5. előadás

2020. március 9.

$\mathbb{R}^n \to \mathbb{R}^m$ függvények totális deriváltja

Emlékeztető. Említettük már azt, hogy a többváltozós függvények körében a deriválás több változatosságot kínál, mint a folytonosság vagy a határérték. Itt többféle derivált fogalommal kell megismerkednünk. A két egyszerűbbről, nevezetesen a *parciális-* és az *iránymenti deriváltakról* már volt szó.

Most $\mathbb{R}^n \to \mathbb{R}^m$ vektor-vektor függvény *totális deriváltját* értelmezzük. Látni fogjuk, hogy ugyan a totális deriválhatóság fogalma pontosan megfelel az egyváltozós deriválhatóság (egyik) definíciójának, de a totális derivált fogalma bonyolultabb, mint egy változóban. Ezzel kapcsolatban emlékeztetünk valós-valós függvények deriválhatóságára.

• Az $f \in \mathbb{R} \to \mathbb{R}$ függvény *differenciálható* vagy *deriválható* az $a \in \operatorname{int} \mathcal{D}_f$ pontban (jelben: $f \in D\{a\}$), ha létezik és véges a

$$\lim_{h \to 0} \frac{f(a+h) - f(a)}{h} \quad \text{határérték.}$$

Ezt a határértéket az $f'(a)$ szimbólummal jelöljük, és az f függvény a pontbeli *deriváltjának* vagy *differenciálhányadosának* nevezzük.

• A többváltozós függvények esetében a differenciahányadosnak nincs közvetlen megfelelője (hiszen vektorok körében nem tudunk osztani), ezért a deriválhatóságot nem tudjuk differenciahányadosok határértékeként értelmezni. Az egyváltozós analízisben láttuk, hogy az elsőfokú plonimomokkal való lokális közelíthetőség ekvivalens a differenciálhatósággal. Nevezetesen: Ha $f \in \mathbb{R} \to \mathbb{R}$ és $a \in \operatorname{int} \mathcal{D}_f$, akkor

$$f \in D\{a\} \quad \Longleftrightarrow \quad \begin{cases} \exists\, A \in \mathbb{R} \ \text{ és } \ \exists\, \varepsilon : \mathcal{D}_f \to \mathbb{R},\ \lim\limits_{0} \varepsilon = 0 : \\ f(a+h) - f(a) = A \cdot h + \varepsilon(h) \cdot h \ \ \big(a + h \in \mathcal{D}_f\big). \end{cases}$$

Az A szám az f függvény $a \in \operatorname{int} \mathcal{D}_f$ pontbeli deriváltja, vagyis $A = f'(a)$.

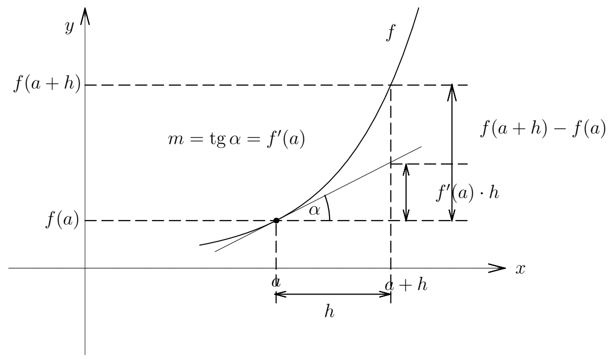
A függvényértékek megváltozása tehát

$$f(a+h) - f(a) = f'(a) \cdot h + \varepsilon(h) \cdot h \ \ \big(a + h \in \mathcal{D}_f\big).$$

A jobb oldalon álló első tag egy lineáris függvény, a második tag pedig a $\lim\limits_{0} \varepsilon = 0$ feltétel miatt az elsőhöz képest „kicsi”. Az f függvény *a pontbeli deriválhatósága tehát azt jelenti, hogy a függvény megváltozása az a pont környezetében„jól” közelíthető lineáris függvénnyel*:

$$f(a+h) - f(a) \sim f'(a) \cdot h, \quad \text{ha } h \sim 0.$$

Ezt szemlélteti a következő ábra.



• Vegyük észre, hogy a fentiekben az ε függvény szerepeltetése „kiküszöbölhető”. Pontosabban: Ha $f \in \mathbb{R} \to \mathbb{R}$ és $a \in \operatorname{int} \mathcal{D}_f$, akkor

$$f \in D\{a\} \quad \Longleftrightarrow \quad \exists\, A \in \mathbb{R}: \quad \lim_{h\to 0} \frac{f(a+h) - f(a) - A \cdot h}{h} = 0.$$

Ha még azt is figyelembe vesszük, hogy egy $g \in \mathbb{R} \to \mathbb{R}$ függvény esetén $\lim\limits_0 g = 0 \iff \lim\limits_0 |g| = 0$, akkor végül azt kapjuk, hogy

$$(1) \qquad \boxed{f \in D\{a\} \quad \Longleftrightarrow \quad \exists\, A \in \mathbb{R}: \quad \lim_{h\to 0} \frac{|f(a+h) - f(a) - A \cdot h|}{|h|} = 0.}$$

Valós-valós függvény deriválhatóságára vonatkozó (1) ekvivalens átfogalmazás már kiterjeszthető vektor-vektor függvényre is, ha (1)-ben az abszolút értéket a megfelelő *nomákkal*, az A valsós számot pedig *$m \times n$-es mátrixszal* helyettesítjük.

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

$$\exists\, A \in \mathbb{R}^{m\times n}: \quad \lim_{h\to \mathbf{0}} \frac{\|f(a+h) - f(a) - A \cdot h\|_m}{\|h\|_n} = 0,$$

ahol $\|\cdot\|_n$, illetve $\|\cdot\|_m$ az euklideszi norma az $\mathbb{R}^n$, illetve az $\mathbb{R}^m$ lineáris téren.
Ekkor A egyértelmű, és $f'(a) := A$ az f függvény *deriváltmátrixa* az a pontban.

Megjegyzés. Az 1. definícióból következik, hogy egy $f \in \mathbb{R}^n \to \mathbb{R}^m$ vektor-vektor függvény $a \in \operatorname{int} \mathcal{D}_f$ pontbeli deriválhatósága az egyváltozós esethez hasonlóan azt jelenti, hogy a függvény megváltozása az a pont környezetében „jól” közelíthető lineáris függvénnyel:

$$f(a+h) - f(a) \sim A \cdot h, \quad \text{ha } h \sim \mathbf{0}.$$

Itt A egy $m \times n$-es mátrix, h egy n-dimenziós oszlopvektor és $\cdot$ a mátrixok közötti szorzás művelete. A jobb oldalon álló

$$L : \mathbb{R}^n \to \mathbb{R}^m, \qquad L(h) := A \cdot h$$

függvény *lineáris*, ami azt jelenti, hogy

$$L(\alpha x+\beta y)=\alpha L(x)+\beta L(y)$$

teljesül minden $x, y \in \mathbb{R}^n$ és $\alpha, \beta \in \mathbb{R}$ esetén.

A definíció alapján a deriváltmátrix előállítása általában nem egyszerű feladat. A következő tétel azonban azt állítja, hogy a könnyen számolható parciális deriváltak segítségével egyszerűen felírható az $f \in \mathbb{R}^n \to \mathbb{R}^m$ függvény a pontbeli deriváltmátrixa.

1. tétel. (A deriváltmátrix előállítása.) Legyen $n, m \in \mathbb{N}$,

$$f=\begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{bmatrix} \in \mathbb{R}^n \to \mathbb{R}^m,$$

ahol $f_i \in \mathbb{R}^n \to \mathbb{R}$ az f függvény i-edik $(i=1,2,\ldots,m)$ koordinátafüggvénye.

Ha $f \in D\{a\}$, akkor

$$\exists\, \partial_j f_i(a) \quad (\forall\, i=1,\ldots,m;\ j=1,\ldots,n) \ \text{ és}$$

$$f'(a)=\begin{bmatrix} \partial_1 f_1(a) & \partial_2 f_1(a) & \cdots & \partial_n f_1(a) \\ \partial_1 f_2(a) & \partial_2 f_2(a) & \cdots & \partial_n f_2(a) \\ \vdots & \vdots & \vdots & \vdots \\ \partial_1 f_m(a) & \partial_2 f_m(a) & \cdots & \partial_n f_m(a) \end{bmatrix} \in \mathbb{R}^{m\times n}$$

a *deriváltmátrix* vagy *Jacobi-mátrix.*

Speciális esetek:

- Ha $m=1$, azaz f egy $\mathbb{R}^n \to \mathbb{R}$ függvény, akkor f deriváltmátrixa az a pontban

$$f'(a)=\begin{bmatrix} \partial_1 f(a) & \partial_2 f(a) & \ldots & \partial_n f(a) \end{bmatrix} \in \mathbb{R}^{1\times n} \approx \mathbb{R}^n,$$

azaz ebben az esetben az $f'(a)$ Jacobi-mátrix tekinthető egy $\mathbb{R}^n$-beli sorvektornak, amit az f függvény a-beli *gradiensének* nevezünk, és a $\operatorname{grad} f(a)$ szimbólummal jelölünk.

- Ha $n=1$, azaz $f=(f_1, f_2, \ldots, f_m)$ egy $\mathbb{R} \to \mathbb{R}^m$ függvény, akkor f deriváltmátrixa az a pontban

$$f'(a)=\begin{bmatrix} f_1'(a) \\ f_2'(a) \\ \vdots \\ f_n'(a) \end{bmatrix} \in \mathbb{R}^{m\times 1} \approx \mathbb{R}^m,$$

azaz ebben az esetben az $f'(a)$ Jacobi-mátrix tekinthető egy $\mathbb{R}^m$-beli oszlopvektornak.

A totális deriválhatóság definíciójából következik, hogy ha $f \in D\{a\}$, akkor f folytonos is az a pontban. Nem meglepő, hogy ennek az állításnak a megfordítása *nem igaz.* Másrészt az 1. tétel alapján, ha $f \in D\{a\}$, akkor az összes parciális derivált létezik az a pontban. Az is sejthető azonban, hogy a parciális deriváltak létezéséből *nem következik* a totális deriválhatóság. Például az

$$f(x,y) := \sqrt{xy} \qquad \big((x,y) \in \mathbb{R}^2\big)$$

függvény folytonos az origóban, itt léteznek a parciális deriváltak, de f nem totálisan deriválható az $a := (0,0)$ pontban.

Szerencsére az egyszerűen számolható parciális deriváltakból is következtethetünk a totális deriválhatóságra. Ehhez a fentiek alapján a pontbeli parciális deriváltak létezésénél többet kell feltennünk. A következő tétel egy ilyen gyakran alkalmazható *elégséges feltételt* ad a függvény totális deriválhatóságára.

2. tétel. (Elégséges feltétel a totális deriválhatóságra.) Legyen $f \in \mathbb{R}^n \to \mathbb{R}$ $(n \in \mathbb{N})$ és $a \in \operatorname{int} \mathcal{D}_f$. Tegyük fel, hogy az a pontnak létezik olyan $K(a) \subset \mathcal{D}_f$ környezete, amelyre minden $i = 1, \ldots, n$ index esetén a következők teljesülnek:

(a) $\exists\, \partial_i f(x)$ minden $x \in K(a)$ pontban,

(b) a $\partial_i f : K(a) \to \mathbb{R}$ parciális deriváltfüggvény folytonos az a pontban. Ekkor az f függvény totálisan deriválható az a pontban.

Felület érintősíkja

Az egyváltozós analízisben láttuk, hogy ha $f \in D\{a\}$, akkor az f függvényt az a pont környezetében jól közelítő elsőfokú polinom nem más, mint a függvénygrafikon $\big(a, f(a)\big)$ pontbeli érintője. Most megvizsgáljuk, hogy mi felel meg ennek az állításnak az $\mathbb{R}^2 \to \mathbb{R}$ függvények körében.

Legyen $f \in \mathbb{R}^2 \to \mathbb{R}$, $(x_0, y_0) \in \operatorname{int} \mathcal{D}_f$, és tegyük fel, hogy $f \in D\{(x_0, y_0)\}$. Ez azt jelenti, hogy

$$f(x,y) - f(x_0,y_0) \sim \partial_1 f(x_0,y_0)(x - x_0) + \partial_2 f(x_0,y_0)(y - y_0), \qquad \text{ha } (x,y) \sim (x_0,y_0).$$

Legyen $z_0 := f(x_0, y_0)$. Ekkor

$$z - z_0 = \partial_1 f(x_0,y_0)(x - x_0) + \partial_2 f(x_0,y_0)(y - y_0) \tag{2}$$

esetén $f(x,y) \sim z$, ha $(x,y) \sim (x_0,y_0)$. A (2) egyenlet egy olyan sík egyenlete a térben, amelyik átmegy az $\big(x_0, y_0, f(x_0,y_0)\big)$ ponton és egy normálvektora

$$\mathbf{n} = \big(\partial_1 f(x_0,y_0), \partial_2 f(x_0,y_0), -1\big).$$

Megjegyzés. A háromdimenziós térben a sík általános egyenlete

$$Ax + By + Cz = D,$$

alakú, ahol az A, B, C együtthatók legalább egyike nem nulla. Ekkor az

$$\mathbf{n} = (A, B, C) \in \mathbb{R}^3$$

egy olyan vektor, amelyik a sík minden vektorára merőleges (ez a sík egy *normálvektora*).

Valóban: Tekintsünk a térben egy S síkot. Legyen $P_0 = (x_0, y_0, z_0)$ az S sík egy tetszőleges pontja, és $\mathbf{x_0}$ ebbe a pontba mutató helyvektor. Legyen $\mathbf{n} = (A, B, C)$ az S síkra merőleges nemnulla vektor (az S sík egy *normálvektora*). A tér geometriájából következik, hogy egy $(x, y, z) \in \mathbb{R}^3$ koordinátájú pont (az ide mutató helyvektor $\mathbf{x}$) akkor és csak akkor eleme S-nek, ha az $\mathbf{x} - \mathbf{x_0}$ vektor merőleges az $\mathbf{n}$ vektorra, azaz

$$\langle \mathbf{x} - \mathbf{x_0}, \mathbf{n} \rangle = A(x - x_0) + B(y - y_0) + C(z - z_0) = 0.$$

Ebből átrendezéssel adódik az állítás, ahol $D = \langle \mathbf{n}, \mathbf{x_0} \rangle$.

Most már könnyen definiálhatjuk az érintőnek megfelelő fogalmat kétváltozós függvényekre.

2. definíció. Az $f \in \mathbb{R}^2 \to \mathbb{R}$ függvény grafikonjának az $\big(x_0, y_0, f(x_0, y_0)\big)$ pontban *van érintősíkja*, ha $f \in D\big\{(x_0, y_0)\big\}$. A érintősík *egyenlete*:

$$z - f(x_0, y_0) = \partial_1 f(x_0, y_0) \cdot (x - x_0) + \partial_2 f(x_0, y_0) \cdot (y - y_0)$$

és egy *normálvektora*:

$$\mathbf{n} = \big(\partial_1 f(x_0, y_0), \partial_2 f(x_0, y_0), -1\big).$$

Deriválási szabályok

$\mathbb{R}^n \to \mathbb{R}^m$ típusú vektor-vektor függvények közötti algebrai műveletek és a totális derivált kapcsolatára az egyváltozós esethez hasonló állítások érvényesek:

- Ha $f, g \in \mathbb{R}^n \to \mathbb{R}^m$ $(n, m \in \mathbb{N})$ és $f, g \in D\{a\}$, akkor $\forall\, \lambda, \mu \in \mathbb{R}$ esetén

$$\big(\lambda f + \mu g\big) \in D\{a\} \quad \text{és} \quad \big(\lambda f + \mu g\big)'(a) = \lambda f'(a) + \mu g'(a).$$

- Ha $m = 1$, akkor akkor az $f \cdot g$ és az $\dfrac{f}{g}$ függvényekre az egyváltozós esethez hasonló deriválási szabályok teljesülnek.

A következő tétel függvények kompozíciójának differenciálhatóságára és deriváltjára vonatkozik.

3. tétel. (Az összetett függvény deriválhatósága, az ún. *láncszabály.*) Legyen $n, m, s \in \mathbb{N}$. Ha $g \in \mathbb{R}^n \to \mathbb{R}^m$ és $g \in D\{a\}$, továbbá $f \in \mathbb{R}^m \to \mathbb{R}^s$ és $f \in D\{g(a)\}$, akkor $f \circ g \in D\{a\}$ és

$$\big(f \circ g\big)'(a) = f'\big(g(a)\big) \cdot g'(a), \tag{3}$$

ahol $\cdot$ a mátrixok közötti szorzás műveletét jelöli.

1. megjegyzés. Figyeljük meg a (3) képletben szereplő deriváltakat:
mivel $f \circ g \in \mathbb{R}^n \to \mathbb{R}^s$, ezért $\big(f \circ g\big)'(a) \in \mathbb{R}^{s \times n}$;
mivel $g \in \mathbb{R}^n \to \mathbb{R}^m$, ezért $g'(a) \in \mathbb{R}^{m \times n}$;
mivel $f \in \mathbb{R}^m \to \mathbb{R}^s$, ezért $f'\big(g(a)\big) \in \mathbb{R}^{s \times m}$.

Ez azt jelenti, hogy (3) bal oldalán egy $s \times n$-es mátrix áll. A jobb oldalon egy $s \times m$-es és egy $m \times n$-es mátrix ebben a sorrendben vett szorzata szerepel, ami egy $s \times n$-es mátrix.

2. megjegyzés. Tekintsük a láncszabályt a $2 \le n, m \in \mathbb{N}$ és $s = 1$ speciális esetben. Ha $g = (g_1, \ldots, g_m) \in \mathbb{R}^n \to \mathbb{R}^m$ és $g \in D\{a\}$, továbbá $f \in \mathbb{R}^m \to \mathbb{R}^1$ és $f \in D\{g(a)\}$, akkor az

$$F(x) := \big(f \circ g\big)(x) = f(g(x)) = f\big(g_1(x), \ldots, g_m(x)\big) \quad \big(x \in \mathcal{D}_g\big)$$

$\mathbb{R}^n \to \mathbb{R}$ típusú függvény differenciálható az a pontban, és a (3) képlet alapján

$$\partial_j F(a) = \sum_{k=1}^{m} \partial_k f\big(g(a)\big) \cdot \partial_j g_k(a) \tag{4}$$

minden $j = 1, 2, \ldots, n$ esetén.

A (4) összefüggés könnyebben megjegyezhető a következő alakban. Jelöljük f, illetve g változóit $y_1, \ldots, y_m$-mel, illetve $x_1, \ldots, x_n$-nel. Ekkor azt kapjuk, hogy ha $j = 1, 2, \ldots, n$, akkor

$$\frac{\partial F}{\partial x_j} = \frac{\partial f}{\partial y_1} \cdot \frac{\partial g_1}{\partial x_j} + \frac{\partial f}{\partial y_2} \cdot \frac{\partial g_2}{\partial x_j} + \cdots + \frac{\partial f}{\partial y_m} \cdot \frac{\partial g_m}{\partial x_j}.$$