Lecture 10

Econ 2001

2015 August 21

Lecture 10 Outline

1. Derivatives and Partial Derivatives
2. Differentiability
3. Tangents to Level Sets

Calculus!
This material will not be included in today's exam.

Announcement*:*
- Test 2 today at 3pm, in WWPH 4716; recitation at 1pm.
The exam will last an hour.

Differentiability: Roadmap

The objective is to define the "derivative" of a function that goes from a multimensional subset of the reals to another multidimensional subset of the reals.

- First, we look at properties of things we already know, namely functions from one interval to the real line.
- We use these properties to motivate the general definitions.
- The central idea is that the derivative of a function at a point is a linear approximation of that function as it "moves away" from that point.
- When the domain is an interval in $\mathbb{R}$ things are easy: "moving away" means only one thing.
- When the domain is an interval in $\mathbb{R}^n$ things are more complicated: "moving away" needs to be described more carefully.

Linear Approximations for Functions of One Variable

MAIN IDEA

Differentiation is about approximating a function with a linear function.

- Think of a line that intersects the graph of f at the points $(x, f(x))$ and the point $(x+\delta, f(x+\delta))$.
- This line has slope given by:

$$\frac{f(x+\delta)-f(x)}{(x+\delta)-x} = \frac{f(x+\delta)-f(x)}{\delta}$$

  - When f is linear, the line with this slope is f itself.
- When δ is small:

$$\lim_{\delta\to 0}\frac{f(x+\delta)-f(x)}{(x+\delta)-x} \quad = \quad \lim_{\delta\to 0}\frac{f(x+\delta)-f(x)}{\delta}$$

  - For the limit to make sense, x must be an interior point of the domain of f.
- If the limit exists, we say that f is differentiable at x and call it the derivative of f at x.

Linear Approximations for Functions of One Variable

Definition

Let $f : I \longrightarrow \mathbf{R}$, where $I \subseteq \mathbf{R}$ is an open interval. f is differentiable at $x_0 \in I$ if there exists an $a \in \mathbf{R}$ such that
$$\lim_{h \longrightarrow 0} \frac{f(x_0 + h) - f(x_0)}{h} = a$$

- Let $x = x_0 + h$ so that the expression above can be rewritten as
$$\lim_{x \longrightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} - a = 0$$
- Define the affine (?) function g as $g(x) = [f(x_0) - ax_0] + ax$ and observe that
$$\begin{aligned} \lim_{x \longrightarrow x_0} \left| \frac{f(x) - g(x)}{x - x_0} \right| &= \lim_{x \longrightarrow x_0} \left| \frac{f(x) - ([f(x_0) - ax_0] + ax)}{x - x_0} \right| \\ &= \lim_{x \longrightarrow x_0} \left| \frac{f(x) - f(x_0) + a(x_0 - x)}{x - x_0} \right| \\ &= \lim_{x \longrightarrow x_0} \left| \frac{f(x) - f(x_0)}{x - x_0} - a \right| \end{aligned}$$
- In other words, the difference between f and the affine function g disappears in the limit.

Differentiability: Functions of One Variable

Definition

Let $f : I \rightarrow \mathbf{R}$, where $I \subseteq \mathbf{R}$ is an open interval. f is differentiable at $x \in I$ if

$$\lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} = a$$

for some $a \in \mathbf{R}$.

- The equation above is equivalent to

$$\begin{aligned}
& \lim_{h \rightarrow 0} \frac{f(x+h) - (f(x) + ah)}{h} = 0 \\
\Leftrightarrow \quad & \forall \varepsilon > 0 \; \exists \delta > 0 \text{ such that } 0 < |h| < \delta \Rightarrow \left| \frac{f(x+h) - (f(x) + ah)}{h} \right| < \varepsilon \\
\Leftrightarrow \quad & \forall \varepsilon > 0 \; \exists \delta > 0 \text{ such that } 0 < |h| < \delta \Rightarrow \frac{|f(x+h) - (f(x) + ah)|}{|h|} < \varepsilon \\
\Leftrightarrow \quad & \lim_{h \rightarrow 0} \frac{|f(x+h) - (f(x) + ah)|}{|h|} = 0
\end{aligned}$$

- The last expression motivates the general definition we will see later.

Partial Derivatives

What happens when f's domain is $\mathbb{R}^n$?

- One possibility is to perturb only one variable; that is, take derivatives 'one dimension at a time'.
- We define this as the derivative when $\mathbf{x}$ is perturbed in the particular way described by $\mathbf{e}_i$ (the vector equal to 0 in all components but i, and equal to 1 in that component).

Definition

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

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

when this limit exists.

- This definition treats every other x_j as a constant: take the derivative as though f were a function of just x_i.
- As in the one variable case, partial derivatives need not exist.
- The partial derivative can be thought of as a function of one variable: h.
  - The one-variable function is of the form $f(\mathbf{x} + h\mathbf{e}_i)$ with $\mathbf{x}$ and $\mathbf{e}_i$ fixed.

Gradient

Definition

Given $f : \mathbb{R}^n \longrightarrow \mathbb{R}$, then the gradient of f at $\mathbf{x} \in \mathbb{R}^n$ is

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

- This is the vector of the n partial derivatives.

Note

The gradient is a vector in $\mathbb{R}^n$.

Directional Derivatives

What happens when f's domain is $\mathbb{R}^n$?

- Another possibility is to perturb the function in a particular direction.
- We define this as the derivative when $\mathbf{x}$ moves along some unit vector that is non zero in more than one component.

Definition

Given $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ and let $\mathbf{v}$ be a unit vector in $\mathbb{R}^n$ ($\|v\| = 1$). The directional derivative of f in the direction $\mathbf{v}$ at $\mathbf{x}$ is defined as

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

- It follows from the definition, that

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

  - The ith partial derivative is the directional derivative in the direction $\mathbf{e}_i$.
- The directional derivative can also be thought of as a function of one variable, namely h.
  - The one-variable function is of the form $f(\mathbf{x} + h\mathbf{v})$ with $\mathbf{x}$ and $\mathbf{v}$ fixed.

What is the direction from x that most increases the value of f?

- Answer: the direction given by the gradient.

Theorem

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

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

This result follows because

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

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

- Any idea why we may care about this?

Differentiability in General

So far, we have seen functions from $\mathbb{R}^n$ to $\mathbb{R}$. Notation gets more messy when the range is $\mathbb{R}^m$.

- One can think of f as a family of functions f^i for $i = 1, 2, ..., m$.
- Despite the messy notation, the conceptual difference is small.
- A function is differentiable when it can be approximated by a linear function. So, the derivatives are given by this linear approximation.
- Also, we need to generalize

$$\lim_{h \to 0} \frac{|f(x+h) - (f(x) + ah)|}{|h|} = 0$$

  when the numerator and denominator can be vectors.
- Finally, the function may not be defined on the whole of $\mathbb{R}^n$, but this is easy to take care of.
- We will see how the previous definitions can be thought of as special cases.

Differentiability

Definition

If $X \subset \mathbf{R}^n$ is open, $f : X \to \mathbf{R}^m$ is differentiable at $x \in X$ if there exist $T_x \in L(\mathbf{R}^n, \mathbf{R}^m)$ such that

$$\lim_{\substack{\mathbf{h} \to 0 \\ \mathbf{h} \in \mathbf{R}^n}} \frac{\|f(\mathbf{x}+\mathbf{h}) - (f(\mathbf{x}) + T_x(\mathbf{h}))\|}{\|\mathbf{h}\|} = 0$$

f is differentiable if it is differentiable at all $\mathbf{x} \in X$.

- T_x is uniquely determined by the equation above.
- The definition requires that the same linear operator T_x works no matter how $\mathbf{h}$ approaches zero.
- $f(\mathbf{x}) + T_x(\mathbf{h})$ is the best linear approximation to $f(\mathbf{x}+\mathbf{h})$ for sufficiently small $\mathbf{h}$.
- Objects inside the $\|\cdot\|$ are vectors: numerator and denominator are lenghts of vectors in $\mathbf{R}^m$ and $\mathbf{R}^n$ respectively.
- $T_x \in L(\mathbf{R}^n, \mathbf{R}^m)$ is a big ugly thing: an $m \times n$ matrix.

Definition

The linear transformation T_x is called the differential of f at x, and is denoted df_x.

Notation: Big-Oh and little-oh

read this "$\mathbf{y}$ is big-Oh of $\|\mathbf{h}\|^n$"

- $\mathbf{y} = O(\|\mathbf{h}\|^n)$ as $\mathbf{h} \to 0$ means
$$\exists K, \delta > 0 \text{ such that } \|\mathbf{h}\|^n < \delta \Rightarrow \|\mathbf{y}\| \le K\|\mathbf{h}\|^n$$

read this "y is little-oh of $\|\mathbf{h}\|^n$"

- $\mathbf{y} = o(\|\mathbf{h}\|^n)$ as $\mathbf{h} \to 0$ means
$$\lim_{h\to 0} \frac{\|\mathbf{y}\|}{\|\mathbf{h}\|^n} = 0$$
- Note that $\mathbf{y} = O(\|\mathbf{h}\|^{n+1})$ as $\mathbf{h} \to 0$ implies $\mathbf{y} = o(\|\mathbf{h}\|^n)$ as $\mathbf{h} \to 0$.
- Using the notation above:
  - f is differentiable at $\mathbf{x} \Leftrightarrow \exists T_x \in L(\mathbf{R}^n, \mathbf{R}^m)$ such that
  $$f(\mathbf{x}+\mathbf{h}) = f(\mathbf{x}) + T_x(\mathbf{h}) + o(\mathbf{h}) \text{ as } \mathbf{h} \to 0$$

Definition

Let $E_f(\mathbf{h}) = f(\mathbf{x}+\mathbf{h}) - (f(\mathbf{x}) + df_x(\mathbf{h}))$ be the error term

Differentiability and Error Term

Using the previous slide's notation:

$$f \text{ is differentiable at } \mathbf{x} \Leftrightarrow E_f(\mathbf{h}) = o(\mathbf{h}) \text{ as } \mathbf{h} \to 0$$

The Jacobian Matrix

- The Jacobian of f at x, denoted as $Df(\mathbf{x})$, is the matrix corresponding to df_x with respect to the standard basis.

- Let $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ be the standard basis of $\mathbf{R}^n$.
- Look in direction $\mathbf{e}_j$ (note that $\|\gamma \mathbf{e}_j\| = |\gamma|$).

$$\begin{aligned}
o(\gamma) &= f(\mathbf{x} + \gamma \mathbf{e}_j) - (f(\mathbf{x}) + T_x(\gamma \mathbf{e}_j)) \\
&= f(\mathbf{x} + \gamma \mathbf{e}_j) - \left(f(\mathbf{x}) + \begin{pmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mj} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \gamma \\ 0 \\ \vdots \\ 0 \end{pmatrix} \right) \\
&= f(\mathbf{x} + \gamma \mathbf{e}_j) - \left(f(\mathbf{x}) + \begin{pmatrix} \gamma a_{1j} \\ \vdots \\ \gamma a_{mj} \end{pmatrix} \right)
\end{aligned}$$

- For $i = 1, \ldots, m$, let f^i denote the i^{th} component of the function f:

$$f^i(\mathbf{x} + \gamma \mathbf{e}_j) - \left(f^i(\mathbf{x}) + \gamma a_{ij} \right) = o(\gamma) \quad \Rightarrow \quad a_{ij} = \frac{\partial f^i}{\partial x_j}(\mathbf{x})$$

  - Go back to the definition of partial derivative to make sure you see this.

Jacobian and Partial Derivatives

All this algebra can be summarized as follows.

Theorem

Suppose $X \subset \mathbf{R}^n$ is open and $f : X \to \mathbf{R}^m$ is differentiable at $\mathbf{x} \in X$. Then $\frac{\partial f^i}{\partial x_j}$ exists for $1 \le i \le m$, $1 \le j \le n$, and

$$Df(\mathbf{x}) = \begin{pmatrix} \frac{\partial f^1}{\partial x_1}(\mathbf{x}) & \cdots & \frac{\partial f^1}{\partial x_n}(\mathbf{x}) \\ \vdots & \ddots & \vdots \\ \frac{\partial f^m}{\partial x_1}(\mathbf{x}) & \cdots & \frac{\partial f^m}{\partial x_n}(\mathbf{x}) \end{pmatrix}$$

In words: the Jacobian is the matrix of partial derivatives.

Directional Derivatives and Partial Derivatives

NOTE

Suppose $X \subset \mathbf{R}^n$ open, $f : X \to \mathbf{R}^m$ is differentiable at x, and take $\|\mathbf{u}\| = 1$.

- Then

$$f(\mathbf{x} + \gamma\mathbf{u}) - (f(\mathbf{x}) + T_x(\gamma\mathbf{u})) = o(\gamma) \text{ as } \gamma \to 0$$

$\Rightarrow$

$$f(\mathbf{x} + \gamma\mathbf{u}) - (f(\mathbf{x}) + \gamma T_x(\mathbf{u})) = o(\gamma) \text{ as } \gamma \to 0$$

$\Rightarrow$

$$\lim_{\gamma \to 0} \frac{f(\mathbf{x} + \gamma\mathbf{u}) - f(\mathbf{x})}{\gamma} = T_x(\mathbf{u}) = Df(\mathbf{x})\mathbf{u}$$

- That is, the directional derivative in the direction $\mathbf{u}$ (with $\|\mathbf{u}\| = 1$) is

$$Df(\mathbf{x})\mathbf{u} \in \mathbf{R}^m$$

- The directional derivative is a weighted average of partial derivatives.

Summary so Far

- f differentiable means $\exists\ T_x \in L(\mathbf{R}^n, \mathbf{R}^m)$ s.t.

$$\lim_{\substack{\mathbf{h}\to 0\\ \mathbf{h}\in\mathbf{R}^n}} \frac{\|f(\mathbf{x}+\mathbf{h}) - (f(\mathbf{x}) + T_x(\mathbf{h}))\|}{\|\mathbf{h}\|} = 0$$

- The differential df_x is the linear transformation T_x
- The Jacobian $Df(\mathbf{x})$ is the matrix corresponding to df_x with respect to the standard basis:

$$Df(\mathbf{x}) = \begin{pmatrix} \frac{\partial f^1}{\partial x_1}(\mathbf{x}) & \cdots & \frac{\partial f^1}{\partial x_n}(\mathbf{x}) \\ \vdots & \ddots & \vdots \\ \frac{\partial f^m}{\partial x_1}(\mathbf{x}) & \cdots & \frac{\partial f^m}{\partial x_n}(\mathbf{x}) \end{pmatrix}$$

- One would hope that existence of all the partial derivates $\frac{\partial f^i}{\partial x_j}(\mathbf{x})$ implies that the function is differentiable... but this is not enough.

REMARK

- If f is differentiable at $\mathbf{x}$, then all first-order partial derivatives $\frac{\partial f^i}{\partial x_j}$ exist at $\mathbf{x}$.
- However, *the converse is false*: existence of all the first-order partial derivatives does not imply that f is differentiable.

Differentiability and Continuity

- The missing piece is continuity of the partial derivatives:

Theorem

If all the first-order partial derivatives $\frac{\partial f^i}{\partial x_j}$ $(1 \leq i \leq m,\ 1 \leq j \leq n)$ *exist and are continuous at* $\mathbf{x}$, *then* f *is differentiable at* $\mathbf{x}$.

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

Definition

Let $X \subset \mathbf{R}^n$ be open. A function $f : X \to \mathbf{R}^m$ is continuously differentiable on X if

1. f is differentiable on X and
2. df_x is a continuous function of x from X to $L(\mathbf{R}^n, \mathbf{R}^m)$.

f is C^k if all partial derivatives of order $\leq k$ exist and are continuous in X.

Theorem

Suppose $X \subset \mathbf{R}^n$ *is open and* $f : X \to \mathbf{R}^m$. *Then* f *is continuously differentiable on* X *if and only if* f *is* C^1.

Properties of the Derivative

Since the derivative is a linear transformation, some of these are easy to prove.

Theorem

If $g, f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ are both differentiable at $\mathbf{x} \in \mathbb{R}^n$, then

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

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

For the case $m = 1$:

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

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

Chain Rule

Theorem (Chain Rule)

Let $X \subset \mathbf{R}^n$, $Y \subset \mathbf{R}^m$ be open, $f : X \to Y$, $g : Y \to \mathbf{R}^p$. Let $x_0 \in X$ and $F = g \circ f$.

If f is differentiable at x_0 and g is differentiable at $f(x_0)$, then $F = g \circ f$ is differentiable at x_0 and

$$dF_{x_0} = dg_{f(x_0)} \circ df_{x_0} \qquad \textit{(composition of linear transformations)}$$

and

$$DF(x_0) = Dg(f(x_0))Df(x_0) \qquad \textit{(matrix multiplication)}$$

Remark: *This mirrors the univariate case (replace the univariate derivative by a linear transformation), and the proof is similar (add linear algebra).*

Chain Rule Special Case

Let $f : \mathbb{R}^m \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}^m$ then the Chain rule says:

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

Chain Rule: Examples

Example

Let $g : \mathbb{R} \longrightarrow \mathbb{R}$ and $f : \mathbb{R} \longrightarrow \mathbb{R}^2$ be defined by

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

- Hence

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

- Having stated the obvious, see how the chain rule would work:

$$Dg(x) = 1 \qquad \text{and} \qquad Df(y) = \begin{pmatrix} 2 \\ 2y \end{pmatrix}$$

- Hence

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

Chain Rule: Examples

Example

$$f(\mathbf{y}) = f(y_1, y_2) = \begin{pmatrix} y_1^2 + y_2 \\ y_1 - y_1 y_2 \end{pmatrix} \quad \text{and} \quad g(\mathbf{x}) = g(x_1, x_2) = \begin{pmatrix} x_1^2 - x_2 \\ x_1 x_2 \end{pmatrix} = \mathbf{y}$$

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

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

- $$\begin{aligned} D[f \circ g](\mathbf{x}) &= Df(g(\mathbf{x}))Dg(\mathbf{x}) \\ &= \begin{pmatrix} \frac{\partial f_1}{\partial y_1} & \frac{\partial f_1}{\partial y_2} \\ \frac{\partial f_2}{\partial y_1} & \frac{\partial f_2}{\partial y_2} \end{pmatrix} \cdot \begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} \end{pmatrix} = \begin{pmatrix} 2y_1 & 1 \\ 1 - y_2 & -y_1 \end{pmatrix} \cdot \begin{pmatrix} 2x_1 & -1 \\ x_2 & x_1 \end{pmatrix} \end{aligned}$$
  - and we know that $y_1 = x_1^2 - x_2$ and $y_2 = x_1 x_2$
- So $$\begin{aligned} D[f \circ g](\mathbf{x}) &= \begin{pmatrix} 2x_1^2 - 2x_2 & 1 \\ 1 - x_1 x_2 & x_2 - x_1^2 \end{pmatrix} \cdot \begin{pmatrix} 2x_1 & -1 \\ x_2 & x_1 \end{pmatrix} \\ &= \begin{pmatrix} 4x_1(x_1^2 - x_2) + x_2 & x_1 - 2(x_1^2 - x_2) \\ 2x_1(1 - x_1 x_2) + x_2(x_2 - x_1^2) & x_1(x_2 - x_1^2) + x_1 x_2 \end{pmatrix} \end{aligned}$$

Graph

Remember from earlier definitions:

The graph of $f : X \to Y$ is given by

$$Gr(f) = \{(x, y) : y = f(x)\}.$$

- 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 Sets and Countours

Definition

The level set of a function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is defined as the set

$$\{\mathbf{x} \in X : f(\mathbf{x}) = c\}$$

for any $c \in \mathbb{R}$.

- This is the set of points such that the function achieves a given value.
- While the graph of the function is a subset of $\mathbb{R}^{n+1}$, the level sets are subsets of $\mathbb{R}^n$.

Definition

The upper contour set of $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ at x_0 is the set

$$\{x \in X : f(x) \geq f(x_0)\}$$

The lower contour set of $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ at x_0 is the set

$$\{x \in X : f(x) \leq f(x_0)\}$$

Tangents to Surfaces

- A surface in $\mathbb{R}^{n+1}$ can be viewed as the solution to a system of equations.
- A point in $\mathbb{R}^{n+1}$ can be represented as a pair (x, y), with $x \in \mathbb{R}^n$ and $y \in \mathbb{R}$.
- If $F : \mathbb{R}^{n+1} \longrightarrow \mathbb{R}$, 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 also 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.

Tangents to Surfaces: Gradients and Level Sets

- Let $F : \mathbb{R}^{n+1} \longrightarrow \mathbb{R}$ be differentiable at (x_0, y_0).
- Consider a function $G : \mathbb{R} \longrightarrow \mathbb{R}^{n+1}$ such that

$$G(0) = (x_0, y_0) \qquad \text{and} \qquad F \circ G(t) \equiv \mathbf{0} \text{ for } t \text{ in a neighborhood of } 0$$

  - 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

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

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

Tangents to Surfaces: Gradients and Level Sets

Definition

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

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

NOTE

Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ be 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.
- Thus, 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.

Tangents to Surfaces: Gradients and Level Sets

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 $f(x)$ 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 the equation on the previous slide and re-arrange terms. □

Tangent to Surfaces: Example

Find the tangent plane to $\{\mathbf{x} \in \mathbb{R}^3 : x_1 x_2 - x_3^2 = 6\} \subset \mathbb{R}^3$ at $\widehat{\mathbf{x}} = (2, 5, 2)$

- If you let $f(\mathbf{x}) = x_1 x_2 - x_3^2$, then this is a level set of f for value 6.

$$\nabla f(\mathbf{x}) = (\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \frac{\partial f}{\partial x_3}) = (x_2, x_1, -2x_3)$$

$$\nabla f(\widehat{\mathbf{x}}) = \nabla f(\mathbf{x}) \mid_{\mathbf{x}=(2,5,2)} = (5, 2, -4)$$

- Tangent Plane:

$$\begin{aligned}\{\mathbf{y} \in \mathbb{R}^3 : \widehat{\mathbf{x}} + \mathbf{y} : \mathbf{y} \cdot \nabla f(\widehat{\mathbf{x}}) = 0\} &= \{(2,5,2) + (y_1, y_2, y_3) : 5y_1 + 2y_2 - 4y_3 = 0\} \\ &= \{\mathbf{x} \in \mathbb{R}^3 : 5x_1 - 10 + 2x_2 - 10 - 4x_3 + 8 = 0\} \\ &= \{\mathbf{x} \in \mathbb{R}^3 : 5x_1 + 2x_2 - 4x_3 = 12\}\end{aligned}$$

Tangent to Surfaces: Example

Example

- Let $f(x, y, z) = 3x^2 + 2xy - z^2$ and consider the level set at $(2, 1, 3)$.
  - Since $f(2, 1, 3) = 7$, this level set is
    $$\{(x, y, z) : f(x, y, z) = 7\}.$$
  - It is a two-dimensional surface in $\mathbb{R}^3$ that can be written as $F(x, y, z) = 0$:
    $$f(x, y, z) - 7 = 0$$
  - The tangent to the level set of f is an hyperplane in $\mathbb{R}^3$
    $$\nabla f(x, y, z) = (6x + 2y, 2x, -2z)$$
  - At the point $(2, 1, 3)$, the hyperplane has normal equal to
    $$\nabla f(2, 1, 3) = (14, 4, -6)$$
  - Hence the equation of the hyperplane to the level set at $(2, 1, 3)$ is :
    $$(14, 4, -6) \cdot (x - 2, y - 1, z - 3) = 0 \qquad \text{or} \qquad 14x + 4y - 6z = 14.$$
- The graph of f is a three-dimensional subset of $\mathbb{R}^4$:
  $$\{(x, y, z, w) : w = f(x, y, z)\}$$
  - A point on this surface is $(2, 1, 3, 7) = (x, y, z, w)$.
  - The tangent hyperplane at this point can be written as:
    $$w - 7 = \nabla f(2, 1, 3) \cdot (x - 2, y - 1, z - 3) = 14x + 4y - 6z - 14$$
    or
    $$14x + 4y - 6z - w = 7.$$

Monday

- We start using calculus in different applications, and get ready for unconstrained optimization.

1. Homogeneous Functions and Euler's Theorem
2. Mean Value Theorem
3. Taylor's Theorem

Problem Set 10

This problem set is extremely important to understand the definitions we covered today.
No need to work on it today before the test, but please work on it during the weekend.
Eric will cover both problem Set 10 and 11 in Section on Monday.