Econ 205 - Slides from Lecture 9

Joel Sobel

September 2, 2010

Representing Functions

- Graph: $Gr(f) = \{(x, y) : y = f(x)\}$.
  If $f : \mathbb{R}^n \longrightarrow \mathbb{R}$, then the graph is a subset of $\mathbb{R}^{n+1}$.
  If $n = 2$, then the graph is a subset of $\mathbb{R}^3$, so someone with a good imagination of a three-dimensional drawing surface could visualize it.
  If $n > 2$ there is no hope. You can get some intuition by looking at "slices" of the graph obtained by holding the function's value constant.
- Level Set:

Definition

A *level set* is the set of points such that the functions achieves the same value. Formally it is defined as the set

$$\{\mathbf{x} \in \mathbb{R}^n \mid f(\mathbf{x}) = c\} \quad \text{for any } c \in \mathbb{R}$$

While the graph of the function is a subset of $\mathbb{R}^{n+1}$, the level sets are subsets of $\mathbb{R}^n$.

Definition

We define the *upper contour set* of f as the set

$$\{\mathbf{x} \in \mathbb{R}^{\mathbf{n}} \mid \mathbf{f}(\mathbf{x}) \geq \mathbf{c}\} \quad \mathbf{c} \in \mathbb{R}$$

And we define the *lower contour set* of f as the set

$$\{\mathbf{x} \in \mathbb{R}^{\mathbf{n}} \mid \mathbf{f}(\mathbf{x}) \leq \mathbf{c}\} \quad \mathbf{c} \in \mathbb{R}$$

Definition (Limit of a Function)

$$f : \mathbb{R}^n \longrightarrow \mathbb{R}$$

$$\lim_{\mathbf{x} \longrightarrow \mathbf{a}} f(\mathbf{x}) = c \in \mathbb{R}$$

$\Longleftrightarrow \forall\, \varepsilon > 0, \exists\, \delta > 0$ such that

$$0 < d(\mathbf{x}, \mathbf{a}) < \delta$$

$$\Longrightarrow |f(\mathbf{x}) - c| < \varepsilon$$

This definition agrees with the earlier definition, although there are two twists. First, a general "distance function" replaces absolute values in the condition that says that x is close to a. For our purposes, the distance function will always be the standard Euclidean distance. Second, there we do not define one-sided continuity.

Definition (Continuity of a Function)

$$f : \mathbb{R}^n \longrightarrow \mathbb{R}$$

is called continuous at a point $\mathbf{a} \in \mathbb{R}^n$ if

$$\lim_{\mathbf{x} \longrightarrow \mathbf{a}} f(\mathbf{x}) = f(\mathbf{a})$$

Again, this definition is a simple generalization of the one-variable definition.
One can also generalize the sequential definition of continuity.

Partial Derivatives and Directional Derivatives

Definition

Take $f : \mathbb{R}^n \longrightarrow \mathbb{R}$. The *ith partial derivative* of f at $\mathbf{x}$ is defined as

$$\frac{\partial f}{\partial x_i}(\mathbf{x}) \quad \equiv \quad \lim_{h \longrightarrow 0} \frac{f(\mathbf{x} + h\mathbf{e}_i) - f(\mathbf{x})}{h}.$$

Treat every other x_j as a constant and take the derivative as though f were a function of just x_i.

As in the one variable case, partial derivatives need not exist. If the ith partial derivative exists, then the function (when viewed as a function of x_i alone) must be continuous.

Intuitions

1. Real functions on $\mathbb{R}^n$ are many functions of one variable.
2. Hold all but one variable fixed and study the function defined in only one "direction."

Directional Derivative

Definition

Take $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ and let $\mathbf{v}$ be a unit vector in $\mathbb{R}^n$. The *directional derivative of f in the direction* $\mathbf{v}$ *at* $\mathbf{x}$ is defined as

$$D_{\mathbf{v}}(\mathbf{x}) \quad \equiv \quad \lim_{h \longrightarrow 0} \frac{f(\mathbf{x} + h\mathbf{v}) - f(\mathbf{x})}{h}.$$

It follows from the definition, that

$$\frac{\partial f}{\partial x_i}(\mathbf{x}) \quad \equiv \quad D_{\mathbf{e}_i}(\mathbf{x}).$$

That is, the ith partial derivative is just a directional derivative in the direction $\mathbf{e}_i$.

A directional derivative is just the derivative of a function of one variable. The one-variable function is the function of h of the form $f(\mathbf{x} + h\mathbf{v})$ with $\mathbf{x}$ and $\mathbf{v}$ fixed.

Differentiability

Definition

A function $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is *differentiable* at $\mathbf{a} \in \mathbb{R}^n$ if and only if there is a linear function $L : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ such that

$$\lim_{\mathbf{x} \longrightarrow \mathbf{a}} \frac{||\overset{m\times 1}{f(\mathbf{x})} - \overset{m\times 1}{f(\mathbf{a})} - L(\overset{n\times 1}{\mathbf{x} - \mathbf{a}})||}{||\underset{n\times 1}{\mathbf{x} - \mathbf{a}}||} = 0.$$

If L exists we call it the derivative of f at $\mathbf{a}$ and denote it by $Df(\mathbf{a})$. In the case $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ this is equivalent to

$$\lim_{\mathbf{x} \longrightarrow \mathbf{a}} \left[\frac{f(\mathbf{x}) - f(\mathbf{a})}{||\mathbf{x} - \mathbf{a}||} - \frac{L(\mathbf{x} - \mathbf{a})}{||\mathbf{x} - \mathbf{a}||} \right] = 0$$

This implies that if f is differentiable, then for any directions defined by $\mathbf{y} \in \mathbb{R}^n$ and a magnitude given by $\alpha \in \mathbb{R}$

$$\begin{aligned}\lim_{\alpha \longrightarrow 0} \frac{f(\mathbf{a} + \alpha \mathbf{y}) - f(\mathbf{a})}{|\alpha| ||\mathbf{y}||} &= \lim_{\alpha \longrightarrow 0} \frac{Df(\mathbf{a})\alpha \mathbf{y}}{|\alpha| ||\mathbf{y}||} \\ &= \frac{Df(\mathbf{a})\mathbf{y}}{||\mathbf{y}||}\end{aligned}$$

That is, if a function in differentiable at a point, then all directional derivatives exist at the point.

Differentiability and Directional Derivatives

Theorem

If $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is differentiable then all of its directional derivatives exist. The directional derivative of f at $\mathbf{a} \in \mathbb{R}^n$ in direction $\mathbf{v} \in \mathbb{R}^n$ is given by

$$Df(\mathbf{a})\mathbf{v}$$

We assume that the direction $\mathbf{v}$ in the definition of directional derivative is of unit length.

1. If a function is differentiable, then the matrix representation of the derivative is just the matrix of partial derivatives.
2. A directional derivative is just a weighted average of partial derivatives.

Comments

1. Definition of differentiability generalizes the one-variable definition.
2. The derivative is the best linear approximation to the function.
3. Euclidean distance replaces absolute values.
4. Objects inside the $|| \cdot ||$ are vectors and not scalars. Otherwise the ratios would not make sense.
5. The linear function is more complicated than in the one variable case (but has the same domain and range as f).
6. The derivative is a linear function from $\mathbb{R}^n$ into $\mathbb{R}^m$.
7. The derivative is represented by matrix multiplication of a matrix with m rows and n columns. $m \times n$ numbers in matrix: the partial derivatives.

Notation

Definition

Given

$$f : \mathbb{R}^n \longrightarrow \mathbb{R}$$

the *gradient* of f at $\mathbf{x} \in \mathbb{R}^n$ is

$$\nabla f(\mathbf{x}) = \left(\frac{\partial f}{\partial x_1}(\mathbf{x}), \frac{\partial f}{\partial x_2}(\mathbf{x}), \ldots, \frac{\partial f}{\partial x_n}(\mathbf{x}) \right)$$

The last theorem states that

$$D_{\mathbf{v}} f(a) = \nabla f(a) \cdot \mathbf{v}. \tag{1}$$

Relationship between partial derivatives and differentiability

If a function is differentiable, then has partial derivatives and that these partial derivatives tell you everything you need to know about the derivative.

What about the converse? That is, if a function has partial derivatives, then do we know it is differentiable? The answer is "not quite."

Theorem

Take $f : \mathbb{R}^n \longrightarrow \mathbb{R}$, if f's partial derivatives exist and are continuous ($f \in C^1$), then f is differentiable.

If a function has partial derivatives in all directions, then the function is differentiable **provided that these partial derivatives are continuous.**

$$f(\mathbf{x}) = x_1 x_2$$

$$\begin{aligned} \Longrightarrow D_1(0,0) &= x_2 \mid_{(x_1=0, x_2=0)} \\ &= 0 \end{aligned}$$

$$\begin{aligned} \Longrightarrow D_2(0,0) &= x_1 \mid_{(x_1=0, x_2=0)} \\ &= 0 \end{aligned}$$

$$\lim_{h \longrightarrow 0} \frac{f((0,0) + h(1,1)) - f(0,0)}{h\sqrt{2}}$$

from $(0,0)$ the distance traveled to the point $(1,1)$ is $\sqrt{2}$, so that is why we normalize and divide in the denominator by $\sqrt{2}$.

$$
\begin{aligned}
&= \lim_{h \longrightarrow 0} \frac{f(h,h) - f(0,0)}{h\sqrt{2}} \\
&= \lim_{h \longrightarrow 0} \frac{h^2}{h\sqrt{2}} \\
&= 0
\end{aligned}
$$

This says that $D_v(0,0) = 0$ when $v = (1,1)/\sqrt{2}$. In fact we know that $D_v(0,0) = 0$ for al v.

$$f(\mathbf{x}) = |x_1|^{\frac{1}{2}}|x_2|^{\frac{1}{2}}$$

So

$$\begin{aligned}
\lim_{h \longrightarrow 0} \frac{f((0,0) + h(1,1)) - f(0,0)}{h\sqrt{2}} &= \lim_{h \longrightarrow 0} \frac{f((h,h)) - f(0,0)}{h\sqrt{2}} \\
&= \lim_{h \longrightarrow 0} \frac{|h|^{\frac{1}{2}}|h|^{\frac{1}{2}}}{h\sqrt{2}} \\
&= \lim_{h \longrightarrow 0} \frac{|h|}{h\sqrt{2}}
\end{aligned}$$

The limit does not exist (one sided limits exist, but they are not equal).
Both partial derivatives of the function at $(0,0)$ exist and are equal to zero.
The formula for computing the direction directive as the average of partials fails because f is not differentiable at $(0,0)$.

Facts

Question: What is the direction from x that most increases the value of f?
Answer: It's the direction given by the gradient.

Theorem

If $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is differentiable at x, then the direction v that maximizes $|D_v f(x)|$ is

$$\mathbf{v} = \nabla f(\mathbf{x})$$

This result follows because

$$|D_v f(x)| = |v \cdot \nabla f(\mathbf{x})| \leq ||v|| \times ||\nabla f(\mathbf{x})||$$

and the last inequality is an equation when $\mathbf{v} = \nabla f(\mathbf{x})$.

Properties of the Derivative

Theorem

$g, f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ both differentiable at $\mathbf{a} \in \mathbb{R}^n$

Then

1. $$D[cf](\mathbf{a}) = cDf(\mathbf{a}) \forall c \in \mathbb{R}$$

2. $$D[f+g](\mathbf{a}) = Df(\mathbf{a}) + Dg(\mathbf{a})$$

For the case m=1

3. $$D[\underset{1\times n}{g \cdot f}](\mathbf{a}) = \underset{1\times 1}{g(\mathbf{a})} \cdot \underset{1\times n}{Df(\mathbf{a})} + \underset{1\times 1}{f(\mathbf{a})} \cdot \underset{1\times n}{Dg(\mathbf{a})}$$

4. $$D\left[\frac{f}{g}\right](\mathbf{a}) = \frac{g(\mathbf{a}) \cdot Df(\mathbf{a}) - f(\mathbf{a}) \cdot Dg(\mathbf{a})}{[g(\mathbf{a})]^2}$$

Chain Rule

Theorem (Chain Rule Theorem)

Suppose $U \subset \mathbb{R}^n$ and $V \subset \mathbb{R}^m$ are open sets and suppose $g \colon U \longrightarrow \mathbb{R}^m$ is differentiable at $\mathbf{x} \in U$ and $f : V \longrightarrow \mathbb{R}^l$ is differentiable at $\mathbf{y} \equiv g(\mathbf{x}) \in V$. Then $f \circ g$ is differentiable at $\mathbf{x}$ and

$$\underbrace{D[f \circ g](\mathbf{x})}_{l \times n} = \underset{l \times m}{Df(\mathbf{y})} \underset{m \times n}{Dg(\mathbf{x})}$$

Chain Rule Special Case in Barbaric Notation

$f : \mathbb{R}^m \to \mathbb{R}$ and
$g : \mathbb{R} \to \mathbb{R}^m$

$$\begin{aligned} D[f \circ g](t) &= \frac{\partial y}{\partial t} \\ &= D(f(g(t))Dg(t) \\ &= \left(\frac{\partial f}{\partial x_1}(g(t)), \ldots, \frac{\partial f}{\partial x_m}(g(t)) \right) \cdot \left(\begin{array}{c} \frac{dg_1}{dt} \\ \vdots \\ \frac{dg_m}{dt} \end{array} \right) \\ &= \sum_{i=1}^{m} \frac{\partial y}{\partial x_i} \cdot \frac{dg_i}{dt} \end{aligned}$$

$$g(x) = x - 1$$
$$f(y) = \begin{pmatrix} 2y \\ y^2 \end{pmatrix}$$

So note that

$$g : \mathbb{R} \longrightarrow \mathbb{R}, \qquad \text{and} \qquad f : \mathbb{R} \longrightarrow \mathbb{R}^2$$
$$[f \circ g](x) = \begin{pmatrix} 2(x-1) \\ (x-1)^2 \end{pmatrix}$$
$$D[f \circ g](x) = \begin{pmatrix} 2 \\ 2(x-1) \end{pmatrix}$$

Now let's see if we get the same answer doing it the chain rule way:

$$Dg(x) = 1$$
$$Df(y) = \begin{pmatrix} 2 \\ 2y \end{pmatrix}$$
$$Df(g(x))Dg(x) = \begin{pmatrix} 2 \\ 2(x-1) \end{pmatrix}$$

$$
\begin{aligned}
f(\mathbf{y}) &= f(y_1, y_2) \\
&= \begin{pmatrix} y_1^2 + y_2 \\ y_1 - y_1 y_2 \end{pmatrix}
\end{aligned}
$$

$$
\begin{aligned}
g(\mathbf{x}) &= g(x_1, x_2) \\
&= \begin{pmatrix} x_1^2 - x_2 \\ x_1 x_2 \end{pmatrix} \\
&= \mathbf{y}
\end{aligned}
$$

Both g and f take in two arguments and spit out a (2×1) vector, so we must have

$$
g : \mathbb{R}^2 \longrightarrow \mathbb{R}^2, \qquad \text{and} \qquad f : \mathbb{R}^2 \longrightarrow \mathbb{R}^2
$$

$$
\begin{aligned}
D[f \circ g](\mathbf{x}) &= Df(g(\mathbf{x})Dg(\mathbf{x}) \\
&= \begin{pmatrix} \frac{\partial f_1}{\partial y_1} & \frac{\partial f_1}{\partial y_2} \\ \frac{\partial f_2}{\partial y_1} & \frac{\partial f_2}{\partial y_2} \end{pmatrix} \cdot \begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} \end{pmatrix} \\
&= \begin{pmatrix} 2y_1 & 1 \\ 1 - y_2 & -y_1 \end{pmatrix} \cdot \begin{pmatrix} 2x_1 & -1 \\ x_2 & x_1 \end{pmatrix}
\end{aligned}
$$

and we know that

$$y_1 = x_1^2 - x_2$$

$$y_2 = x_1 x_2$$

So

$$
= \begin{pmatrix} 2x_1^2 - 2x_2 & 1 \\ 1 - x_1 x_2 & x_2 - x_1^2 \end{pmatrix} \cdot \begin{pmatrix} 2x_1 & -1 \\ x_2 & x_1 \end{pmatrix}
$$

$$
= \begin{pmatrix} 4x_1(x_1^2 - x_2) + x_2 & x_1 - 2(x_1^2 - x_2) \\ 2x_1(1 - x_1 x_2) + x_2(x_2 - x_1^2) & x_1(x_2 - x_1^2) + x_1 x_2 \end{pmatrix}
$$

Gradients and Level Sets

$$f : \mathbb{R}^3 \longrightarrow \mathbb{R}$$

so the graph is in $\mathbb{R}^4$ but the graph of the level set is in $\mathbb{R}^3$.

$$f(\mathbf{x}) = x_1^2 + x_2^2 + x_3^2 = 1$$

This is just a sphere of radius 1.

Tangents to Surfaces

A *surface* in $\mathbb{R}^{n+1}$ can be viewed as the solution to a system of equations. Represent a point in $\mathbb{R}^{n+1}$ as a pair (x, y), with $x \in \mathbb{R}^n$ and $y \in \mathbb{R}$. If $F\colon \mathbb{R}^{n+1} \longrightarrow \mathbb{R}$, then the set $\{(x, y) : F(x, y) = 0\}$ is typically an n dimensional set.

What is a tangent to this surface?

The tangent at (x_0, y_0) should be an n dimensional linear manifold in $\mathbb{R}^{n+1}$ that contains (x_0, y_0).

It should satisfy the approximation property: if (x, y) is a point on the surface that is close to (x_0, y_0), then it should be approximated up to first order by a point on the tangent.

Consider a function $G\colon \mathbb{R} \longrightarrow \mathbb{R}^{n+1}$ such that $G(0) = (x_0, y_0)$ and $F \circ G(t) \equiv 0$ for t in a neighborhood of 0.

G defines a curve on the surface through (x_0, y_0). A direction on the surface at (x_0, y_0) is just a direction of a curve through (x_0, y_0) or $DG(0)$.

By the chain rule it follows that

$$\nabla F(x_0, y_0) \cdot DG(0) = 0,$$

which implies that $\nabla F(x_0, y_0)$ is orthogonal to all of the directions on the surface. This generates a non-trivial hyperplane provided that $DF(x_0, y_0) \neq \mathbf{0}$.

Definition

Let $F\colon \mathbb{R}^{n+1} \longrightarrow \mathbb{R}$ be differentiable at the point (x_0, y_0). Assume that $F(x_0, y_0) = 0$ and that $DF(x_0, y_0) \neq \mathbf{0}$. The equation of the hyperplane tangent to the surface $F(x, y) = 0$ at the point (x_0, y_0) is

$$\nabla F(x_0, y_0) \cdot ((x, y) - (x_0, y_0)) = 0. \tag{2}$$

Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is differentiable at $x \in \mathbb{R}^n$. Consider the function $F(x, y) = f(x) - y$.
The surface $F(x, y) = 0$ is exactly the graph of f.
Hence the tangent to the surface is the tangent to the graph of f.
This means that the formula for the equation of the tangent hyperplane given above can be used to find the formula for the equation of the tangent to the graph of a function.

Theorem

If $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is differentiable at $x_0 \in \mathbb{R}^n$ then the vector $\nabla f(x_0)$ is normal (perpendicular) to the tangent vector of the level set of f at value $f(\mathbf{x})$ at point $x \in \mathbb{R}^n$ and the equation of the hyperplane tangent to the graph of f at the point $(x_0, f(x_0))$ is

$$\nabla f(x_0) \cdot (x - x_0)) = y - y_0.$$

Proof.

Substitute $\nabla F(x_0, y_0) = (\nabla f(x_0, -1)$ into equation (**??**) and re-arrange terms. □