Announcements

- PS 4 due Tuesday
- comments on exam in lecture Thursday

Econ 204 2023

Lecture 11

Outline

1. Derivatives
2. Chain Rule
3. Mean Value Theorem
4. Taylor's Theorem

Derivatives

Definition 1. *Let $f : I \to \mathbf{R}$, where $I \subseteq \mathbf{R}$ is an open interval. f is* differentiable *at $x \in I$ if*

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

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

$a = f'(x)$

This is equivalent to $\exists a \in \mathbf{R}$ such that

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

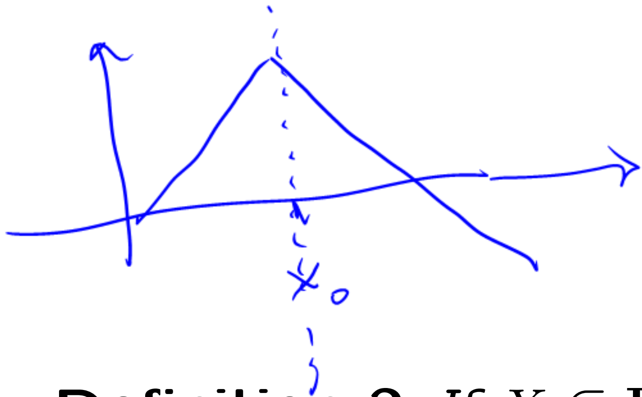
$$\Leftrightarrow \quad \forall \varepsilon > 0 \ \exists \delta > 0 \text{ s.t. } 0 < |h| < \delta \Rightarrow \left| \frac{f(x+h) - (f(x) + ah)}{h} \right| < \varepsilon$$

$$\Leftrightarrow \quad \forall \varepsilon > 0 \ \exists \delta > 0 \text{ s.t. } 0 < |h| < \delta \Rightarrow \frac{|f(x+h) - (f(x) + ah)|}{|h|} < \varepsilon$$

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

Notice: $T : \mathbb{R} \to \mathbb{R}$ is a linear transformation

$\Longleftrightarrow$ $T(h) = rh$ for some $r \in \mathbb{R}$



Derivatives

Definition 2. *If* $X \subseteq \mathbf{R}^n$ *is open,* $f : X \to \mathbf{R}^m$ *is* differentiable at $x \in X$ *if* $\exists T_x \in L(\mathbf{R}^n, \mathbf{R}^m)$ *such that*

$$\lim_{h \to 0, h \in \mathbf{R}^n} \frac{\|f(x+h) - (f(x) + T_x(h))\|}{\|h\|} = 0 \qquad (1)$$

f *is* differentiable *if it is differentiable at all* $x \in X$.

Note that T_x is uniquely determined by Equation (1).

The definition requires that **one** linear ~~operator~~ transformation T_x works no matter how h approaches zero.

In this case, $f(x) + T_x(h)$ is the best linear approximation to $f(x+h)$ for sufficiently small h.

$$f(x+h) - f(x) \approx T_x(h)$$

$x \in \mathbb{R}^n \qquad |x| = \|x\|$

Big-Oh and little-oh

Notation:

- $y = O(\|h\|^n)$ as $h \to 0$ – read "y is big-Oh of $\|h\|^n$" – means

$$\exists K, \delta > 0 \text{ s.t. } 0 < \|h\| < \delta \Rightarrow \|y\| \le K\|h\|^n$$

$\frac{\|y\|}{\|h\|^n}$ is bounded as $h \to 0$

- $y = o(\|h\|^n)$ as $h \to 0$ – read "y is little-oh of $\|h\|^n$" – means

$$\lim_{h \to 0} \frac{\|y\|}{\|h\|^n} = 0$$

$\frac{\|y\|}{\|h\|^n} \to 0$ as $h \to 0$

- Nested: $o(\|h\|^n) \Rightarrow O(\|h\|^n)$
- Note that $y = O(|h|^{n+1})$ as $h \to 0$ implies $y = o(|h|^n)$ as $h \to 0$.
- Also $y = O(\|h\|^n)$ or $y = o(\|h\|^n) \Rightarrow y \to 0$ as $h \to 0$.

Using this notation: f is differentiable at $x \Leftrightarrow \exists T_x \in L(\mathbf{R}^n, \mathbf{R}^m)$ such that

$$f(x+h) = f(x) + T_x(h) + o(|h|) \text{ as } h \to 0$$

More Notation

Notation:

- df_x is the linear transformation T_x
- $Df(x)$ is the matrix representation of df_x with respect to the standard basis.

  This is called the *Jacobian* or *Jacobian matrix* of f at x
- $E_f(h) = f(x+h) - (f(x) + df_x(h))$ is the *error term*

Using this notation,

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

What's $Df(x)$?

Now compute $Df(x) = (a_{ij})$. Let $\{e_1, \ldots, e_n\}$ be the standard basis of $\mathbf{R}^n$. Look in direction e_j (note that $\|\gamma e_j\| = |\gamma|$). $\gamma e_j \to 0$ as $\gamma \to 0$

$$o(\gamma) = f(x+\gamma e_j) - \Big(f(x) + \underbrace{T_x(\gamma e_j)}\Big)$$

$df_x(\gamma e_j)$

$$= f(x+\gamma e_j) - \left(f(x) + \begin{pmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mj} & \cdots & a_{mn} \end{pmatrix} \begin{array}{c} \left(\begin{array}{c} 0 \\ \vdots \\ 0 \\ \gamma \\ 0 \\ \vdots \\ 0 \end{array}\right) \end{array} \begin{array}{l} \\ \\ \\ \leftarrow j^{\text{th}} \text{ spot} \\ \\ \\ \\ \end{array} \right)$$

$$= f(x+\gamma e_j) - \left(f(x) + \begin{pmatrix} \gamma a_{1j} \\ \vdots \\ \gamma a_{mj} \end{pmatrix} \right)$$

$f : \mathbb{R}^n \to \mathbb{R}^m$, $f = (f^1, \ldots, f^m)$ where

$f^i : \mathbb{R}^n \to \mathbb{R} \quad \forall i$

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

$$\forall i: \quad f^i(x + \gamma e_j) - \left(f^i(x) + \gamma a_{ij}\right) = o(\gamma)$$

$$\text{so } a_{ij} = \frac{\partial f^i}{\partial x_j}(x)$$

Derivatives and Partial Derivatives

Theorem 1 (Thm. 3.3)**.** *Suppose $X \subseteq \mathbf{R}^n$ is open and $f : X \to \mathbf{R}^m$ is differentiable at $x \in X$. Then $\frac{\partial f^i}{\partial x_j}(x)$ exists for $1 \le i \le m$, $1 \le j \le n$, and*

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

of f   of f

i.e. the Jacobian at x is the matrix of partial derivatives at x .

h

$$f(x_1, x_2) = \begin{cases} \dfrac{x_1 x_2}{x_1^2 + x_2^2} & x \neq (0,0) \\ 0 & x = (0,0) \end{cases}$$

Derivatives and Partial Derivatives

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

The missing piece is continuity of the partial derivatives:

Theorem 2 (Thm. 3.4). *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 x, then f is differentiable at x.*

Directional Derivatives

Suppose $X \subseteq \mathbf{R}^n$ open, $f : X \to \mathbf{R}^m$ is differentiable at x, and $\|u\| = 1$. $u \in \mathbb{R}^n$, $\gamma u \to 0$ as $\gamma \to 0$: $\|\gamma u\| = |\gamma| \, \|u\| = |\gamma|$

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

$$\Rightarrow \; f(x + \gamma u) - (f(x) + \gamma T_x(u)) = o(\gamma) \text{ as } \gamma \to 0 \qquad (T \text{ linear})$$

$$\Rightarrow \; \lim_{\gamma \to 0} \frac{f(x + \gamma u) - f(x)}{\gamma} = \underbrace{T_x(u)}_{= df_x(u)} = Df(x)u$$

i.e. the directional derivative in the direction u (with $\|u\| = 1$) is

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

Chain Rule

Theorem 3 (Thm. 3.5, Chain Rule). *Let $X \subseteq \mathbf{R}^n$, $Y \subseteq \mathbf{R}^m$ 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*

$$\begin{aligned} dF_{x_0} &= dg_{f(x_0)} \circ df_{x_0} \\ &\quad \textit{(composition of linear transformations)} \\ DF(x_0) &= Dg(f(x_0))Df(x_0) \\ &\quad \textit{(matrix multiplication)} \end{aligned}$$

Remark: The statement is exactly the same as in the univariate case, except we replace the univariate derivative by a linear transformation. The proof is more or less the same, with a bit of linear algebra added.

Mean Value Theorem

Theorem 4 (Thm. 1.7, Mean Value Theorem, Univariate Case)**.**
$a<b$ *Let* $a, b \in \mathbf{R}$. *Suppose* $f : [a, b] \to \mathbf{R}$ *is continuous on* $[a, b]$ *and differentiable on* (a, b). *Then there exists* $c \in (a, b)$ *such that*

$$\frac{f(b) - f(a)}{b - a} = f'(c)$$

that is, such that

$$f(b) - f(a) = f'(c)(b - a)$$

Proof. Consider the function $g : [a,b] \to \mathbb{R}$

$$g(x) = f(x) - f(a) - \frac{f(b) - f(a)}{b - a}(x - a)$$

Note $g(a) = g(b) = 0$.

Then $g(a) = 0 = g(b)$. Note that for $x \in (a, b)$,

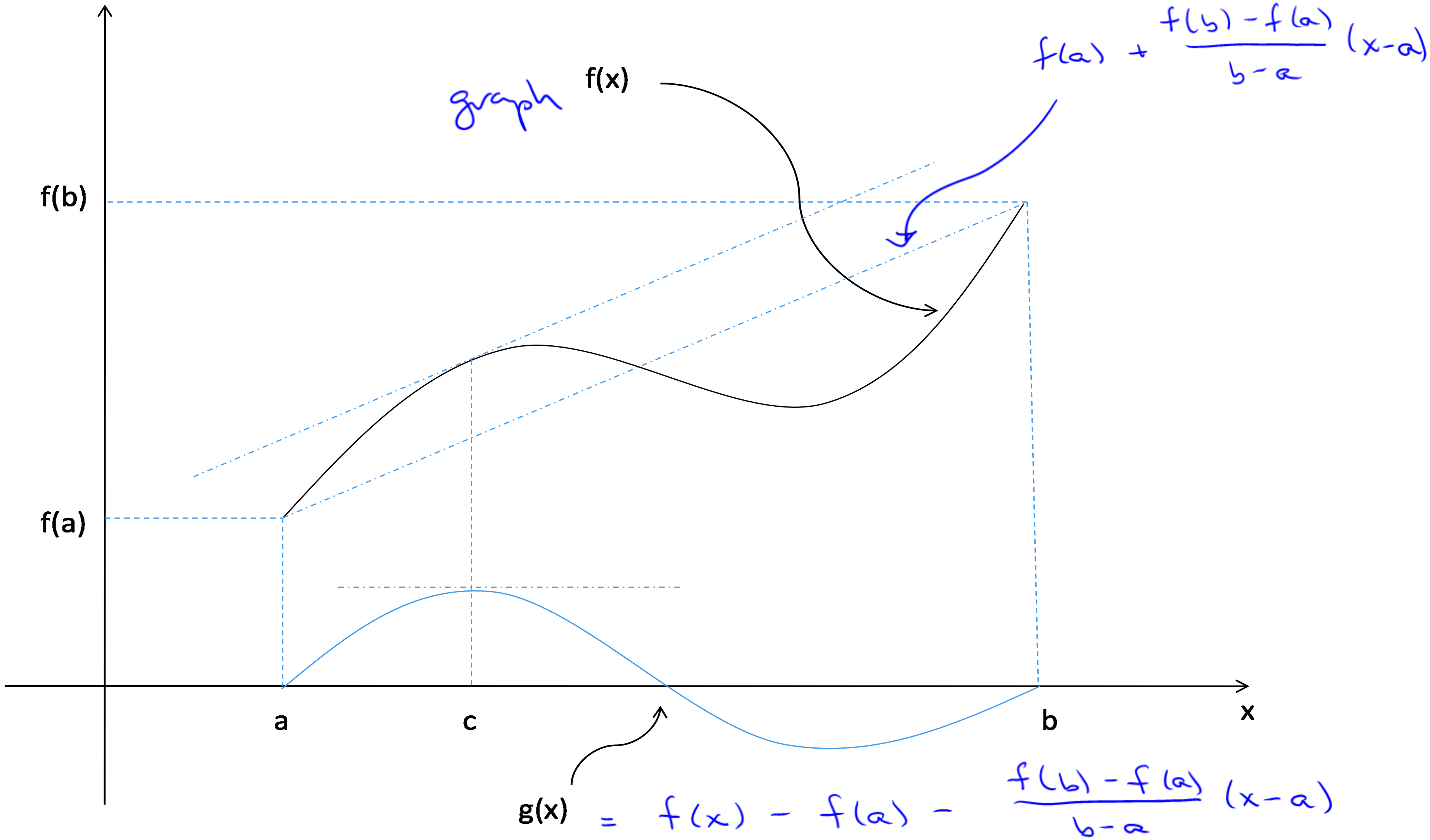
$$g'(x) = f'(x) - \frac{f(b) - f(a)}{b - a}$$

so it suffices to find $c \in (a, b)$ such that $g'(c) = 0$.

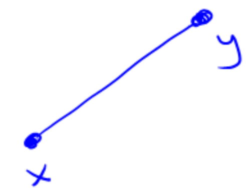
Case I: If $g(x) = 0$ for all $x \in [a, b]$, choose an arbitrary $c \in (a, b)$, and note that $g'(c) = 0$, so we are done.

Case II: Suppose $g(x) > 0$ for some $x \in [a, b]$. Since g is continuous on $[a, b]$, it attains its maximum at some point $c \in (a, b)$. Since g is differentiable at c and c is an interior point of the domain of g, we have $g'(c) = 0$, and we are done.

Case III: If $g(x) < 0$ for some $x \in [a, b]$, the argument is similar to that in Case II. $\square$


f(b)
f(a)
graph f(x)
$f(a) + \frac{f(b)-f(a)}{b-a}(x-a)$
a
c
b
x
$g(x) = f(x) - f(a) - \frac{f(b)-f(a)}{b-a}(x-a)$

Mean Value Theorem



Notation:

$$\ell(x, y) = \{\alpha x + (1 - \alpha)y : \alpha \in [0, 1]\}$$

is the line segment from x to y. $x, y \in \mathbb{R}^n$

Theorem 5 (Mean Value Theorem)**.** *Suppose $f : \mathbf{R}^n \to \mathbf{R}$ is differentiable on an open set $X \subseteq \mathbf{R}^n$, $x, y \in X$ and $\ell(x, y) \subseteq X$. Then there exists $z \in \ell(x, y)$ such that*

$$f(y) - f(x) = Df(z)(y - x)$$

$$= df_z(y - x)$$

Consider $\bar{f} : [0,1] \to \mathbb{R}$ given by

$$\bar{f}(\alpha) = f(\alpha x + (1-\alpha)y)$$

$f = (f^1, \dots, f^m)$, $f^i : \mathbb{R}^n \to \mathbb{R}$ $\forall i$

Notice that the statement is exactly the same as in the univariate case. For $f : \mathbf{R}^n \to \mathbf{R}^m$, we can apply the Mean Value Theorem to each component, to obtain $z_1, \dots, z_m \in \ell(x, y)$ such that

$$f^i(y) - f^i(x) = Df^i(z_i)(y - x)$$

However, we cannot always find a single z which works for every component.

Note that each $z_i \in \ell(x, y) \subset \mathbf{R}^n$; there are m of them, one for each component in the range.

Mean Value Theorem

Theorem 6. *Suppose $X \subset \mathbf{R}^n$ is open and $f : X \to \mathbf{R}^m$ is differentiable. If $x, y \in X$ and $\ell(x,y) \subseteq X$, then there exists $z \in \ell(x,y)$ such that*

$$\begin{aligned} \|f(y) - f(x)\| &\leq \|df_z(y-x)\| \quad = \|Df(z)(y-x)\| \\ &\leq \|df_z\|\|y-x\| \end{aligned}$$

$$\leq \|Df(z)\| \, \|y-x\|$$

norm on linear transformations from Friday

Mean Value Theorem

Remark: To understand why we don't get equality, consider $f : [0,1] \to \mathbf{R}^2$ defined by

$$f(t) = (\cos 2\pi t, \sin 2\pi t)$$

f maps $[0,1]$ to the unit circle in $\mathbf{R}^2$. Note that $f(0) = f(1) = (1,0)$, so $|f(1) - f(0)| = 0$. However, for any $z \in [0,1]$,

$$\begin{aligned} |df_z(1-0)| &= |2\pi(-\sin 2\pi z, \cos 2\pi z)| \\ &= 2\pi\sqrt{\sin^2 2\pi z + \cos^2 2\pi z} \\ &= 2\pi \quad \neq |f(1)-f(0)| = 0 \end{aligned}$$

So $\quad f(1) - f(0) \neq Df(z)\cdot(1-0) \quad \forall z \in [0,1]$

Taylor's Theorem – $\mathbf{R}$

Theorem 7 (Thm. 1.9, Taylor's Theorem in $\mathbf{R}$). *Let $f : I \to \mathbf{R}$ be n-times differentiable, where $I \subseteq \mathbf{R}$ is an open interval. If $x, x+h \in I$, then*

$$f(x+h) = f(x) + \sum_{k=1}^{n-1} \frac{f^{(k)}(x)h^k}{k!} + E_n$$

where $f^{(k)}$ is the k^{th} derivative of f and

$$E_n = \frac{f^{(n)}(x+\lambda h)h^n}{n!} \text{ for some } \lambda \in (0,1)$$

n^{th} order error term or "remainder"

Motivation: Let

$$
\begin{aligned}
T_n(h) &= f(x) + \sum_{k=1}^{n} \frac{f^{(k)}(x)h^k}{k!} \\
&= f(x) + f'(x)h + \frac{f''(x)h^2}{2} + \cdots + \frac{f^{(n)}(x)h^n}{n!} \\
T_n(0) &= f(x) \\
T_n'(h) &= f'(x) + f''(x)h + \cdots + \frac{f^{(n)}(x)h^{n-1}}{(n-1)!} \\
T_n'(0) &= f'(x) \\
T_n''(h) &= f''(x) + \cdots + \frac{f^{(n)}(x)h^{n-2}}{(n-2)!} \\
T_n''(0) &= f''(x) \\
&\vdots \\
T_n^{(n)}(0) &= f^{(n)}(x)
\end{aligned}
$$

$n^{\underline{th}}$ order polynomial in h

so $T_n(h)$ is the unique n^{th} degree polynomial such that

$$\begin{aligned} T_n(0) &= f(x) \\ T_n'(0) &= f'(x) \\ &\vdots \\ T_n^{(n)}(0) &= f^{(n)}(x) \end{aligned}$$

Taylor's Theorem – $\mathbf{R}$

Theorem 8 (Alternate Taylor's Theorem in $\mathbf{R}$)**.** *Let $f : I \to \mathbf{R}$ be n times differentiable, where $I \subseteq \mathbf{R}$ is an open interval and $x \in I$. Then*

$$f(x+h) = f(x) + \sum_{k=1}^{n} \frac{f^{(k)}(x)h^k}{k!} + o\left(h^n\right) \text{ as } h \to 0$$

If f is $(n+1)$ times continuously differentiable, then

$$f(x+h) = f(x) + \sum_{k=1}^{n} \frac{f^{(k)}(x)h^k}{k!} + O\left(h^{n+1}\right) \text{ as } h \to 0$$

Remark: The first equation in the statement of the theorem is essentially a restatement of the definition of the n^{th} derivative. The second statement is proven from Theorem 1.9, and the continuity of the derivative.

C^k Functions

Definition 3. *Let $X \subseteq \mathbf{R}^n$ be open. A function $f : X \to \mathbf{R}^m$ is* continuously differentiable *on X if*

- *f is differentiable on X and*

$df : X \to L(\mathbb{R}^n, \mathbb{R}^m)$

- *df_x is a continuous function of x from X to $L(\mathbf{R}^n, \mathbf{R}^m)$, with respect to the operator norm $\|df_x\|$*

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

C^k Functions

Theorem 9 (Thm. 4.3)**.** *Suppose* $X \subseteq \mathbf{R}^n$ *is open and* $f : X \to \mathbf{R}^m$*. Then* f *is continuously differentiable on* X *if and only if* f *is* C^1*.*

Taylor's Theorem – Linear Terms

Theorem 10. *Suppose* $X \subseteq \mathbf{R}^n$ *is open and* $x \in X$*. If* $f : X \to \mathbf{R}^m$ *is differentiable, then*

$$f(x+h) = f(x) + Df(x)h + o(|h|) \text{ as } h \to 0$$

This is essentially a restatement of the definition of differentiability.

Taylor's Theorem – Linear Terms

Theorem 11 (Corollary of 4.4). *Suppose* $X \subseteq \mathbf{R}^n$ *is open and* $x \in X$. *If* $f : X \to \mathbf{R}^m$ *is* C^2, *then*

$$f(x+h) = f(x) + Df(x)h + O\left(|h|^2\right) \text{ as } h \to 0$$

Taylor's Theorem – Quadratic Terms

We treat each component of the function separately, so consider $f : X \to \mathbf{R}$, $X \subseteq \mathbf{R}^n$ an open set. Let

$$D^2 f(x) \;=\; \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \frac{\partial^2 f}{\partial x_2 \partial x_1}(x) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1}(x) \\ \frac{\partial^2 f}{\partial x_1 \partial x_2}(x) & \frac{\partial^2 f}{\partial x_2^2}(x) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_2}(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) & \cdots & \cdots & \frac{\partial^2 f}{\partial x_n^2}(x) \end{pmatrix}$$

$$\begin{aligned} f \in C^2 \;&\Rightarrow\; \frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \frac{\partial^2 f}{\partial x_j \partial x_i}(x) \\ &\Rightarrow\; D^2 f(x) \text{ is symmetric} \\ &\Rightarrow\; D^2 f(x) \text{ has eigenvectors that are an orthonormal basis} \\ &\qquad \text{and thus can be diagonalized} \end{aligned}$$

Taylor's Theorem – Quadratic Terms

Theorem 12 (Stronger Version of Thm. 4.4). *Let $X \subseteq \mathbf{R}^n$ be open, $f : X \to \mathbf{R}$, $f \in C^2(X)$, and $x \in X$. Then*

$$f(x+h) = f(x) + Df(x)h + \frac{1}{2}h^\top (D^2 f(x))h + o\left(|h|^2\right) \text{ as } h \to 0$$

If $f \in C^3$,

$$f(x+h) = f(x) + Df(x)h + \frac{1}{2}h^\top (D^2 f(x))h + O\left(|h|^3\right) \text{ as } h \to 0$$

$f : \mathbb{R}^n \to \mathbb{R}$ differentiable

Characterizing Critical Points

Definition 4. *We say f has a* saddle *at x if $Df(x) = 0$ but f has neither a local maximum nor a local minimum at x.*

f has a critical point at x if $Df(x) = 0$.

Characterizing Critical Points

Corollary 1. *Suppose* $X \subseteq \mathbf{R}^n$ *is open and* $x \in X$. *If* $f : X \to \mathbf{R}$ *is* C^2, *there is an orthonormal basis* $\{v_1, \ldots, v_n\}$ *and corresponding eigenvalues* $\lambda_1, \ldots, \lambda_n \in \mathbf{R}$ *of* $D^2 f(x)$ *such that*

$h = \sum \gamma_i v_i$

$$\begin{aligned} f(x+h) &= f(x + \gamma_1 v_1 + \cdots + \gamma_n v_n) \\ &= f(x) + \sum_{i=1}^{n} \left(Df(x) v_i\right) \gamma_i + \frac{1}{2} \sum_{i=1}^{n} \lambda_i \gamma_i^2 + o\left(|\gamma|^2\right) \end{aligned}$$

where $\gamma_i = h \cdot v_i$.

$= f(x) + Df(x)h + \frac{1}{2} h^T D^2 f(x) h + o(\|h\|^2)$

1. *If* $f \in C^3$, *we may strengthen* $o\left(|\gamma|^2\right)$ *to* $O\left(|\gamma|^3\right)$.

2. *If* f *has a local maximum or local minimum at* x, *then*

$$Df(x) = 0$$

3. *If* $Df(x) = 0$, *then*

- $\lambda_1, \ldots, \lambda_n > 0 \Rightarrow f$ *has a local minimum at* x
- $\lambda_1, \ldots, \lambda_n < 0 \Rightarrow f$ *has a local maximum at* x
- $\lambda_i < 0$ *for some* i, $\lambda_j > 0$ *for some* $j \Rightarrow f$ *has a saddle at* x
- $\lambda_1, \ldots, \lambda_n \geq 0$, $\lambda_i > 0$ *for some* $i \Rightarrow f$ *has a local minimum or a saddle at* x
- $\lambda_1, \ldots, \lambda_n \leq 0$, $\lambda_i < 0$ *for some* $i \Rightarrow f$ *has a local maximum or a saddle at* x
- $\lambda_1 = \cdots = \lambda_n = 0$ *gives no information.*

Proof. (Sketch) From our study of quadratic forms, we know the behavior of the quadratic terms is determined by the signs of the eigenvalues. If $\lambda_i = 0$ for some i, then we know that the quadratic form arising from the second partial derivatives is identically zero in the direction v_i, and the higher derivatives will determine the behavior of the function f in the direction v_i. For example, if $f(x) = x^3$, then $f'(0) = 0$, $f''(0) = 0$, but we know that f has a saddle at $x = 0$; however, if $f(x) = x^4$, then again $f'(0) = 0$ and $f''(0) = 0$ but f has a local (and global) minimum at $x = 0$. $\square$