CALCULUS

PART I : DIFFERENTIAL CALCULUS

LECTURE 1 : DIFFERENTIAL CALCULUS FOR FUNCTIONS OF A SINGLE VARIABLE. TAYLOR'S FORMULA.

1. WEIERSTRASS THEOREM :

$$\left.\begin{array}{l} f:[a,b]\to\mathbb{R} \\ f \text{ continuous} \end{array}\right\} \Rightarrow f \text{ is bounded and attains its minimum } (m) \text{ and maximum } (M) \text{ in the sense that : } \operatorname{Im} f = [m,M]$$

$\operatorname{Im} f = f([a,b])$ – different notations

$f([a,b]) = \{ y\in\mathbb{R} : \exists x\in[a,b] \text{ such that } y = f(x)\}$

Remark : $\forall x\in[a,b]\ \exists y\in[m,M]$ such that $y=f(x)$

2. DIFFERENTIABILITY DEFINITION :

$f:(a,b)\to\mathbb{R}$ is differentiable at $x^*\in(a,b)$ if the limit

$$\lim_{h\to 0}\frac{f(x^*+h)-f(x^*)}{h}$$ exists and is finite.

If the limit exists but is infinite, then we say that f has derivative at x^* (but f is not differentiable at x^*).

3. LOCAL EXTREMA DEFINITION :

x^* is a local minimum of f, $f:(a,b)\to\mathbb{R}$ if $f(x^*)< f(x)$, $\forall x\in[x^*-\varepsilon, x^*+\varepsilon]$, $\varepsilon>0$

x^* is a local maximum of f, $f:(a,b)\to\mathbb{R}$ if $f(x^*)> f(x)$, $\forall x\in[x^*-\varepsilon, x^*+\varepsilon]$, $\varepsilon>0$

4. CRITICAL OR STATIONARY POINTS DEFINITION :

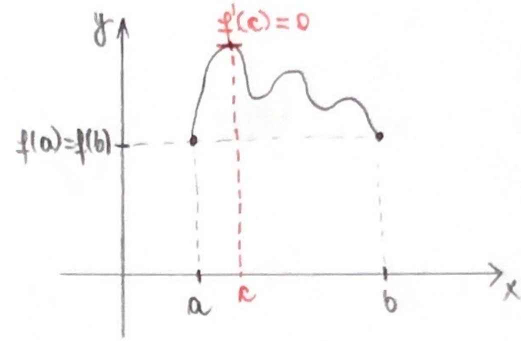
x^* is a critical point of f if $f'(x^*)=0$

Remark: local extrema and critical points do not coincide

5. FERMAT THEOREM :

$f:(a,b) \to \mathbb{R}$ has a local minimum / maximum at $x^* \in (a,b)$. If f is differentiable at x^* then $f'(x^*) = 0$.

6. ROLLE THEOREM :



Let $f:(a,b) \to \mathbb{R}$. If :

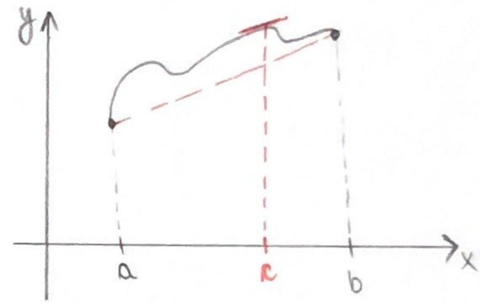
(1) f is continuous on $[a,b]$

(2) f is differentiable on (a,b)

(3) $f(a) = f(b)$

Then $\exists\, c \in (a,b)$ such that $f'(c) = 0$.

7. LAGRANGE THEOREM :



Let $f:(a,b) \to \mathbb{R}$. If :

(1) f is continuous on $[a,b]$

(2) f is differentiable on (a,b)

Then $\exists\, c \in (a,b)$ such that $f'(c) = \dfrac{f(b)-f(a)}{b-a}$

8. CAUCHY THEOREM :

Let $f, g : [a,b] \to \mathbb{R}$. If :

(1) f, g are continuous on $[a,b]$

(2) f, g are differentiable on (a,b)

(3) $g'(x) \neq 0,\ \forall x \in (a,b)$

Then $\exists\, c \in (a,b)$ such that : $\dfrac{f(b)-f(a)}{g(b)-g(a)} = \dfrac{f'(c)}{g'(c)}$

9. TAYLOR THEOREM :

$f:(a,b) \to \mathbb{R}$ is $(n+1)$ times differentiable, $x_0 \in (a,b)$. Then $\forall x \in (a,b)\ \exists\, c$ between x and x_0 such that :

$$T_n(x) = f(x_0) + \frac{f'(x_0)}{1!}(x-x_0) + \frac{f''(x_0)}{2!}(x-x_0)^2 + \dots + \frac{f^{(n)}(x_0)}{n!}(x-x_0)^n$$

$$R_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!}(x-x_0)^{n+1}$$

$$f(x) = T_n(x) + R_n(x)$$

for approximation $R_n(x) \to 0$.

LECTURE 2: CALCULUS FOR FUNCTIONS OF SEVERAL VARIABLES I. THE GEOMETRY OF $\mathbb{R}^d$, PARTIAL DERIVATIVES, THE GRADIENT.

1. THE GEOMETRY OF $\mathbb{R}^d$:

a) „+" addition:

$x = (x_1, x_2, \dots, x_d) \qquad y = (y_1, y_2, \dots, y_d)$

$x + y = (x_1 + y_1, x_2 + y_2, \dots, x_d + y_d)$

b) multiplication by a scalar:

$\lambda x = (\lambda x_1, \lambda x_2, \dots, \lambda x_d)$

c) the dot product (inner product):

$x \cdot y = x_1 y_1 + x_2 y_2 + \dots + x_d y_d$

($\langle x, y \rangle$ alternative notation)

d) norm of x (length):

$\|x\| = \sqrt{x \cdot x} = \sqrt{x_1^2 + x_2^2 + \dots + x_d^2}$

e) distance:

$dist(x, y) = \|x - y\|$

$N_1: \|x\| \geq 0$ and $\|x\| = 0 \Leftrightarrow x = 0_{\mathbb{R}^d}$

$N_2: \|\lambda x\| = |\lambda| \|x\|, \forall \lambda \in \mathbb{R}, x \in \mathbb{R}^d$

$N_3: \|x + y\| \leq \|x\| + \|y\|, \forall x, y \in \mathbb{R}^d$

f) open ball centered at x and of radius r:

$B_r(x) = \{ y \in \mathbb{R}^d : \|x - y\| < r \}, r > 0$. ! is a convex set

g) orthogonality in $\mathbb{R}^d$

$x \perp y \Leftrightarrow x \cdot y = 0$

h) segments in $\mathbb{R}^d$

$$\underset{\text{segment}}{[x, y]} = \{ \underbrace{(1-\alpha)x + \alpha y}_{\text{convex comb (or average)}} : \alpha \in [0,1] \in \mathbb{R} \} \subset \mathbb{R}^d$$

i) C convex set if $\forall x, y \in C$ we have $[x, y] \subset C$.

$\mathbb{R}^d$ can not be ordered completely.

2. PARTIAL DERIVATIVES DEFINITION :

$f : \mathbb{R}^d \to \mathbb{R}$, f has partial derivative with respect to x_k at a point $x = (x_1, \dots, x_d)$

if $$\lim_{h \to 0} \frac{f(x_1, x_2, \dots, x_{k-1}, x_k + h, x_{k+1}, \dots x_d) - f(x_1, \dots, x_{k-1}, x_k, x_{k+1}, \dots, x_d)}{h}$$

exits.

Notation : $\frac{\partial f}{\partial x_k}$

3. THE GRADIENT :

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}(x), \frac{\partial f}{\partial x_2}(x), \dots, \frac{\partial f}{\partial x_d}(x) \right), \ x \in \mathbb{R}^d$$

LECTURE 3: DIFFERENTIAL CALCULUS FOR FUNCTIONS OF SEVERAL VARIABLES II: DIFFERENTIABILITY AND PROPERTIES.

1. CONTINUITY DEFINITION:

$f: \mathbb{R}^d \to \mathbb{R}$ is continuous at $x \in \mathbb{R}^d$ if $\forall \varepsilon > 0$ $\exists \delta > 0$ such that $|f(x) - f(y)| < \varepsilon$ $\forall y \in \mathbb{R}^d$ with $\|x - y\| < \delta$

2. CHAIN RULE THEOREM:

$f: \mathbb{R}^d \to \mathbb{R}$ has continuous partial derivatives and $x_1, \dots, x_d : [a,b] \subset \mathbb{R} \to \mathbb{R}$ are all differentiable.
Then $F: [a,b] \to \mathbb{R}$, $F(t) = f(x_1(t), \dots, x_d(t))$ $(F = f \circ (x_1, \dots, x_d))$ is differentiable and $\frac{d}{dt} F(t) = \nabla f(x_1(t), \dots, x_d(t)) = \frac{d}{dt} x(t)$

3. LAGRANGE THEOREM ($d > 1$):

If $D \leq \mathbb{R}^d$ convex, $a, b \in D (a \neq b)$.
$f: D \to \mathbb{R}$ has continuous partial derivatives
then $\exists c \in (a,b)$ such that: $f(b) - f(a) = \nabla f(c) \cdot (b-a)$

PROOF:

Let $a, b \in D$ and $g: [0,1] \to \mathbb{R}$, $g(t) = f(a + t(b-a))$, $t \in [0,1]$.
Taking account of the Lagrange theorem for functions of one variable:
$\exists t_0 \in (0,1)$ such that $g(1) - g(0) = g'(t_0)$. (1)
If $a = (a_1, a_2, \dots, a_m)$, $b = (b_1, b_2, \dots, b_m)$ then:

$$g(t) = f(\underbrace{a_1 + t(b_1 - a_1)}_{u_1}, \dots, \underbrace{a_m + t(b_m - a_m)}_{u_m})$$

and

$$g'(t) = f'_{u_1}(a + t(b-a)) u_1'(t) + \dots + f'_{u_n}(a + t(b-a)) u_n'(t) =$$
$$= f'_{u_1}(a + t(b-a))(b_1 - a_1) + \dots + f'_{u_n}(a + t(b-a))(b_n - a_n) =$$
$$= \nabla f(a)(a + t(b-a))(b-a). \quad (2)$$

Taking $c = a + t_0(b-a)$ to (1) and (2) $\Rightarrow$ $f(b) - f(a) = \nabla f(c)(b-a)$

4. SCHWARZ THEOREM :

$f:\mathbb{R}^d \to \mathbb{R}$ admits continuous mixed second order partial derivatives (on a small ball around x) then :

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

5. HESSIAN MATRIX :

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

$$H_f = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(x)\right)_{i,j=\overline{1,d}} = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_d} \\ \vdots & & & \\ \frac{\partial^2 f}{\partial x_d \partial x_1} & \dots & & \frac{\partial^2 f}{\partial x_d \partial x_d} \end{pmatrix}$$

6. LINEAR FUNCTIONS :

$T:\mathbb{R}^d \to \mathbb{R}$ is called linear if :

$$\left.\begin{array}{l} i)\ T(x+y) = T(x)+T(y) \quad \forall x,y \in \mathbb{R}^d \\ ii)\ T(\alpha x) = \alpha T(x), \ \forall x \in \mathbb{R}^d, \forall \alpha \in \mathbb{R} \end{array}\right\} \Rightarrow iii)\ T(\alpha x + \beta y) = \alpha T(x) + \beta T(y)$$

THEOREM :

$\forall T:\mathbb{R}^d \to \mathbb{R}$ linear function $\Rightarrow$ there exist unique $a_\kappa \in \mathbb{R}^d$ such that $T(x) = a_\kappa \cdot x$

EXAMPLE :

$a = (a_1, \dots, a_d) \in \mathbb{R}^d$

$T_a(x) = a \cdot x \to$ linear function

7. QUADRATIC FUNCTIONS :

$Q:\mathbb{R}^d \to \mathbb{R}$ is called quadratic if :

$$Q(x) = \sum_{i,j=1}^{d} a_{ij} x_i x_j, \quad a_{ij} = a_{ji}, \ i,j = \overline{1,d}$$

$$A = (a_{ij})_{i,j=\overline{1,d}} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1d} \\ \vdots & & & \\ a_{d1} & \dots & & a_{dd} \end{pmatrix}$$ – the matrix of the quadratic function

EXAMPLE :

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \Rightarrow Q(x_1,x_2) = x_1^2 + x_1x_2 + x_2x_1 + x_2^2 = x_1^2 + 2x_1x_2 + x_2^2$$

LECTURE 4: OPTIMIZATION FOR FUNCTIONS OF SEVERAL VARIABLES I.

1. FERMAT THEOREM:

$f : \mathbb{R}^d \to \mathbb{R}$ is Fréchet differentiable in $x^* \in \mathbb{R}^d$.

If x^* is a local minimum / maximum then $\nabla f(x^*) = 0$.

A quadratic function $Q : \mathbb{R}^m \to \mathbb{R}$ (with matrix $A = (a_{ij})$) is:

- positive definite if $Q(x) > 0 \quad \forall x \in \mathbb{R}^d \setminus \{0_{\mathbb{R}^d}\}$
- negative definite if $Q(x) < 0 \quad \forall x \in \mathbb{R}^d \setminus \{0_{\mathbb{R}^d}\}$
- indefinite if $Q(x_1) > 0$, $Q(x_2) < 0$

2. SYLVESTER THEOREM: (crit. for positive / negative definite)

If $A = (a_{ij})$ is the matrix of Q.

Then:

1) $a_{11} > 0$, $\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0, \dots, \begin{vmatrix} a_{11} & \cdots & \\ \vdots & \ddots & \\ & & a_{dd} \end{vmatrix} > 0 \Rightarrow Q$ is positive definite

2) $a_{11} < 0$, $\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0, \dots, (-1)^d \begin{vmatrix} a_{11} & \cdots & \\ \vdots & \ddots & \\ & & a_{dd} \end{vmatrix} > 0 \Rightarrow Q$ is negative definite

3) otherwise the crit. is not effective

3. THEOREM:

$f : \mathbb{R}^d \to \mathbb{R}$ twice Fréchet differentiable in x^*. If $\nabla f(x^*) = 0_{\mathbb{R}^d}$ and

$H_f(x^*) = \nabla^2 f(x^*)$ is
- positive definite $\Rightarrow x^*$ is minimum
- negative definite $\Rightarrow x^*$ is maximum

4. LEAST SQUARES METHOD :

Given :

- a set of data (measurement)

x	$x_1 \dots x_i \dots x_m$
y	$y_1 \dots y_i \dots y_m$

- a model $f(x) = ax+b$

Goal : find a^*, b^* such that a^*x+b^* is the best fit for the given data

$$E(a,b) = \sum_{i=1}^{m} \left(y_i - (ax_i + b)\right)^2 \longrightarrow \min$$

LECTURE 5: CONSTRAINT OPTIMIZATION

1. PLANAR CURVES:

a) Implicit form: $f(x_1, x_2) = 0$

ex: circle : $x_1^2 + x_2^2 = 0$

folium : $x_1^3 + x_2^3 - 3x_1x_2 = 0$

b) Explicit form (solve the implicit equation): $x_2 = \varphi_f(x_1)$

ex: circle : $x_2 = \pm\sqrt{1-x_1^2}$

c) Parametric form ("add a parameter") : $\begin{cases} x_1 = x_1(t) \\ x_2 = x_2(t) \end{cases}$, $t \in [0, T)$

ex: circle $\begin{cases} x_1 = \cos t \\ x_2 = \sin t \end{cases}$ $t \in [0, 2\pi)$

folium $\begin{cases} x_1 = \dfrac{3t}{1+t^3} \\ x_2 = \dfrac{3t^2}{1+t^3} \end{cases}$

2. LEVEL SETS:

$f: \mathbb{R}^d \to \mathbb{R}$, $c \in \mathbb{R} \Rightarrow \Gamma_c = \{(x_1, \dots, x_d) \in \mathbb{R}^d : f(x_1, \dots, x_d) = c\} \to$ c - level set of f

you can also talk about sub-level sets:

$\Gamma_{\leq c} = \{(x_1, \dots, x_d) \in \mathbb{R}^d : f(x_1, \dots, x_d) \leq c\}$

The gradient is orthogonal to level sets:

$f: \mathbb{R}^2 \to \mathbb{R}$, ∇f continuous, $\Gamma_c \neq \emptyset$, $c \in \mathbb{R}$

If $\Gamma \begin{cases} x_1 = x_1(t) \\ x_2 = x_2(t) \end{cases}$ $t \in [0, T)$ is a differentiable parametric curve, then

the tangent to Γ is given by: $\frac{d}{dt}(x_1(t), x_2(t)) = (x_1'(t), x_2'(t))$

3. LAGRANGE MULTIPLIER METHOD:

Let $f, g : \mathbb{R}^2 \to \mathbb{R}$ continuous and differentiable, x^* conditional minimum. Then there exists $\lambda^* \in \mathbb{R}$ such that $(x_1^*, x_2^*, \lambda^*)$ is a local (unconditional) minimum for $L(x_1, x_2, \lambda) = f(x_1, x_2) - \lambda g(x_1, x_2)$ that is

$$\nabla L(x_1, x_2, \lambda) = 0_{\mathbb{R}^3} \iff \begin{cases} \dfrac{\partial L}{\partial x_1} = 0 \\ \dfrac{\partial L}{\partial x_2} = 0 \\ \dfrac{\partial L}{\partial \lambda} = 0 \end{cases}$$

Geometric insight: At the conditional minimum point the $g = 0$ and $f = c$ contour lines are tangent to eachother.

The only good case is: $\nabla f \perp$ tangent but also $\nabla g \perp$ tangent $\Rightarrow$ $\Rightarrow \nabla f, \nabla g$ colinear $\Rightarrow \exists \lambda^* \in \mathbb{R}$ s.t. $\nabla f(x^*) = \lambda^* g(x^*) \iff \nabla f(x^*) - \lambda^* g(x^*) = 0 \iff$ $\iff \nabla L = 0$