The Chain Rule and The Gradient

Department of Mathematics and Statistics

October 31, 2012

The Chain Rule

Their are various versions of the chain rule for multivariable functions. For example,

Theorem (Version I)

Given functions $z = f(x, y)$, $x = u(t)$ *and* $y = v(t)$, *for all values of* t *at which* u *and* v *are differentiable and* f *is differentiable at* $(u(t), v(t))$, *we have*

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}.$$

Theorem (Version II)

Given functions $z = f(x, y)$, $x = u(s, t)$ *and* $y = v(s, t)$, *for all values of* s *and* t *at which* u *and* v *are differentiable and* f *is differentiable at* $(u(s, t), v(s, t))$, *we have*

$$\frac{\partial z}{\partial s} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial s} \text{ and } \frac{\partial z}{\partial t} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial t}.$$

The Gradient

Definition

Let $z = f(x, y)$ be a function of two variables, the **gradient** of f is the vector function defined by

$$\nabla f(x,y) = \frac{\partial f}{\partial x}\mathbf{i} + \frac{\partial f}{\partial y}\mathbf{j} = \langle f_x(x,y), f_y(x,y)\rangle.$$

Similarly, if $w = f(x, y, z)$ is a function of three variables the **gradient** of f is the vector function defined by

$$\nabla f(x,y,z) = \frac{\partial f}{\partial x}\mathbf{i} + \frac{\partial f}{\partial y}\mathbf{j} + \frac{\partial f}{\partial z}\mathbf{k} = \langle f_x(x,y,z), f_y(x,y,z), f_z(x,y,z)\rangle.$$

The domain of the gradient is the set of all points in the domain of f at which the partial derivatives exist.

The symbol ∇f is read "the gradient of f", "grad f" or "del f".

Uses of the Gradient

Our primary uses for the gradient will be:

- Locating extrema – places where $\nabla f(x, y) = \mathbf{0}$ or where $\nabla f(x, y)$ doesn't exist are candidates for the location of extrema.
- Providing a shortcut for computing the directional derivative.
- Finding the direction of most rapid increase and decrease of the function.

Computing the Directional Derivative

Theorem

Let $f(x, y)$ be a function of two variables and (x_0, y_0) be a point in the domain of f at which the first-order partial derivatives of f exist. If $\mathbf{u} \in \mathbb{R}^2$ is a unit vector for which the directional derivative $D_{\mathbf{u}}f(x_0, y_0)$ also exists, then

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

Similarly, if $f(x, y, z)$ is a function of three variables and (x_0, y_0, z_0) is a point in the domain of f at which the first-order partial derivatives of f exist and $\mathbf{u} \in \mathbb{R}^3$ is a unit vector for which the directional derivative $D_{\mathbf{u}}f(x_0, y_0, z_0)$ also exists, then

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

The Geometry of the Gradient

Theorem (The Gradient Points in the Direction of Greatest Increase)

Let f be a function of two or three variables and let P be a point in the domain of f at which f is differentiable. The gradient of f at P points in the direction in which f increases most rapidly.

Theorem (Gradient Vectors are Orthogonal to Level Curves)

Let f be a function of two variables and let (x_0, y_0) be a point in the domain of f at which f is differentiable. If $\mathcal{C}$ is the level curve containing the point $c_0 = f(x_0, y_0)$, then $\nabla f(x_0, y_0)$ and $\mathcal{C}$ are orthogonal at $f(x_0, y_0)$.