Announcements

- PS 4 due Tuesday
- comments about exam Thursday in lecture
- exam: available 9 am 8/19
  due 9 am 8/20
  (Berkeley time)
Derivatives

**Definition 1.** Let $f : I \to \mathbb{R}$, where $I \subseteq \mathbb{R}$ 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 $a \in \mathbb{R}$. 

This is equivalent to \( \exists a \in \mathbb{R} \) such that

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

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

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

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

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

\( \iff T(h) = \lambda h \text{ for some } \lambda \in \mathbb{R} \)
Derivatives

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

\[
\lim_{h \to 0, h \in \mathbb{R}^n} \frac{\|f(x + h) - (f(x) + T_x(h))\|}{\|h\|} = 0 \tag{1}
\]

\( f \) 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 \( T_x \) 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 \( h \).
Big-Oh and little-oh

Notation:

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

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

\( \frac{\|y\|}{\|h\|^n} \) 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 $x$ $⇔$ $\exists T_x \in L(\mathbb{R}^n, \mathbb{R}^m)$ such that

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

$$g(h) = f(x + h) - (f(x) + T_x(h))$$
More Notation

Notation:

- $df_x$ is the linear transformation $T_x$
- $Df(x)$ is the matrix 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,

$f$ is differentiable at $x \iff E_f(h) = o(h)$ 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 $\mathbb{R}^n$. Look in direction $e_j$ (note that $|\gamma e_j| = |\gamma|$).

$$o(\gamma) = f(x + \gamma e_j) - \left(f(x) + T_x(\gamma e_j)\right) = f(x + \gamma e_j) - \left(f(x) + \begin{pmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mj} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} 0 \\ \vdots \end{pmatrix} + \begin{pmatrix} \gamma a_{1j} \\ \vdots \\ \gamma a_{mj} \end{pmatrix}\right)$$
For \( i = 1, \ldots, m \), let \( f^i \) denote the \( i^{th} \) component of the function \( f: \mathbb{R}^n \to \mathbb{R}^m \) where

\[
\begin{aligned}
f(x) &= (f^1(x), \ldots, f^m(x)) \\
f: \mathbb{R}^n &\to \mathbb{R}^m \\
\end{aligned}
\]

\[
\begin{aligned}
f^i(x + \gamma e_j) - (f^i(x) + \gamma a_{ij}) &= o(\gamma) \\
\text{so } a_{ij} &= \frac{\partial f^i}{\partial x_j}(x)
\end{aligned}
\]
Derivatives and Partial Derivatives

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

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

i.e. the Jacobian at $x$ is the matrix of partial derivatives at $x$. 
Derivatives and Partial Derivatives

Remark: If $f$ is differentiable at $x$, then all first-order partial derivatives $\frac{\partial f}{\partial x_j}$ 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 $x$. 

$$
\begin{cases}
\frac{x_1 x_2}{x_1^2 + x_2^2} & (x_1, x_2) \neq (0, 0) \\
0 & (x_1, x_2) = (0, 0)
\end{cases}
$$
Directional Derivatives

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

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

$$\Rightarrow f(x + \gamma u) - (f(x) + \gamma T_x(u)) = o(\gamma) \text{ as } \gamma \to 0 \quad \text{(Taylor Linear)}$$

$$\Rightarrow \lim_{\gamma \to 0} \frac{f(x + \gamma u) - f(x)}{\gamma} = T_x(u) = Df(x)u \quad \text{\small by } df_x(u)$$

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

$$Df(x)u \in \mathbb{R}^m$$
Chain Rule

**Theorem 3** (Thm. 3.5, Chain Rule). Let $X \subseteq \mathbb{R}^n$, $Y \subseteq \mathbb{R}^m$ be open, $f : X \to Y$, $g : Y \to \mathbb{R}^p$. 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

$$dF_{x_0} = dg_{f(x_0)} \circ df_{x_0}$$

*(composition of linear transformations)*

$$DF(x_0) = Dg(f(x_0))Df(x_0)$$

*(matrix multiplication)*

**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). Let \( a, b \in \mathbb{R} \). Suppose \( f : [a, b] \to \mathbb{R} \) 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. □
$f(x) = f(a) + \frac{f(b) - f(a)}{b - a} (x - a)$

$g(x) = f(x) - f(a) - \frac{f(b) - f(a)}{b - a} (x - a)$
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 \( f : \mathbb{R}^n \to \mathbb{R} \) is differentiable on an open set \( X \subseteq \mathbb{R}^n \), \( x, y \in X \) and \( \ell(x, y) \subseteq X \). Then there exists \( z \in \ell(x, y) \) such that*

\[
f(y) - f(x) = Df(z)(y - x)
\]

Consider \( \overline{f} : [0, 1] \to \mathbb{R} \) given by

\[
\overline{f}(\alpha) = f(\alpha x + (1-\alpha)y)
\]
\[ f = (f_1, \ldots, f_m), \quad f_i : \mathbb{R}^n \to \mathbb{R} \] \[ \] Notice that the statement is exactly the same as in the univariate case. For \( f : \mathbb{R}^n \to \mathbb{R}^m \), we can apply the Mean Value Theorem to each component, to obtain \( z_1, \ldots, z_m \in \ell(x, y) \) such that
\[ f^i(y) - f^i(x) = Df^i(z_i)(y - x) \]
However, we cannot find a single \( z \) which works for every component.

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

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

\[
\|f(y) - f(x)\| \leq \|df_z(y - x)\| = \|Df(z)(y - x)\| \\
\leq \|df_z\|\|y - x\|
\]
Mean Value Theorem

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

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

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

$$|df_z(1 - 0)| = |2\pi (-\sin 2\pi z, \cos 2\pi z)|$$

$$= 2\pi \sqrt{\sin^2 2\pi z + \cos^2 2\pi z}$$

$$= 2\pi$$

So $f(1) - f(0) \neq df_z(1 - 0) \forall z \in [0, 1]$. 
Taylor’s Theorem – $\mathbb{R}$

**Theorem 7** (Thm. 1.9, Taylor’s Theorem in $\mathbb{R}$). Let $f : I \to \mathbb{R}$ be $n$-times differentiable, where $I \subseteq \mathbb{R}$ 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

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

order error term or "remainder"
Motivation: Let

\[ T_n(h) = f(x) + \sum_{k=1}^{n} \frac{f^{(k)}(x)h^k}{k!} \]

\[ = f(x) + f'(x)h + \frac{f''(x)h^2}{2} + \cdots + \frac{f^{(n)}(x)h^n}{n!} \]

\[ T_n(0) = f(x) \]

\[ T'_n(h) = f'(x) + f''(x)h + \cdots + \frac{f^{(n)}(x)h^{n-1}}{(n-1)!} \]

\[ T'_n(0) = f'(x) \]

\[ T''_n(h) = f''(x) + \cdots + \frac{f^{(n)}(x)h^{n-2}}{(n-2)!} \]

\[ T''_n(0) = f''(x) \]

\[ \vdots \]

\[ T^{(n)}_n(0) = f^{(n)}(x) \]
so $T_n(h)$ is the unique $n^{th}$ degree polynomial such that

\[
\begin{align*}
T_n(0) &= f(x) \\
T_n'(0) &= f'(x) \\
& \vdots \\
T_n^{(n)}(0) &= f^{(n)}(x)
\end{align*}
\]
Taylor’s Theorem – R

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

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

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 \( X \subseteq \mathbb{R}^n \) be open. A function \( f : X \to \mathbb{R}^m \) 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 \( L(\mathbb{R}^n, \mathbb{R}^m) \), 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 $X \subseteq \mathbb{R}^n$ is open and $f : X \rightarrow \mathbb{R}^m$. Then $f$ is continuously differentiable on $X$ if and only if $f$ is $C^1$. 
Taylor’s Theorem – Linear Terms

**Theorem 10.** Suppose $X \subseteq \mathbb{R}^n$ is open and $x \in X$. If $f : X \to \mathbb{R}^m$ 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 $X \subseteq \mathbb{R}^n$ is open and $x \in X$. If $f : X \to \mathbb{R}^m$ is $C^2$, then

$$f(x + h) = f(x) + Df(x)h + O(|h|^2) \text{ as } h \to 0$$
Taylor’s Theorem – Quadratic Terms

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

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

$f \in C^2 \Rightarrow \frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \frac{\partial^2 f}{\partial x_j \partial x_i}(x)$

$\Rightarrow D^2 f(x)$ is symmetric

$\Rightarrow D^2 f(x)$ has eigenvectors that are an orthonormal basis and thus can be diagonalized
Taylor’s Theorem – Quadratic Terms

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

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

If $f \in C^3$,

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

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

$f$ has a critical point at $x$ if $Df(x) = 0$. \hfill 27
Characterizing Critical Points

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

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

where $\gamma_i = h \cdot v_i$.

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

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 \( x = 0 \). \( \square \)
\[ (-1)^n (\lambda - c_1) \cdots (\lambda - c_n) = \det (A - \lambda I) \]

\[ \prod_{i=1}^{n} (c_i - \lambda) = \det (A - \lambda I) \]

\[ V = \{ v_\lambda : \lambda \in \mathbb{R} \} \quad U = \{ u_\lambda : \lambda \in \mathbb{R} \} \]

\[ x \in X \Rightarrow \exists \{ u_\lambda \} \in U \quad \forall \phi \in F \]

\[ x = \sum_{i=1}^{n} a_i v_i \quad \text{s.t.} \quad x = \sum_{i=1}^{n} a_i v_i \]
\[ u = T(v), \quad u \neq 0 \]

\[ \Rightarrow T(\lambda u) = T^2(v) = T(v) = u \]

\[ \lambda = 1 \]

\[ x \in \ker(T), \quad x \neq 0, \quad T(x) = 0 = 0 \cdot x \Rightarrow x \text{ eigenvector} \]

\[ \lambda = 0 \]

\[ \text{Range } T = \mathbb{R}^r, \quad r \leq n \]

Choose \( \{u_1, \ldots, u_r\} \) basis for \( \text{Im } T \)

\[ \dim \ker T = n - r \]
\[ W = \{ x \in \mathbb{R}^3 : x_1 = x_2 = 0 \} = \{ x \in \mathbb{R}^3 : (0, 0, z) \} \]

\[ W \subseteq \mathbb{R}^3 \text{ vector subspace for some } z \in \mathbb{R} \]

\[ \dim W = 1 \]

\[ (0, 0, 1) \]

\[ W = U \]