}^2 \to \mathbb{R}$, and a *unit* vector $\mathbf{u} = \langle a, b\rangle \in \mathbb{R}^2$, we define the *directional derivative* of f at $(x_0, y_0) \in D$ in the direction of $\mathbf{u}$ to be

$$D_{\mathbf{u}}f(x_0, y_0) = \lim_{h\to 0} \frac{f(x_0 + ah, y_0 + bh) - f(x_0, y_0)}{h}.$$

When $\mathbf{u} = \mathbf{i} = \langle 1, 0\rangle$, then $D_{\mathbf{u}}f = f_x$, and when $\mathbf{u} = \mathbf{j} = \langle 0, 1\rangle$, then $D_{\mathbf{u}}f = f_y$. For general $\mathbf{u}$, $D_{\mathbf{u}}f(x_0, y_0)$ represents the instantaneous rate of change of f as (x, y) change in the direction of $\mathbf{u}$ from the point (x_0, y_0).

Because it is cumbersome to compute a directional derivative using the definition directly, it is desirable to be able to relate the directional derivative to the partial derivatives, which can be computed easily using differentiation rules. We have

$$\begin{aligned}
D_{\mathbf{u}}f(x_0, y_0) &= \lim_{h\to 0} \frac{f(x_0 + ah, y_0 + bh) - f(x_0, y_0)}{h} \\
&= \lim_{h\to 0} \frac{f(x_0 + ah, y_0 + bh) - f(x_0, y_0 + bh) + f(x_0, y_0 + bh) - f(x_0, y_0)}{h} \\
&= \lim_{h\to 0} \frac{f(x_0 + ah, y_0 + bh) - f(x_0, y_0 + bh)}{h} + \frac{f(x_0, y_0 + bh) - f(x_0, y_0)}{h} \\
&= \lim_{h\to 0} \frac{f(x_0 + ah, y_0 + bh) - f(x_0, y_0 + bh)}{ah}a + \frac{f(x_0, y_0 + bh) - f(x_0, y_0)}{bh}b \\
&= f_x(x_0, y_0)a + f_y(x_0, y_0)b \\
&= \nabla f(x_0, y_0) \cdot \mathbf{u}.
\end{aligned}$$

That is, the directional derivative in the direction of $\mathbf{u}$ is the *dot product* of the gradient with $\mathbf{u}$. It can be shown that this is the case for any number of variables: given $f : D \subseteq \mathbb{R}^n \to \mathbb{R}$, and a unit vector $\mathbf{u} \in \mathbb{R}^n$, the directional derivative of f at $\mathbf{x}_0 \in \mathbb{R}^n$ in the direction of $\mathbf{u}$ is given by

$$D_{\mathbf{u}}f(\mathbf{x}_0) = \nabla f(\mathbf{x}_0) \cdot \mathbf{u}.$$

Because the dot product $\mathbf{a} \cdot \mathbf{b}$ can also be defined as

$$\mathbf{a} \cdot \mathbf{b} = \|\mathbf{a}\|\|\mathbf{b}\| \cos\theta,$$

where θ is the angle between $\mathbf{a}$ and $\mathbf{b}$, the directional derivative can be used to determine the direction along which f increases most rapidly, decreases most rapidly, or does not change at all.

We first note that if θ is the angle between $\nabla f(\mathbf{x}_0)$ and $\mathbf{u}$, then

$$D_{\mathbf{u}} f(\mathbf{x}_0) = \nabla f(\mathbf{x}_0) \cdot \mathbf{u} = \|\nabla f(\mathbf{x}_0)\| \cos\theta.$$

Then we have the following:

- When $\theta = 0$, $\cos\theta = 1$, so $D_{\mathbf{u}} f$ is maximized, and its value is $\|\nabla f(\mathbf{x}_0)\|$. In this case,

  $$\mathbf{u} = \frac{\nabla f(\mathbf{x}_0)}{\|\nabla f(\mathbf{x}_0)\|},$$

  and this is called the *direction of steepest ascent*.

- When $\theta = \pi$, $\cos\theta = -1$, so $D_{\mathbf{u}} f$ is minimized, and its value is $-\|\nabla f(\mathbf{x}_0)\|$. In this case,

  $$\mathbf{u} = -\frac{\nabla f(\mathbf{x}_0)}{\|\nabla f(\mathbf{x}_0)\|},$$

  and this is called the *direction of steepest descent*.

- When $\theta = \pm\pi/2$, $\cos\theta = 0$, so $D_{\mathbf{u}} = 0$. In this case, $\mathbf{u}$ is a unit vector that is *orthogonal* (perpendicular) to $\nabla f(\mathbf{x}_0)$. Since f is not changing at all along this direction, it follows that $\mathbf{u}$ indicates the direction of a *level set* of f, on which $f(\mathbf{x}_0) = f(\mathbf{x}_0)$.

The direction of steepest descent is of particular interest in applications in which the goal is to find the minimum value of f. From a starting point $\mathbf{x}_0$, one can choose a new point $\mathbf{x}_1 = \mathbf{x}_0 + \alpha\mathbf{u}$, where $\mathbf{u} = -\nabla f(\mathbf{x}_0)$ is the direction of steepest descent, by choosing α so as to minimize $f(\mathbf{x}_1)$. Then, this process can be repeated using the direction of steepest descent at $\mathbf{x}_1$, which is $-\nabla f(\mathbf{x}_1)$, to compute a new point $\mathbf{x}_2$, and so on, until a minimum is found. This process is called the *method of steepest descent*.

While not used very often in practice, it serves as a useful building block for some of the most powerful methods that are used in practice for minimizing functions.

Example Let $f(x, y) = x^2 y + y^3$, and let $(x_0, y_0) = (2, -2)$. Then

$$\nabla f(x, y) = \begin{bmatrix} f_x(x, y) & f_y(x, y) \end{bmatrix} = \begin{bmatrix} 2xy & x^2 + 3y^2 \end{bmatrix},$$

which yields $\nabla f(x_0,y_0) = \langle f_x(2,-2), f_y(2,-2)\rangle = \langle -8,16\rangle$. It follows that the direction of steepest ascent is

$$\mathbf{u} = \frac{\nabla f(2,-2)}{\|\nabla f(2,-2)\|} = \frac{\langle -8,16\rangle}{\sqrt{(-8)^2+16^2}} = \frac{\langle -8,16\rangle}{\sqrt{320}} = \frac{\langle -8,16\rangle}{8\sqrt{5}} = \left\langle -\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}}\right\rangle.$$

For this $\mathbf{u}$, we have $D_{\mathbf{u}}f(2,-2) = \|\nabla f(2,-2)\| = 8\sqrt{5}$.

Furthermore, the direction of steepest descent is

$$\mathbf{u} = \left\langle \frac{1}{\sqrt{5}}, -\frac{2}{\sqrt{5}}\right\rangle,$$

and along this direction, we have $D_{\mathbf{u}}f(2,-2) = -\|\nabla f(2,-2)\| = -8\sqrt{5}$. Finally, the directions along which f does not change at all are those that are orthogonal to the directions of steepest ascent and descent,

$$\mathbf{u} = \pm\left\langle \frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}}\right\rangle.$$

The level curve defined by the equation $f(x,y) = f(2,-2) = -16$ proceeds along these directions from the point $(2,-2)$. □

Tangent Planes to Level Surfaces

Let $F : D \subseteq \mathbb{R}^3 \to \mathbb{R}$ be a function of three variables x, y and z that implicitly defines a surface through the equation $F(x,y,z) = 0$, and let (x_0,y_0,z_0) be a point on that surface. If F satisfies the conditions of the Implicit Function Theorem at (x_0,y_0,z_0), then the equation of the plane that is tangent to the surface at this point can be obtained using the fact that z is implicitly defined as a function of x and y near this point. It then follows that the equation of the tangent plane is

$$z - z_0 = z_x(x_0,y_0)(x-x_0) + z_y(x_0,y_0)(y-y_0),$$

where, by the Chain Rule,

$$z_x(x_0,y_0) = -\frac{F_x(x_0,y_0,z_0)}{F_z(x_0,y_0,z_0)}, \quad z_y(x_0,y_0) = -\frac{F_y(x_0,y_0,z_0)}{F_z(x_0,y_0,z_0)}.$$

This is not possible if $F_z(x_0,y_0,z_0) = 0$, because then the Implicit Function Theorem does not apply.

It would be desirable to be able to obtain the equation of the tangent plane even if $F_z(x_0,y_0,z_0) = 0$, because the level surface still has a tangent plane at that point even if z cannot be implicitly defined as a function of x and y. To that end, we note that any direction $\mathbf{u}$ within the tangent plane is parallel to the tangent vector of some curve that lies within the surface and passes through

(x_0, y_0, z_0). Because $F(x, y, z) = 0$ on this surface, it follows that $D_{\mathbf{u}}F(x_0, y_0, z_0) = 0$. However, this implies that $\nabla F(x_0, y_0, z_0)$ must be orthogonal to $\mathbf{u}$, in view of

$$D_{\mathbf{u}}F(x_0, y_0, z_0) = \nabla F(x_0, y_0, z_0) \cdot \mathbf{u} = 0.$$

Since this is the case for *any* direction $\mathbf{u}$ within the tangent plane, we conclude that $\nabla F(x_0, y_0, z_0)$ is *normal* to the tangent plane, and therefore the equation of this plane is

$$F_x(x_0, y_0, z_0)(x - x_0) + F_y(x_0, y_0, z_0)(y - y_0) + F_z(x_0, y_0, z_0)(z - z_0) = 0.$$

Note that this equation is equivalent to that obtained using the Chain Rule, when $F_z(x_0, y_0, z_0) \neq 0$.

The gradient not only provides the normal vector to the tangent plane, but also the direction numbers of the *normal line* to the surface at (x_0, y_0, z_0), which is the line that passes through the surface at this point and is perpendicular to the tangent plane. The equation of this line, in parametric form, is

$$x = x_0 + tF_x(x_0, y_0, z_0), \quad y = y_0 + tF_y(x_0, y_0, z_0), \quad z = z_0 + tF_z(x_0, y_0, z_0).$$

Example Let $F(x, y, z) = x^2 + y^2 + z^2 - 2x - 4y - 4$. Then the equation $F(x, y, z) = 0$ defines a sphere of radius 3 centered at $(1, 2, 0)$. At the point $(x_0, y_0, z_0) = (3, 3, 2)$, we have

$$\begin{aligned} \nabla F(x_0, y_0, z_0) &= \begin{bmatrix} F_x(x_0, y_0, z_0) & F_y(x_0, y_0, z_0) & F_z(x_0, y_0, z_0) \end{bmatrix} \\ &= \begin{bmatrix} 2x_0 - 2 & 2y_0 - 4 & 2z_0 \end{bmatrix} \\ &= \langle 4, 2, 4 \rangle. \end{aligned}$$

It follows that the equation of the plane that is tangent to the sphere at $(3, 3, 2)$ is

$$4(x - x_0) + 2(y - y_0) + 4(z - z_0) = 0,$$

and the equation of the normal line, in parametric form, is

$$x = x_0 + tF_x(x_0, y_0, z_0) = 3 + 4t, \quad y = y_0 + tF_y(x_0, y_0, z_0) = 3 + 2t, \quad z = z_0 + tF_z(x_0, y_0, z_0) = 2 + 4t.$$

Equivalently, we can describe the normal line using its symmetric equations,

$$\frac{x - 3}{4} = \frac{y - 3}{2} = \frac{z - 2}{4}.$$

□

Practice Problems

1. Compute the directional derivatives of each function at the indicated point, in the direction of the given vector.

   (a) $f(x, y, z) = x^2yz^3 + x^3y^2z$, $(x_0, y_0, z_0) = (1, -1, 2)$, $\mathbf{u} = \langle 3/\sqrt{50}, 4/\sqrt{50}, 5/\sqrt{50} \rangle$

   (b) $f(x, y) = 4x^2 + 9y^2$, $(x_0, y_0) = (3, 2)$, $\mathbf{u}$ is the unit vector in the xy-plane that makes the angle $\theta = \pi/6$ with the positive x-axis

2. For each of the following functions, compute the direction along which the function increases most rapidly from the given point.

   (a) $f(x, y) = x^2 + y^2$, $(x_0, y_0) = (1, 4)$

   (b) $f(x, y, z) = e^z \cos x \sin y$, $(x_0, y_0, z_0) = (\pi/6, \pi/6, 1)$

3. For each of the following implicitly defined surfaces, compute the equations of the tangent plane and normal line at the indicated point.

   (a) $F(x, y, z) = x^2 + y^2 + z^2 - 9 = 0$, $(x_0, y_0, z_0) = (2, -1, 2)$

   (b) $F(x, y, z) = x^2 + y^2 + z^2 + 16 - 8\sqrt{x^2 + y^2} - s^2 = 0$, $(x_0, y_0, z_0) = (1, 4, \sqrt{32 - 8\sqrt{17}})$

Additional Practice Problems

Additional practice problems from the recommended textbooks are:

- Stewart: Section 11.6, Exercises 1-17 odd, 31, 33
- Marsden/Tromba: Section 2.6, Exercises 1, 3, 5, 9