MAT237Y1 – LEC5201

Multivariable Calculus

DIFFERENTIABILITY: A SUMMARY

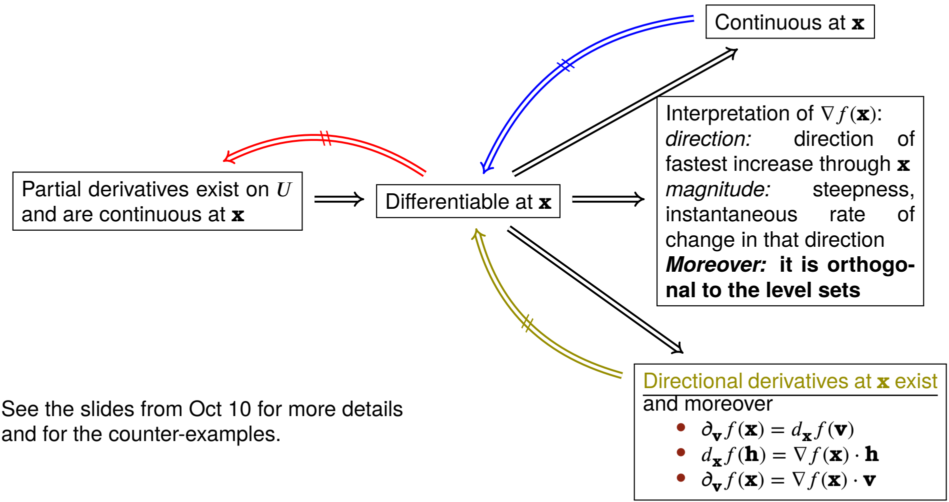
November 14th, 2019

Real-valued case – $U \subset \mathbb{R}^n$ open and $f : U \to \mathbb{R}$.

Name	Nature	Notation and definition
Directional derivative at $\mathbf{x} \in U$ along $\mathbf{v} \in \mathbb{R}^n$	Real number	$\partial_{\mathbf{v}} f(\mathbf{x}) = \lim_{t\to 0} \frac{f(\mathbf{x}+t\mathbf{v}) - f(\mathbf{x})}{t}$
i-th partial derivative at $\mathbf{x} \in U$	Real number	$\frac{\partial f}{\partial x_i}(\mathbf{x}) = \partial_{\mathbf{e}_i} f(\mathbf{x})$
Gradient at $\mathbf{x} \in U$	Vector in $\mathbb{R}^n$	$\nabla f(\mathbf{x}) = \left(\frac{\partial f}{\partial x_1}(\mathbf{x}), \dots, \frac{\partial f}{\partial x_n}(\mathbf{x})\right)$
Differential (or total derivative) at $\mathbf{x} \in U$ *“f is differentiable at $\mathbf{x}$”*	Linear function $d_{\mathbf{x}} f : \mathbb{R}^n \to \mathbb{R}$	$f(\mathbf{x}+\mathbf{h}) = f(\mathbf{x}) + d_{\mathbf{x}} f(\mathbf{h}) + E(\mathbf{h})$ with $\lim_{\mathbf{h}\to\mathbf{0}} \frac{E(\mathbf{h})}{\lVert\mathbf{h}\rVert} = 0$

See the slides from Oct 10 for the geometric intuitions about these objects.

Real-valued case – $f : U \to \mathbb{R}$, $U \subset \mathbb{R}^n$ open, $\mathbf{x} \in U$



Continuous at $\mathbf{x}$

Partial derivatives exist on U and are continuous at $\mathbf{x}$

Differentiable at $\mathbf{x}$

Interpretation of $\nabla f(\mathbf{x})$:
direction: direction of fastest increase through $\mathbf{x}$
magnitude: steepness, instantaneous rate of change in that direction
***Moreover:* it is orthogonal to the level sets**

Directional derivatives at $\mathbf{x}$ exist
and moreover

- $\partial_{\mathbf{v}} f(\mathbf{x}) = d_{\mathbf{x}} f(\mathbf{v})$
- $d_{\mathbf{x}} f(\mathbf{h}) = \nabla f(\mathbf{x}) \cdot \mathbf{h}$
- $\partial_{\mathbf{v}} f(\mathbf{x}) = \nabla f(\mathbf{x}) \cdot \mathbf{v}$

See the slides from Oct 10 for more details and for the counter-examples.

Partial derivatives at $\mathbf{x}$ exist $\not\Longrightarrow$ Directional derivatives at $\mathbf{x}$ exist

All the directional derivatives at $\mathbf{x}$ exist $\not\Longrightarrow$ Continuous at $\mathbf{x}$

Vector-valued case – $U \subset \mathbb{R}^n$ open and $\mathbf{f} : U \to \mathbb{R}^k$.

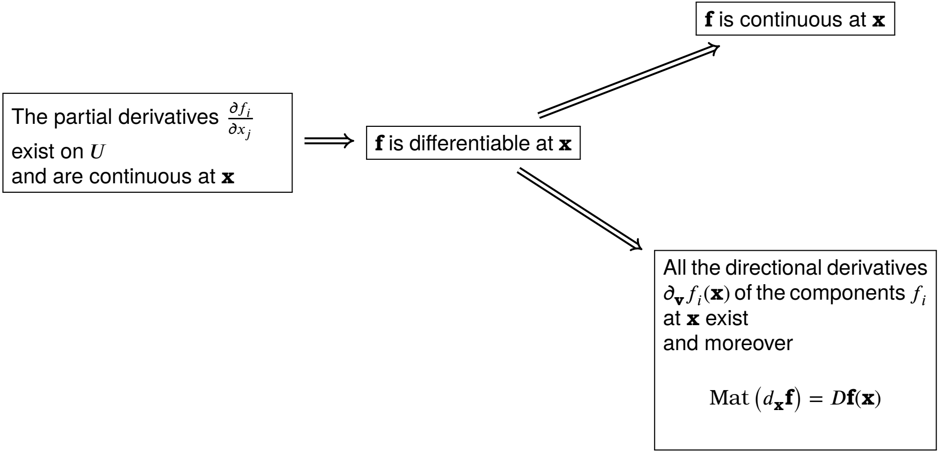
We denote by f_i the components of **f**, i.e. $\mathbf{f} = (f_1, \dots, f_k) : U \to \mathbb{R}^k$

Name	Nature	Notation and definition
Differential (or total derivative) at $\mathbf{x} \in U$ “**f** *is differentiable at* $\mathbf{x}$”	Linear function $d_\mathbf{x}\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^k$	$\mathbf{f}(\mathbf{x}+\mathbf{h}) = \mathbf{f}(\mathbf{x}) + d_\mathbf{x}\mathbf{f}(\mathbf{h}) + \mathbf{E}(\mathbf{h})$ with $\lim_{\mathbf{h}\to\mathbf{0}} \frac{\mathbf{E}(\mathbf{h})}{\lVert\mathbf{h}\rVert} = \mathbf{0}$
Jacobian matrix of **f** at $\mathbf{x} \in U$	$(k \times n)$-matrix	$D\mathbf{f}(\mathbf{x}) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{x}) & \cdots & \frac{\partial f_1}{\partial x_n}(\mathbf{x}) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_k}{\partial x_1}(\mathbf{x}) & \cdots & \frac{\partial f_k}{\partial x_n}(\mathbf{x}) \end{pmatrix}$

Vector-valued case – $\mathbf{f} : U \to \mathbb{R}^k$, $U \subset \mathbb{R}^n$ open, $\mathbf{x} \in U$

We denote by $f_i : U \to \mathbb{R}$ the components of $\mathbf{f}$, i.e. $\mathbf{f} = (f_1, \dots, f_k) : U \to \mathbb{R}^k$

We proved that $\mathbf{f}$ is differentiable at $\mathbf{x}$ if and only if its components f_i are too.
It allowed us to use the results from the real-valued case to prove the following theorems:



The partial derivatives $\frac{\partial f_i}{\partial x_j}$ exist on U and are continuous at $\mathbf{x}$

$\Longrightarrow$

$\mathbf{f}$ is differentiable at $\mathbf{x}$

$\mathbf{f}$ is continuous at $\mathbf{x}$

All the directional derivatives $\partial_{\mathbf{v}} f_i(\mathbf{x})$ of the components f_i at $\mathbf{x}$ exist and moreover

$$\operatorname{Mat}\left(d_{\mathbf{x}}\mathbf{f}\right) = D\mathbf{f}(\mathbf{x})$$

The Chain Rule

Let $U \subset \mathbb{R}^n$ open, $\mathbf{f} : \begin{array}{ccc} U & \to & \mathbb{R}^l \\ \mathbf{x} & \mapsto & \mathbf{f}(\mathbf{x}) \end{array}$, $V \subset \mathbb{R}^l$ open, $\mathbf{g} : \begin{array}{ccc} V & \to & \mathbb{R}^k \\ \mathbf{y} & \mapsto & \mathbf{g}(\mathbf{y}) \end{array}$.
Assume that $\mathbf{f}(U) \subset V$ so that $\mathbf{g} \circ \mathbf{f} : U \to \mathbb{R}^k$ is well-defined.

Let $\mathbf{x} \in U$.
If $\mathbf{f}$ is differentiable at $\mathbf{x}$ and $\mathbf{g}$ is differentiable at $\mathbf{f}(\mathbf{x})$ then $\mathbf{g} \circ \mathbf{f}$ is differentiable at $\mathbf{x}$.

- Chain rule formula for the differentials:

$$d_{\mathbf{x}}(\mathbf{g} \circ \mathbf{f}) = \left(d_{\mathbf{f}(\mathbf{x})}\mathbf{g}\right) \circ \left(d_{\mathbf{x}}\mathbf{f}\right)$$

- Chain rule formula for the Jacobian matrices:

$$D(\mathbf{g} \circ \mathbf{f})(\mathbf{x}) = D(\mathbf{g})(\mathbf{f}(\mathbf{x})) \cdot D(\mathbf{f})(\mathbf{x})$$

- Chain rule for the partial derivatives:

$$\frac{\partial(g_i \circ \mathbf{f})}{\partial x_j}(\mathbf{x}) = \sum_{\alpha=1}^{l} \frac{\partial g_i}{\partial y_\alpha}(\mathbf{f}(\mathbf{x})) \cdot \frac{\partial f_\alpha}{\partial x_j}(\mathbf{x})$$

The Chain Rule

Let $U \subset \mathbb{R}^n$ open, $\mathbf{f} : \begin{array}{ccc} U & \to & \mathbb{R}^l \\ \mathbf{x} & \mapsto & \mathbf{f}(\mathbf{x}) \end{array}$, $V \subset \mathbb{R}^l$ open, $\mathbf{g} : \begin{array}{ccc} V & \to & \mathbb{R}^k \\ \mathbf{y} & \mapsto & \mathbf{g}(\mathbf{y}) \end{array}$.
Assume that $\mathbf{f}(U) \subset V$ so that $\mathbf{g} \circ \mathbf{f} : U \to \mathbb{R}^k$ is well-defined.

Let $\mathbf{x} \in U$.
If $\mathbf{f}$ is differentiable at $\mathbf{x}$ and $\mathbf{g}$ is differentiable at $\mathbf{f}(\mathbf{x})$ then $\mathbf{g} \circ \mathbf{f}$ is differentiable at $\mathbf{x}$.

- Chain rule formula for the differentials:

$$d_{\mathbf{x}}(\mathbf{g} \circ \mathbf{f}) = \left(d_{\mathbf{f}(\mathbf{x})}\mathbf{g}\right) \circ \left(d_{\mathbf{x}}\mathbf{f}\right)$$

- Chain rule formula for the Jacobian matrices:

$$D(\mathbf{g} \circ \mathbf{f})(\mathbf{x}) = D(\mathbf{g})(\mathbf{f}(\mathbf{x})) \cdot D(\mathbf{f})(\mathbf{x})$$

- Chain rule for the partial derivatives:

$$\frac{\partial (g_i \circ \mathbf{f})}{\partial x_j}(\mathbf{x}) = \sum_{\alpha=1}^{l} \frac{\partial g_i}{\partial y_\alpha}(\mathbf{f}(\mathbf{x})) \cdot \frac{\partial f_\alpha}{\partial x_j}(\mathbf{x})$$

We derive the second formula from the first one by noticing that $D(\mathbf{f})(\mathbf{x}) = \mathrm{Mat}(d_{\mathbf{x}}\mathbf{f})$. And we derive the third formula from the second one by looking at the (i, j)-component of the matrices (the RHS is just the matrix multiplication formula).

The Chain Rule

Let $U \subset \mathbb{R}^n$ open, $\mathbf{f} : \begin{array}{ccc} U & \to & \mathbb{R}^l \\ \mathbf{x} & \mapsto & \mathbf{f}(\mathbf{x}) \end{array}$, $V \subset \mathbb{R}^l$ open, $\mathbf{g} : \begin{array}{ccc} V & \to & \mathbb{R}^k \\ \mathbf{y} & \mapsto & \mathbf{g}(\mathbf{y}) \end{array}$.
Assume that $\mathbf{f}(U) \subset V$ so that $\mathbf{g} \circ \mathbf{f} : U \to \mathbb{R}^k$ is well-defined.

Let $\mathbf{x} \in U$.
If $\mathbf{f}$ is differentiable at $\mathbf{x}$ and $\mathbf{g}$ is differentiable at $\mathbf{f}(\mathbf{x})$ then $\mathbf{g} \circ \mathbf{f}$ is differentiable at $\mathbf{x}$.

- Chain rule formula for the differentials:

$$d_{\mathbf{x}}(\mathbf{g} \circ \mathbf{f}) = \left(d_{\mathbf{f}(\mathbf{x})}\mathbf{g}\right) \circ \left(d_{\mathbf{x}}\mathbf{f}\right)$$

- Chain rule formula for the Jacobian matrices:

$$D(\mathbf{g} \circ \mathbf{f})(\mathbf{x}) = D(\mathbf{g})(\mathbf{f}(\mathbf{x})) \cdot D(\mathbf{f})(\mathbf{x})$$

- Chain rule for the partial derivatives:

$$\frac{\partial (g_i \circ \mathbf{f})}{\partial x_j}(\mathbf{x}) = \sum_{\alpha=1}^{l} \frac{\partial g_i}{\partial y_\alpha}(\mathbf{f}(\mathbf{x})) \cdot \frac{\partial f_\alpha}{\partial x_j}(\mathbf{x})$$

The last formula may seem *difficult* but after using it several times you'll notice that it is easy to use in practice, it generalizes the chain rule from MAT135/137/157 in a natural way.

Beware!

Your worst enemy in calculus is going to be the notation!

- There are as many notations as people: if you pick two different textbooks/mathematicians randomly, they probably don't use the same notations for the directional derivatives, the partial derivatives, the differentials, the Jacobian matrices…
  For instance, below are some notations more or less commonly used for the partial derivative of $f : \mathbb{R}^2 \to \mathbb{R}$ with respect to the first variable (i.e. the directional derivative along $\mathbf{e}_1$):

$$\frac{\partial f}{\partial x},\ \partial_x f,\ \partial_{\mathbf{e}_1} f,\ \partial_1 f,\ f_x,\ f'_x,\ D_x f,\ D_{\mathbf{e}_1} f,\ D_1 f,\ D^1 f,\ D^{\mathbf{e}_1} f, \dots$$

- The notations might be confusing at first: be sure that you understand what you are reading and/or writing! Rely on the context to avoid any confusion!
  For instance, given a function $f : \mathbb{R}^2 \to \mathbb{R}$, $\frac{\partial f}{\partial x}$ simply denotes the derivative with respect to the first variable (i.e. the directional derivative along $\mathbf{e}_1$), do not try to interpret the x in the denominator ∂x, that's just a notation.
  Therefore, if you see $\frac{\partial f}{\partial x}(x^2, xyz)$, it means that you **first** compute the partial derivative and **then** that you evaluate it at (x^2, xyz).
  You should **not** compute $f(x^2, xyz)$ and then take the derivative with respect to x.

The MVT

Theorem. *The MVT (one-variable case)*
Let $f : [a,b] \to \mathbb{R}$ continuous on $[a,b]$ and differentiable on (a,b).
Then there exists $c \in (a,b)$ such that $f(b) - f(a) = f'(c)(b-a)$.

Theorem. *The MVT (multivariable case)*
$U \subset \mathbb{R}^n$ open, $f : U \to \mathbb{R}$ differentiable on U.
Let $\mathbf{a}, \mathbf{b} \in U$ such that $L_{\mathbf{a},\mathbf{b}} = \{(1-t)\mathbf{a} + t\mathbf{b} \ : \ t \in [0,1]\} \subset U$.
Then there exists $\mathbf{c} \in L_{\mathbf{a},\mathbf{b}}$ such that $f(\mathbf{b}) - f(\mathbf{a}) = \nabla f(\mathbf{c}) \cdot (\mathbf{b} - \mathbf{a})$

Corollary. $U \subset \mathbb{R}^n$ open and **convex**, $f : U \to \mathbb{R}$ differentiable on U.
If there exists $M > 0$ such that $\forall \mathbf{x} \in U$, $\|\nabla f(\mathbf{x})\| \leq M$ then
$\forall \mathbf{a}, \mathbf{b} \in U$, $|f(\mathbf{b}) - f(\mathbf{a})| \leq M \|\mathbf{b} - \mathbf{a}\|$

Corollary. $U \subset \mathbb{R}^n$ open and **path-connected**, $f : U \to \mathbb{R}$ differentiable.
If $\forall \mathbf{x} \in U$, $\nabla f(\mathbf{x}) = \mathbf{0}$ then f is constant on U.

Higher-order partial derivatives – 1

Notation. $U \subset \mathbb{R}^n$ open, $f : U \to \mathbb{R}$, $\mathbf{a} \in U$.
Assume that $\frac{\partial f}{\partial x_i}$ exists in a small ball centered at $\mathbf{a}$ and that it admits a directional derivative at $\mathbf{a}$ along $\mathbf{e}_j$, then we denote the *second-order partial derivatives* by

$$\frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{a}) := \frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right)(\mathbf{a})$$

Be careful, we first take the partial derivative w.r.t. x_i and then w.r.t. x_j.

Theorem. *Clairaut's theorem*.
$U \subset \mathbb{R}^n$ open, $f : U \to \mathbb{R}$ of class C^2, $\mathbf{a} \in U$. Then

$$\frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{a}) = \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a})$$

"If the second-order partial derivatives are continuous then the order doesn't matter."

Example. The C^2 assumption is crucial here!
Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by $f(x, y) = \begin{cases} xy\frac{x^2-y^2}{x^2+y^2} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{otherwise} \end{cases}$

Then $\frac{\partial^2 f}{\partial x \partial y}(0,0) = 1 \neq -1 = \frac{\partial^2 f}{\partial y \partial x}(0,0)$.

Higher-order partial derivatives – 2

Notation. Similarly we set $\dfrac{\partial^k f}{\partial x_{i_k} \partial x_{i_{k-1}} \cdots \partial x_{i_1}}(\mathbf{a}) = \dfrac{\partial}{\partial x_{i_k}} \left(\dfrac{\partial}{\partial x_{i_{k-1}}} \left(\cdots \left(\dfrac{\partial f}{\partial x_{i_1}} \right) \cdots \right) \right) (\mathbf{a})$.

Another possible notation is $\partial_{x_{i_k}} \partial_{x_{i_{k-1}}} \cdots \partial_{x_{i_1}} f(\mathbf{a})$.

Be careful we read from right to left (that's a composition): first we differentiate w.r.t. x_{i_1}, then w.r.t. x_{i_2}, and so on…

Definition. $U \subset \mathbb{R}^n$ open, $f : U \to \mathbb{R}$.
We say that f is of class C^k if all its partial derivatives of order less than or equal to k exist and are continuous *(don't forget the continuity assumption!)*.

Comment. C^0 means *continuous*.
Comment. C^1 is read *continuously differentiable*.

Theorem. C^k functions are closed under elementary operations.

Theorem. $U \subset \mathbb{R}^n$ open, $f : U \to \mathbb{R}$ of class C^k, $\mathbf{a} \in U$.
Then $\dfrac{\partial^k f}{\partial x_{i_k} \partial x_{i_{k-1}} \cdots \partial x_{i_1}}(\mathbf{a})$ doesn't depend on the order of the $i_1, \ldots, i_k$.

Taylor theorem – The one-variable case

Definition. $I \subset \mathbb{R}$ interval, $f : I \to \mathbb{R}$, $a \in I$, assume that f is k-times differentiable at a then we define the *k-th order Taylor polynormal of f at a* by

$$P_{a,k}(h) := \sum_{j=0}^{k} \frac{f^{(j)}(a)}{j!} h^j = f(a) + f'(a)h + \frac{f''(a)}{2}h^2 + \cdots + \frac{f^{(k)}(a)}{k!}h^k$$

Comment. $P_{a,k}$ is the unique polynomial of degree at most k such that
$P_{a,k}(0) = f(a)$, $P'_{a,k}(0) = f'(a)$, $P''_{a,k}(0) = f''(a)$, $\ldots$, $P^{(k)}_{a,k}(0) = f^{(k)}(a)$

Theorem. *Taylor's theorem* or *Taylor–Young's theorem*
$I \subset \mathbb{R}$ interval, $f : I \to \mathbb{R}$ of class C^{k-1} on I, and $a \in I$.
If $f^{(k)}(a)$ exists then $f(a+h) = P_{a,k}(h) + E(h)$ where $\displaystyle\lim_{h\to 0} \frac{E(h)}{h^k} = 0$.

Theorem. *Taylor–Lagrange's theorem*
Let $I \subset \mathbb{R}$ interval, $f : I \to \mathbb{R}$ be $(k+1)$-times differentiable on I, and $a \in I$.
Let $h \in \mathbb{R} \setminus \{0\}$ such that $\begin{cases} [a, a+h] \subset I \text{ if } h > 0 \text{ or} \\ [a+h, a] \subset I \text{ if } h < 0 \end{cases}$ then $\begin{cases} \exists \xi \in (a, a+h) \text{ if } h > 0 \text{ or} \\ \exists \xi \in (a+h, a) \text{ if } h < 0 \end{cases}$
such that $f(a+h) = P_{a,k}(h) + \dfrac{f^{(k+1)}(\xi)}{(k+1)!} h^{k+1}$.

Taylor theorem – The multi-variable case – 1

Theorem. *Taylor's theorem at order* 1
$U \subset \mathbb{R}^n$ open, $f : U \to \mathbb{R}$ differentiable at $\mathbf{a} \in U$.
Then $f(\mathbf{a}+\mathbf{h}) = f(\mathbf{a}) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{a})h_i + E(\mathbf{h})$ where $\lim_{\mathbf{h}\to 0} \frac{E(\mathbf{h})}{\|\mathbf{h}\|} = 0$.

Theorem. *Taylor–Lagrange's theorem at order* 2
$U \subset \mathbb{R}^n$ open, $f : U \to \mathbb{R}$ of class C^2, $\mathbf{a} \in U$, $\mathbf{h} \in \mathbb{R}^n$.
Assume that $\forall t \in [0,1]$, $\mathbf{a} + t\mathbf{h} \in U$ then there exists $\theta \in (0,1)$ such that

$$f(\mathbf{a}+\mathbf{h}) = f(\mathbf{a}) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{a})h_i + \frac{1}{2}\sum_{j=1}^{n}\sum_{i=1}^{n} \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{a}+\theta\mathbf{h})h_i h_j.$$

Theorem. *Taylor's theorem at order* 2
$U \subset \mathbb{R}^n$ open, $f : U \to \mathbb{R}$ of class C^2, $\mathbf{a} \in U$.
Then $f(\mathbf{a}+\mathbf{h}) = f(\mathbf{a}) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{a})h_i + \frac{1}{2}\sum_{j=1}^{n}\sum_{i=1}^{n} \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{a})h_i h_j + E(\mathbf{h})$ where $\lim_{\mathbf{h}\to 0} \frac{E(\mathbf{h})}{\|\mathbf{h}\|^2} = 0$.

Taylor theorem – The multi-variable case – 2

We define the *Hessian matrix of* f *at* $\mathbf{a}$ by

$$H_f(\mathbf{a}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(\mathbf{a}) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(\mathbf{a}) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(\mathbf{a}) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n}(\mathbf{a}) \end{pmatrix}$$

Notice that

$$\begin{aligned} \sum_{j=1}^{n} \sum_{i=1}^{n} \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{a}) h_i h_j &= \begin{pmatrix} h_1 & \cdots & h_n \end{pmatrix} \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(\mathbf{a}) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(\mathbf{a}) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(\mathbf{a}) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n}(\mathbf{a}) \end{pmatrix} \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix} \\ &= \mathbf{h} \cdot \left(H_f(\mathbf{a}) \mathbf{h} \right) \end{aligned}$$

Taylor theorem – The multi-variable case – 3

You do **NOT** need to know the following formula (see next slide)!

Theorem. *(Taylor's theorem)*
$U \subset \mathbb{R}^n$ open, $f : U \to \mathbb{R}$ of class C^k, $\mathbf{a} \in U$. Then

$$f(\mathbf{a}+\mathbf{h}) = \sum_{|\boldsymbol{\alpha}| \leq k} \frac{\partial^{\boldsymbol{\alpha}} f(\mathbf{a})}{\boldsymbol{\alpha}!}\mathbf{h}^{\boldsymbol{\alpha}} + E(\mathbf{h})$$

with $\lim\limits_{\mathbf{h}\to 0} \dfrac{E(\mathbf{h})}{\|\mathbf{h}\|^k} = 0$.

Where $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}^k_{\geq 0}$, $|\boldsymbol{\alpha}| = \alpha_1 + \dots + \alpha_n$, $\boldsymbol{\alpha}! = \alpha_1! \cdots \alpha_n!$,
$\mathbf{h}^{\boldsymbol{\alpha}} = h_1^{\alpha_1} \cdots h_n^{\alpha_n}$ and

$$\partial^{\boldsymbol{\alpha}} f(\mathbf{a}) = \frac{\partial^{|\boldsymbol{\alpha}|} f}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}}(\mathbf{a})$$

(Since the function is of class C^k the order doesn't matter and we gather together the differentiation w.r.t. a same variable)

Taylor theorem – The multi-variable case – 4 and last

In practice it would be very inefficient to use the formula from the previous slide since it involves too many computations of partial derivatives...
Instead you should rely on the Taylor series you already know.

Example.

$$\begin{aligned}\frac{e^{x-2y}}{1+x^2-y} &= \frac{e^{x-2y}}{1-(y-x^2)}\\ &= \left(1+(x-2y)+\frac{(x-2y)^2}{2}+\cdots\right)\left(1+(y-x^2)+(y-x^2)^2+\cdots\right)\\ &= 1+x-y-\frac{x^2}{2}-xy+y^2+E(x,y)\end{aligned}$$

with $\displaystyle\lim_{(x,y)\to(0,0)}\frac{E(x,y)}{\|(x,y)\|^2}=0$.

Hence $P_{\mathbf{0},2}(x,y)=1+x-y-\frac{x^2}{2}-xy+y^2$.

Some power series from MAT135/137/157

① $\forall x \in \mathbb{R},\ e^x = \sum_{n=0}^{+\infty} \frac{x^n}{n!}$ (recall that $0! = 1$)

② $\forall x \in \mathbb{R},\ \cos(x) = \sum_{n=0}^{+\infty} \frac{(-1)^n}{(2n)!} x^{2n}$ and $\sin(x) = \sum_{n=0}^{+\infty} \frac{(-1)^n}{(2n+1)!} x^{2n+1}$

③ $\forall x \in (-1, 1],\ \ln(1+x) = \sum_{n=1}^{+\infty} \frac{(-1)^{n+1}}{n} x^n$

④ $\forall x \in (-1, 1),\ \frac{1}{1-x} = \sum_{n=0}^{+\infty} x^n$

⑤ $\forall x \in (-1, 1),\ (1+x)^\alpha = 1 + \sum_{n=1}^{+\infty} \frac{\alpha(\alpha-1)\cdots(\alpha-n+1)}{n!} x^n$

(The last one holds for $x \in \mathbb{R}$ when $\alpha \in \mathbb{N}$.)

Keep in mind that power series behave well with respect to the usual operations: use them to reduce to the above results.

Local extrema & critical points – 1

Local extremum

Let $U \subset \mathbb{R}^n$ be open, $f : U \to \mathbb{R}$ and $\mathbf{a} \in U$.
We say that $\mathbf{a}$ is a *local min* (resp. *local max*) of f if

$$\exists r > 0, \forall \mathbf{x} \in U, \|\mathbf{x} - \mathbf{a}\| < r \implies f(\mathbf{a}) \leq f(\mathbf{x})$$

$$\left(\text{resp. } \exists r > 0, \forall \mathbf{x} \in U, \|\mathbf{x} - \mathbf{a}\| < r \implies f(\mathbf{a}) \geq f(\mathbf{x})\right)$$

Critical point

Let $U \subset \mathbb{R}^n$ be open, $f : U \to \mathbb{R}$ be differentiable and $\mathbf{a} \in U$.
We say that $\mathbf{a}$ is a *critical point* of f if $\nabla f(\mathbf{a}) = \mathbf{0}$.

Theorem: first derivative test

Let $U \subset \mathbb{R}^n$ be open, $f : U \to \mathbb{R}$ be differentiable and $\mathbf{a} \in U$.
If $\mathbf{a}$ is a local extremum of f then it is a critical point of f.

Hence the local extrema of f are among the critical points of f.

Local extrema & critical points – 2

Theorem: second derivative test

Let $U \subset \mathbb{R}^n$ be open, $f : U \to \mathbb{R}$ be of class C^2 and $\mathbf{a} \in U$ be a critical point.

- If $H_f(\mathbf{a})$ is positive definite then $\mathbf{a}$ is a local min of f.
- If $H_f(\mathbf{a})$ is negative definite then $\mathbf{a}$ is a local max of f.
- If $H_f(\mathbf{a})$ is indefinite then $\mathbf{a}$ is a *saddle point* of f.

▲ Indeed

$$f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) = \frac{1}{2}\mathbf{h} \cdot \left(H_f(\mathbf{a})\mathbf{h}\right) + E(\mathbf{h})$$

where $\lim_{\mathbf{h}\to 0} \frac{E(\mathbf{h})}{\|\mathbf{h}\|^2} = 0$. ■

In all the remaining cases, the Hessian matrix is not enough to conclude about the nature of $\mathbf{a}$.

Example: The Hessian matrices of $f(x,y) = x^2$ and $g(x,y) = x^3$ at $\mathbf{0}$ are both non-negative definite but f has a local min at $\mathbf{0}$ whereas g has no local extremum at $\mathbf{0}$.

Local extrema & critical points – 3

Recall that for $A \in M_{n,n}(\mathbb{R})$ symmetric

A positive definite $\Leftrightarrow$ $\forall \mathbf{h} \in \mathbb{R}^n \setminus \{\mathbf{0}\},\ \mathbf{h}^t A \mathbf{h} > 0$
$\Leftrightarrow$ All the eigenvalues of A are > 0
$\Leftrightarrow$ There exists $\lambda > 0$, $\forall \mathbf{h} \in \mathbb{R}^n$, $\mathbf{h}^t A \mathbf{h} \geq \lambda \|\mathbf{h}\|^2$

A negative definite $\Leftrightarrow$ $\forall \mathbf{h} \in \mathbb{R}^n \setminus \{\mathbf{0}\},\ \mathbf{h}^t A \mathbf{h} < 0$
$\Leftrightarrow$ All the eigenvalues of A are < 0
$\Leftrightarrow$ There exists $\lambda < 0$, $\forall \mathbf{h} \in \mathbb{R}^n$, $\mathbf{h}^t A \mathbf{h} \leq \lambda \|\mathbf{h}\|^2$

A indefinite $\Leftrightarrow$ $\exists \mathbf{h}, \mathbf{k} \in \mathbb{R}^n$, $\mathbf{h}^t A \mathbf{h} < 0 < \mathbf{k}^t A \mathbf{k}$
$\Leftrightarrow$ A has a positive eigenvalue and a negative eigenvalue

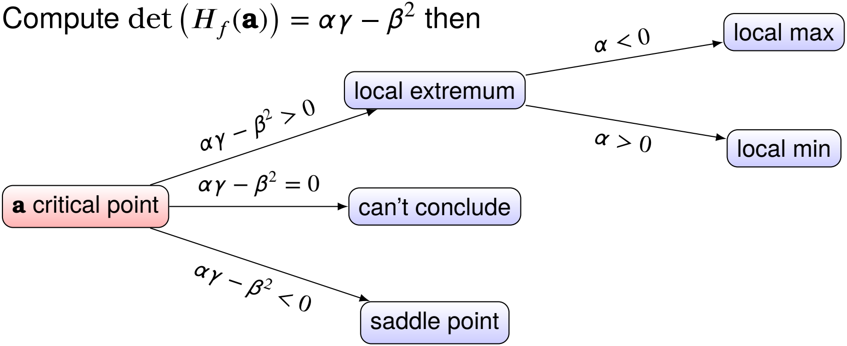
Local extrema & critical points – 4 – two-variable case

Let $U \subset \mathbb{R}^2$ be open, $f : U \to \mathbb{R}$ be of class C^2 and $\mathbf{a} \in U$ be a critical point (i.e. $\frac{\partial f}{\partial x}(\mathbf{a}) = \frac{\partial f}{\partial y}(\mathbf{a}) = 0$).

By Clairaut's theorem $H_f(\mathbf{a}) = \begin{pmatrix} \alpha & \beta \\ \beta & \gamma \end{pmatrix}$ is symmetric where

$$\alpha = \frac{\partial^2 f}{\partial x^2}(\mathbf{a}) \qquad \beta = \frac{\partial^2 f}{\partial x \partial y}(\mathbf{a}) = \frac{\partial^2 f}{\partial y \partial x}(\mathbf{a}) \qquad \gamma = \frac{\partial^2 f}{\partial y^2}(\mathbf{a})$$

Compute $\det\left(H_f(\mathbf{a})\right) = \alpha\gamma - \beta^2$ then



a critical point

$\alpha\gamma - \beta^2 > 0$

local extremum

$\alpha < 0$

local max

$\alpha > 0$

local min

$\alpha\gamma - \beta^2 = 0$

can't conclude

$\alpha\gamma - \beta^2 < 0$

saddle point