6. előadás

2020. március 23.

Magasabb rendű deriváltak

Emlékeztetünk arra, hogy az $f \in \mathbb{R} \to \mathbb{R}$ valós-valós függvények körében „nem okozott gondot” az f függvény 2-szer, 3-szor, ... való differenciálhatóságának teljes indukcióval történő értelmezése. Először a kétszer deriválhatóság fogalmát definiáltuk. Azt mondtuk, hogy a szóban forgó függvény *kétszer differenciálható az* $a \in \operatorname{int} \mathcal{D}_f$ *pontban* (ezt röviden az $f \in D^2\{a\}$ szimbólummal jelöltük), ha f az a pont egy $K(a) \subset \mathcal{D}_f$ környezetében deriválható (értelmezve van tehát az $f' : K(a) \to \mathbb{R}$ deriváltfüggvény) és az f' deriváltfüggvény differenciálható az a pontban, azaz $f' \in D\{a\}$. Ekkor az $f''(a) := \big(f'\big)'(a)$ számot az f függvény *a pontbeli második deriváltjának neveztük.* Teljes indukcióval hasonlóan értelmeztük a 2-nél magasabb rendű deriválhatóság fogalmát.

Vegyük észre, hogy a kétszeri differenciálhatóságnak ez az értelmezése átvihető az $f \in \mathbb{R}^n \to \mathbb{R}$ függvényekre akkor is, ha $2 \le n \in \mathbb{N}$. Ti., ha $a \in \operatorname{int} \mathcal{D}_f$ és $\exists\, K(a) \subset \mathcal{D}_f$, hogy $f \in D\big(K(a)\big)$, akkor értelmezhető az

$$f' = \big(\partial_1 f, \partial_2 f, \ldots, \partial_n f\big) = \operatorname{grad} f \in \mathbb{R}^n \to \mathbb{R}^n$$

deriváltfüggvény, és $\mathbb{R}^n \to \mathbb{R}^n$ vektor-vektor függvények körében értelmeztük a deriválhatóságot. Tehát az $f \in \mathbb{R}^n \to \mathbb{R}$ függvényre minden további nélkül előírhatjuk, hogy $f' = \operatorname{grad} f \in D\{a\}$ teljesüljön. Ekkor $f''(a) := \big(f'\big)'(a) = \big(\operatorname{grad} f\big)'(a)$ az f függvény második deriváltja az a pontban. Világos, hogy egy $f \in \mathbb{R}^n \to \mathbb{R}$ függvény esetén $f''(a)$ egy $(n \times n)$-es mátrix.

Az is nyilvánvaló azonban, hogy ezen az úton nem tudunk eljutni a 2-nél magasabb rendű deriválhatósághoz, hiszen pl. $\mathbb{R}^n \to \mathbb{R}^{n\times n}$ típusú függvények differenciálhatóságát nem értelmeztük.

Emlékeztetünk ugyanakkor arra, hogy

$$f' = \operatorname{grad} f = \big(\partial_1 f, \partial_2 f, \ldots, \partial_n f\big) \in D\{a\}$$

azzal ekvivalens, hogy a $\partial_i f$ $(i = 1, 2, \ldots, n)$ parciális deriváltfüggvények differenciálhatók az a pontban, azaz

$$\partial_i f \in D\{a\} \qquad (i = 1, 2, \ldots, n).$$

Éppen ezért az alábbiak szerint definiáljuk egy $f \in \mathbb{R}^n \to \mathbb{R}$ függvény magasabb rendű deriválhatóságát. Az értelmezés teljes indukcióval történik, első lépésként a kétszer deriválhatóságot definiálva.

1. definícó. Az $f \in \mathbb{R}^n \to \mathbb{R}$ $(n \in \mathbb{N})$ függvény *kétszer deriválható* (vagy *differenciálható*) az $a \in \operatorname{int} \mathcal{D}_f$ pontban (jelben: $f \in D^2\{a\}$), ha

(a) $\exists\, K(a) \subset \mathcal{D}_f$, hogy $f \in D\{x\}$ minden $x \in K(a)$ pontban, és

(b) $\forall\, i = 1, 2, \ldots, n$ indexre $\partial_i f \in D\{a\}$.

Az (a) feltételt röviden úgy is írhatjuk, hogy $\exists\, K(a) \subset \mathcal{D}_f : \; f \in D\big(K(a)\big)$. Ebből következik, hogy $K(a)$ környezetben létezik az

$$f' = \big(\partial_1 f, \partial_2 f, \dots, \partial_n f\big) = \operatorname{grad} f \in \mathbb{R}^n \to \mathbb{R}^n$$

deriváltfüggvény, ami tehát már egy vektor-vektor függvény.

A (b) feltétel pedig azzal ekvivalens, hogy $f' \in D\{a\}$. Így minden $i = 1, 2, \dots, n$ esetén a $\partial_i f \in \mathbb{R}^n \to \mathbb{R}$ parciális deriváltfüggvényeknek léteznek az a pontban mindegyik változó szerinti parciális deriváltjai:

$$\partial_j\big(\partial_i f\big)(a) \quad (j = 1, 2, \dots, n).$$

Ezeket a számokat az f függvény *a pontbeli, i-edik és j-edik változó szerinti másodrendű* (vagy *második*) *parciális deriváltjának* nevezzük.

Az f függvény *a pontbeli második deriváltját* így értelmezzük:

$$f''(a) := \big(\operatorname{grad} f\big)'(a).$$

Világos, hogy egy $f \in \mathbb{R}^n \to \mathbb{R}$ függvény esetén $f''(a)$ egy $(n \times n)$-es mátrix.

2. definícó. Ha az $f \in \mathbb{R}^n \to \mathbb{R}$ $(2 \le n \in \mathbb{N})$ függvény kétszer deriválható az $a \in \operatorname{int} \mathcal{D}_f$ pontban, akkor

$$f''(a) = \begin{bmatrix} \partial_{11} f(a) & \partial_{12} f(a) & \dots & \partial_{1n} f(a) \\ \partial_{21} f(a) & \partial_{22} f(a) & \dots & \partial_{2n} f(a) \\ \vdots & \vdots & \vdots & \vdots \\ \partial_{n1} f(a) & \partial_{n2} f(a) & \dots & \partial_{nn} f(a) \end{bmatrix} \in \mathbb{R}^{n \times n}$$

az f függvény *a pontbeli Hesse-féle mátrixa*, ahol

$$\partial_{ij} f(a) = \partial_j\big(\partial_i f\big)(a) \quad \big(i = 1, 2, \dots, n;\; j = 1, 2, \dots, n\big).$$

A 2-nél magasabb rendű deriválhatóságot teljes indukcióval így értelmezzük:

3. definícó. Az $f \in \mathbb{R}^n \to \mathbb{R}$ $(2 \le n \in \mathbb{N})$ függvény s-szer $(2 \le s \in \mathbb{N})$ deriválható az $a \in \operatorname{int} \mathcal{D}_f$ pontban (jelben $f \in D^s\{a\}$), ha

(a) $\exists\, K(a) \subset \mathcal{D}_f$, hogy $f \in D^{s-1}\big(K(a)\big)$ és

(b) minden $(s-1)$-edrendű

$$\partial_{i_1} \partial_{i_2} \dots \partial_{i_{s-1}} f \quad \big(1 \le i_1, i_2, \dots, i_{s-1} \le n\big)$$

parciális deriváltfüggvény deriválható az a pontban.

A magasabb rendű parciális deriváltakkal kapcsolatban joggal merül fel a *kérdés*, hogy azok kiszámításakor van-e szerepe a változók sorrendjének? A következő tétel azt állítja, hogy ha az f függvény a szóban forgó a helyen „elég sokszor” deriválható, akkor a sorrend elveszti a jelentőségét.

1. tétel. (Young-tétel.) Ha $f \in \mathbb{R}^n \to \mathbb{R}$ $(2 \le n \in \mathbb{N})$ és $f \in D^2\{a\}$, akkor

$$\partial_{ij} f(a) = \partial_{ji} f(a) \quad \forall\, i, j = 1, \ldots, n \;\text{ indexre.}$$

Példa. Az $f \in D^2\{a\}$ feltétel hiánya esetén a parciális deriváltak képzésének a sorrendje általában nem cserélhető fel. Ha például

$$f(x,y) := \begin{cases} xy\dfrac{x^2-y^2}{x^2+y^2}, & \text{ha } (x,y) \in \mathbb{R}^2 \setminus \{(0,0)\} \\ 0, & \text{ha } (x,y) = (0,0), \end{cases}$$

akkor

$$\partial_{12} f(0,0) = -1, \qquad \partial_{21} f(0,0) = 1.$$

Érdemes meggondolni azt, hogy $f \notin D^2\{a\}$.

Megjegyzés. Teljes indukcióval igazolható a Young-tétel következő általánosítása: *Legyen az $f \in \mathbb{R}^n \to \mathbb{R}$ $(n \in \mathbb{N})$ függvény s-szer $(s \in \mathbb{N})$ differenciálható az $a \in \operatorname{int} \mathcal{D}_f$ pontban. Tegyük fel, hogy $i_k \in \mathbb{N}$, $1 \le i_k \le n$ $(k = 1, 2, \ldots, s)$ és $j_1, j_2, \ldots, j_s$ az $i_1, i_2, \ldots, i_s$ indexek egy tetszőleges permutációja. Ekkor*

$$\partial_{i_1} \partial_{i_2} \cdots \partial_{i_s} f(a) = \partial_{j_1} \partial_{j_2} \cdots \partial_{j_s} f(a).$$

Taylor-polinomok. Taylor-formula

Induljunk ki az egyváltozós ismereteinkből. *Motivációként* akkor azt a fontos *problémát* vetettük fel, hogy egy adott „bonyolult” függvényt vajon meg lehet-e közelíteni egyszerű szerkezetű függvényekkel, például a jól kezelhető és könnyen számolható polinomokkal.

A valós-valós esetben beláttuk, hogy egy függvény akkor és csak akkor differenciálható egy pontban, ha annak a pontnak egy környezetében a függvény lokálisan jól közelíthető elsőfokú polinommal. Ennek általánosításaként jutottunk el ahhoz a fontos eredményhez, hogy ha egy függvény (mondjuk) m-szer deriválható a szóban forgó pont egy környezetében, akkor ebben a környezetben a függvény jól közelíthető egy alkalmasan megválasztott $(m-1)$-edfokú polinommal, nevezetesen a függvény adott ponthoz tartozó *Taylor-polinomjával.* Ezzel kapcsolatos alapvető eredményünk volt a *Taylor-formulára* vonatkozó alábbi állítás:

Ha $f \in \mathbb{R} \to \mathbb{R}$, $m \in \mathbb{N}$ és egy $K(a) \subset \mathcal{D}_f$ környezetben $f \in D^m\big(K(a)\big)$, akkor minden $h > 0$ $(a + h \in K(a))$ számhoz létezik olyan $\nu \in (0,1)$ szám, amelyre az

$$f(a+h) = \sum_{k=0}^{m-1} \frac{f^{(k)}(a)}{k!} h^k + \frac{f^{(m)}(a+\nu h)}{m!} h^m \tag{1}$$

egyenlőség teljesül.

A legfeljebb $(m-1)$-edfokú

$$T_{a,m-1}(f,h) := \sum_{k=0}^{m-1} \frac{f^{(k)}(a)}{k!} h^k \qquad (h \in \mathbb{R})$$

polinom az f függvény *a ponthoz tartozó* $(m-1)$*-edik Taylor-polinomja*, az (1) egyenlőség jobb oldalán álló összeg utolsó tagja pedig a *Taylor-formula Lagrange-féle maradéktagja.* Az elnevezést az a tény indokolja, hogy ez a tag $T_{a,m-1}(f,h)$-hoz képest kicsi, ha h közel van 0-hoz, azaz

$$f(a+h) \sim T_{a,m-1}(f,h) = \sum_{k=0}^{m-1} \frac{f^{(k)}(a)}{k!}h^k, \quad \text{ha } h \sim 0;$$

vagyis egy elég sima függvény az a pont környezetében lokálisan jól közelíthető egy elég magas fokszámú polinommal, nevezetesen az f függvény a ponthoz tartozó Taylor-polinomjával.

Figyeljük meg, hogy a fenti egyváltozós eredmény $f \in \mathbb{R}^n \to \mathbb{R}$ függvényekre vonatkozó általánosításához *egyrészt* értelmezni kellene $h \in \mathbb{R}^n$ esetén a h^k hatványokat, *másrészt* az $f^{(k)}(a)$ deriváltakat, ha $k = 1, 2, \ldots, m$.

A továbbiakban ezt a kiterjesztést csak az $n = 2$ és az $m = 2$ speciális esetekben *vázoljuk*. Legyen tehát $f \in \mathbb{R}^2 \to \mathbb{R}$ egy adott függvény, $a = (a_1, a_2) \in \operatorname{int} \mathcal{D}_f$ és $h = (h_1, h_2) \in \mathbb{R}^2$. Tegyük fel egyelőre még azt is, hogy egy $K(a) \subset \mathcal{D}_f$ környezetben az f függvény kétszer *folytonosan deriválható*, azaz $f \in C^2\big(K(a)\big)$. Vegyük a síkon az a és az $a + h \in K(a)$ pontokat összekötő egyenes $a + th \ \ (t \in \mathbb{R})$ pontjait. Tekintsük a

$$F(t) := f(a+th) \quad \big(t \in \mathbb{R}\big)$$

valós-valós függvényt. Mivel $f \in C^2\big(K(a)\big)$, ezért az összetett függvény deriválására vonatkozó tételből következik, hogy a F és a F' függvény folytonos a $[0,1]$ intervallumon és F' differencálható $(0,1)$-en.

Alkalmazzuk a F függvényre az (1) alatti Taylor-formulát a $[0,1]$ intervallumon: létezik tehát olyan $\nu \in (0,1)$, hogy

$$F(1) = F(0) + F'(0) + \frac{1}{2!}F''(\nu).$$

Most kiszámítjuk az $F(0)$, az $F'(0)$ és az $F''(\nu)$ értékeket. Mivel

$$F(t) = f(a+th) = f\big(a_1 + th_1, a_2 + th_2\big),$$

ezért

$$F(0) = f(a).$$

Az összetett függvény deriválására vonatkozó tétel (a láncszabály) alapján

$$F'(t) = \partial_1 f(a+th) \cdot h_1 + \partial_2 f(a+th) \cdot h_2 = \langle f'(a+th), h \rangle,$$

következésképpen

$$F'(0) = \langle f'(a), h \rangle.$$

A F függvény második deriváltja:

$$\begin{aligned} F''(t) &= \big(\partial_1 f(a+th)\cdot h_1 + \partial_2 f(a+th)\cdot h_2\big)' = \\ &= \Big(\partial_{11} f(a+th)\cdot h_1 + \partial_{12} f(a+th)\cdot h_2\Big)\cdot h_1 + \\ &\quad + \Big(\partial_{21} f(a+th)\cdot h_1 + \partial_{22} f(a+th)\cdot h_2\Big)\cdot h_2 = \end{aligned}$$

$$= \partial_{11} f(a+th)\cdot h_1^2 + \partial_{12} f(a+th)\cdot h_1 h_2 + \partial_{21} f(a+th)\cdot h_1 h_2 + \partial_{22} f(a+th)\cdot h_2^2 =$$

$$= \langle f''(a+th)\cdot h, h\rangle.$$

(Az utolsó egyenlőséget gondolja végig a mátrixszorzás és a skaláris szorzat definícióinak birtokában. Itt $f''(a+th)$ egy (2×2)-es mátrix, $h = \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}$ oszlopvektor, tehát $f''(a+th)\cdot h \in \mathbb{R}^{2\times 1} \approx \mathbb{R}^2$ egy oszlopvektor.) Így

$$F''(\nu) = \langle f''(a+\nu h)\cdot h, h\rangle.$$

A feltételeinkből következik, hogy az $f'' \in \mathbb{R}^2 \to \mathbb{R}^{2\times 2}$ függvény folytonos az $a \in \mathcal{D}_{\mathbf{f}}$ pontban, ezért

$$f''(a+\nu h) = f''(a) + \eta(h),$$

ahol $\eta = \big[\eta_{ij}\big] \in \mathbb{R}^2 \to \mathbb{R}^{2\times 2}$ egy olyan függvény, amelyre $\lim_{\mathbf{0}} \eta = \mathbf{0}$.

A fentiek alapján tehát

$$F(1) = f(a+h) = f(a) + \langle f'(a), h\rangle + \tfrac{1}{2!}\langle f''(a)\cdot h, h\rangle + \tfrac{1}{2!}\langle \eta(h)\cdot h, h\rangle.$$

Az utolsó tagot még tovább alakítjuk:

$$\langle \eta(h)\cdot h, h\rangle = \sum_{i,j=1}^{2} \eta_{ij}(h)\cdot h_i h_j = \|h\|^2 \cdot \sum_{i,j=1}^{2} \eta_{ij}(h)\cdot \frac{h_i h_j}{\|h\|^2}.$$

Mivel $\dfrac{|h_i h_j|}{\|h\|^2} \le M$ $(i,j=1,2)$ és $\lim_{h\to\mathbf{0}} \eta_{ij} = 0$, ezért

$$\tfrac{1}{2!}\langle \eta(h)\cdot h, h\rangle = \varepsilon(h)\cdot \|h\|^2,$$

ahol $\varepsilon \in \mathbb{R}^2 \to \mathbb{R}$ olyan függvény, amelyre $\lim_{\mathbf{0}} \varepsilon = 0$ teljesül.

Az eddigieket összefoglalva azt kaptuk, hogy van olyan $\varepsilon \in \mathbb{R}^2 \to \mathbb{R}$, a $\lim_{\mathbf{0}} \varepsilon = 0$ feltételt kielégítő függvény, amelyre

$$f(a+h) = f(a) + \langle f'(a), h\rangle + \tfrac{1}{2}\,\langle f''(a)h, h\rangle + \varepsilon(h)\|h\|^2. \tag{2}$$

A fenti gondolatmenetet követve viszonylag egyszerűen belátható, hogy az (2) képlet tetszőleges $f \in \mathbb{R}^n \to \mathbb{R}$ $(2 < n \in \mathbb{N})$ függvényre is teljesül. Nehezebb már annak a bizonyítása, hogy az állítás az $f \in C^2\big(K(a)\big)$ helyett az $f \in D^2\{a\}$ feltétel mellet is igaz.

2. tétel. (A Taylor-formula a Peano-féle maradéktaggal.) Tegyük fel, hogy $f \in \mathbb{R}^n \to \mathbb{R}$ $(n \in \mathbb{N})$ és az $a \in \operatorname{int} \mathcal{D}_f$ pontban $f \in D^2\{a\}$. Ekkor van olyan $\varepsilon \in \mathbb{R}^n \to \mathbb{R}$, a $\lim_{\mathbf{0}} \varepsilon = 0$ feltételnek eleget tevő függvény, hogy

$$(3) \quad f(a+h) = f(a) + \langle f'(a), h\rangle + \tfrac{1}{2}\langle f''(a)h, h\rangle + \varepsilon(h)\cdot\|h\|^2 \quad \big(h \in \mathbb{R}^n,\ a+h \in \mathcal{D}_f\big).$$

A jobb oldal első három tagjának az összegét (ez egy n-változós legfeljebb másodfokú polinom) az f függvény *a ponthoz tartozó második Taylor-polinomjának* nevezzük, és így jelöljük:

$$\begin{aligned} T_{a,m}(f,h) &:= f(a) + \langle f'(a), h\rangle + \tfrac{1}{2}\langle f''(a)h, h\rangle = \\ &= f(a) + \sum_{i=1}^{n} \partial_i f(a) h_i + \frac{1}{2}\sum_{i,j=1}^{n} \partial_{ij} f(a) h_i h_j. \end{aligned}$$

A (3) képletben az $\varepsilon(h)\cdot\|h\|^2$ tagot a *Taylor-formula Peano-féle maradéktagjának* nevezzük.

Megjegyzések. 1. A 2. tétel jelentősége egyrészt abban van, hogy a felhasználásával „bonyolult” n-változós valós értékű függvények helyettesítési értékeire lehet „jó” közelítő értékeket adni. Másrészt, azt a következő órán látni fogjuk, hogy a 2. tétel fontos szerepet játszik $\mathbb{R}^n \to \mathbb{R}$ típusú függvények szélsőérték-problémáinak a vizsgálatánál.

2. A 2. tétel általánosítható arra az esetre is, ha $f \in D^s\{a\}$, aholl $s > 2$ tetszőleges természetes szám.