Többváltozós függvények differenciálása
4. sillabusz a Többváltozós függvények kurzushoz

Mi az, hogy „sillabusz"? Ez egy olyan iromány, ami segédanyagnak készült. Vázlatos, pontatlan, (szándékoltan) hiányos. Segíti a tanulást, vázlatot, ötleteket ad. Ugyanakkor, egy része a törzsanyagon túlmutat(hat).

Differenciálásnak látszó dolgok

Az egyváltozós esetből emlékszünk a derivált

$$f'(a) := \lim_{h\to 0} \frac{f(a+h)-f(a)}{h}$$

definíciójára. Igen ám, de ezt szó szerint általánosítani nem tudjuk (pl. azért sem, mert számot vektorral osztani nem tudunk). Az első ötletünk az lehet, hogy a feladatot valahogyan „egyváltozósítjuk".

1. Definíció. Legyen az f k-változós függvény értelmezve az $\underline{a}$ pont egy környezetében. Az *i-edik változója szerinti parciális deriváltja* az $\underline{a}$ helyen

$$f'_{x_i}(\underline{a}) := \lim_{h\to 0} \frac{f(\underline{a}+h\cdot \underline{e}_i)-f(\underline{a})}{h}, \qquad \text{ahol} \qquad \underline{e}_i := (0,0,\dots,\overset{i}{1},\dots,0),$$

feltéve, hogy a limes létezik és véges.

Magyarul, csak az i-edik változót változtatjuk, a többit rögzítjük. Ez például azt jelenti, hogy ha az $f(x,y)$ kétváltozós függvény grafikonja egy felület az x,y sík felett, és az (a,b) pontban állított, az y, z tengelyekkel párhuzamos sík ezt a felületet egy görbében metszi, a görbe meredeksége az (a,b) pont fölött éppen $f'_y(a,b)$.

Sajnos, *ez nem igazi* derivált. Például a parciális differenciálhatóságból nem következik a folytonosság. (Példa: $f(x,y) := \operatorname{sign} xy$.) Előnyös viszont, hogy a parciális derivált az egyváltozós esetből ismert szabályok szerint könnyen számolható.

2. Definíció. Legyen az f k-változós függvény értelmezve az $\underline{a}$ pont egy környezetében. A függvény *$\underline{v}$ irány szerinti deriváltja* az $\underline{a}$ helyen

$$f'_{\underline{v}}(\underline{a}) := \lim_{t\to 0^+} \frac{f(\underline{a}+t\cdot \underline{v})-f(\underline{a})}{t\|\underline{v}\|},$$

feltéve, hogy a limes létezik és véges.

Ez például azt jelenti, hogy azon a bizonyos felületen az (a,b) pont felett állva és szigorúan a $\underline{v}$ irányába nézve, éppen $f'_{\underline{v}}(a,b)$ meredekségű emelkedést látunk magunk előtt. (A $t\to 0^+$ miatt ez inkább a „féloldali" deriváltra emlékeztet.)

Sajnos, *ez sem igazi* derivált. Például ebből sem következik a folytonosság, sem a parciális differenciálhatóság. (Példák: $f_1(x,y) := \begin{cases} 1, & \text{ha } x^2\le y \text{ vagy } y\le 0, \\ 0, & \text{ha } 0<y<x^2. \end{cases}$, $f_2(x,y) := \sqrt{x^2+y^2}$.)

Azért vannak „derivált-szerű" tulajdonságaik is.

1. Tétel. (Lagrange-féle középértéktétel-variáció.) *Legyen az f k-változós függvény értelmezve az $\underline{a}$ pont egy környezetében. Ekkor a környezet bármely $\underline{x}$ pontjára igaz, hogy vannak olyan $\underline{c}_1, \ldots,$ $\underline{c}_k$ pontok, hogy*

$$f(\underline{x}) - f(\underline{a}) = f'_{x_1}(\underline{c}_1)(x_1 - a_1) + f'_{x_2}(\underline{c}_2)(x_2 - a_2) + \cdots + f'_{x_k}(\underline{c}_k)(x_k - a_k),$$

ahol $\forall\, i: \ d(\underline{c}_i, \underline{a}) < d(\underline{x}, \underline{a})$ *és persze* $\underline{x} = (x_1, \ldots, x_k)$, $\underline{a} = (a_1, \ldots, a_k)$.

2. Tétel. (ha egy függvény deriváltja nulla ...) *Legyen az f k-változós függvény egy összefüggő nyílt D tartományon olyan, hogy* $\forall\, \underline{x} \in D \ \forall\, i: \ f'_{x_i}(\underline{x}) = 0$. *Ekkor f az egész D-n állandó.*

Differenciálás

Mi a differenciálhatóság igazi jelentése? „*Kicsiben minden lineáris*” azaz a függvény lokálisan egy lineáris függvénnyel közelíthető.

Az egyváltozós esetben ezt így fogalmaztuk meg:

3. Definíció. Az f az a pontban differenciálható, ha az a egy kis környezetében értelmezve van, és létezik olyan q szám, hogy ebben a környezetben

$$f(x) = f(a) + q \cdot (x - a) + H(x), \qquad \text{ahol a } H(x) \text{ hiba nagyon kicsi, ha } x \to a,$$

pontosabban, $H(x) = \omega(x) \cdot (x - a)$, $\omega(x) \to 0$, ha $x \to a$.

Ebben a formulában $\ell(x) := f(a) + q \cdot (x - a)$ a „közelítő lineáris függvény”, azaz az érintőegyenes (lásd „adott ponton átmenő, adott meredekségű egyenes egyenlete”) és q az a-beli derivált.

Ez a definíció könnyen általánosítható.

4. Definíció. Az f k-változós függvény az $\underline{a}$ pontban („totálisan”) differenciálható, ha az $\underline{a}$ egy kis környezetében értelmezve van, és léteznek olyan q_i számok, hogy ebben a környezetben

$$f(\underline{x}) = f(\underline{a}) + q_1 \cdot (x_1 - a_1) + q_2 \cdot (x_2 - a_2) + \cdots + q_k \cdot (x_k - a_k) + H(\underline{x}),$$

ahol a $H(\underline{x})$ hiba nagyon kicsi, ha $\underline{x} \to \underline{a}$.

A „nagyon kicsi” hibatag

$$H(\underline{x}) = \omega_1(\underline{x})(x_1 - a_1) + \cdots + \omega_k(\underline{x})(x_k - a_k) = \underline{\omega}(\underline{x}) \odot (\underline{x} - \underline{a}),$$

ahol $\omega_i : \mathbb{R}^k \to \mathbb{R}$ típusú, illetve $\underline{\omega} : \mathbb{R}^k \to \mathbb{R}^k$ típusú függvények, amelyekre

$$\forall\, i: \ \omega_i(\underline{x}) \to 0 \ \ (\underline{x} \to \underline{a}), \qquad \text{vagy} \qquad \underline{\omega}(\underline{x}) \to \underline{o} \ \ (\underline{x} \to \underline{a})$$

és „ $\odot$ ” a két vektor skaláris szorzását jelöli.

Ugyanez tömörebben:

4′. Definíció. Az f k-változós függvény az $\underline{a}$ pontban differenciálható, ha az $\underline{a}$ egy kis környezetében értelmezve van, és létezik olyan $\underline{q}$ vektor, hogy ebben a környezetben

$$f(\underline{x}) = f(\underline{a}) + \underline{q} \odot (\underline{x} - \underline{a}) + H(\underline{x})\,.$$

A „nagyon kicsi” hibatag most is, mint fent, $H(\underline{x}) = \underline{\omega}(\underline{x}) \odot (\underline{x} - \underline{a})$, $\underline{\omega}(\underline{x}) \to \underline{o}$ $(\underline{x} \to \underline{a})$, kényelmesebb alakban (az ekvivalencia könnyen igazolható),

$$(*) \qquad H(\underline{x}) = \Omega(\underline{x}) \cdot \|\underline{x} - \underline{a}\|, \qquad \text{ahol } \Omega(\underline{x}) \to 0, \text{ ha } \underline{x} \to \underline{a},$$

ahol Ω egy $\mathbb{R}^k \to \mathbb{R}$ típusú függvény.

Ezekben a definíciókban $\ell(\underline{x}) := f(\underline{a}) + \underline{q} \odot (\underline{x} - \underline{a})$ a „közelítő lineáris függvény”, azaz az érintősík (hipersík, ha $k > 2$).

Könnyű belátni, hogy ha a függvény differenciálható, akkor parciálisan is differenciálható (valljuk be, ez nem hangzik meglepően), és $\forall\, i: \ q_i = f'_{x_i}(\underline{a})$.

A $\underline{q}$ vektort az f gradiensének nevezzük; tehát $\operatorname{grad} f$ egy $\mathbb{R}^k \to \mathbb{R}^k$ típusú függvény (vektor-vektor függvény). A gradiensvektor merőleges az $f(\underline{x}) = f(\underline{a})$ egyenletű nívóhalmazra (szintvonalra, szintfelületre ...).

Ez az *igazi* differenciálás. A (totális) differenciálhatóságból következik a folytonosság, a parciális és az irány szerinti deriválhatóság. Ez kell az olyan, alapvetően fontos tételekhez, mint az összetett függvény differenciálhatósága, a Young-tétel, a Taylor-formuláról, szélsőértékekről, vagy implicit függvényekről szóló tételek.

Mi a helyzet a vektorfüggvényekkel?

Most már tudjuk, mi a differenciálás; semmi sem akadályozhat meg abban, hogy a fogalmat tovább általánosítsuk.

Legyen $\vec{F} : \mathbb{R}^k \to \mathbb{R}^l$, azaz $\vec{F}$ egy olyan l-dimenziós vektorfüggvény, aminek k változója van. Nyilván írhatjuk

$$\vec{F}(\underline{x}) = \big(F_1(\underline{x}), F_2(\underline{x}), \ldots, F_l(\underline{x})\big)$$

alakban is, ahol $j = 1, 2, \ldots, l$ esetén $F_j : \mathbb{R}^k \to \mathbb{R}$, azaz F_j egy k-változós (skalár)függvény, de igazából nem ezt akarjuk csinálni.

Mi lehet a „közelítő lineáris függvény” vektorterek között? Lineáris algebrából ezeket jól ismerjük; tudjuk például, hogy ha $\mathcal{Q}$ egy lineáris transzformáció $\mathbb{R}^k$-ból $\mathbb{R}^l$-be, akkor ő megadható egy $\mathbf{Q}$ mátrixszal (aminek l sora és k oszlopa van, és balról szorozzuk vele az $\underline{x}$ oszlopvektort).

Csapjunk bele a közepébe:

5. Definíció. Az $\vec{F} : \mathbb{R}^k \to \mathbb{R}^l$ k-változós vektorfüggvény az $\underline{a} \in \mathbb{R}^k$ pontban differenciálható, ha az $\underline{a}$ egy kis környezetében értelmezve van, és létezik olyan $\mathcal{Q} : \mathbb{R}^k \to \mathbb{R}^l$ lineáris transzformáció, hogy ebben a környezetben

$$\vec{F}(\underline{x}) = \vec{F}(\underline{a}) + \mathcal{Q}(\underline{x} - \underline{a}) + \vec{H}(\underline{x})$$

ahol a $\vec{H}$ hibafüggvény $\mathbb{R}^k \to \mathbb{R}^l$ típusú, és persze „nagyon kicsi”; pontosabban, $(*)$ mintájára,

$$\frac{\|\vec{H}(\underline{x})\|_l}{\|\underline{x} - \underline{a}\|_k} \to 0\,, \qquad \text{ha} \quad \underline{x} \to \underline{a}\,,$$

itt $\|\cdot\|_k$, illetve $\|\cdot\|_l$ a k, illetve l-dimenziós vektornormát jelölik.

Ha akarjuk, kiírhatjuk a komponenseket is:

$$\begin{pmatrix} F_1(\underline{x}) \\ \vdots \\ F_l(\underline{x}) \end{pmatrix} = \begin{pmatrix} F_1(\underline{a}) \\ \vdots \\ F_l(\underline{a}) \end{pmatrix} + \begin{pmatrix} q_{11} & \ldots & q_{1k} \\ & \ddots & \\ q_{l1} & \ldots & q_{lk} \end{pmatrix} \begin{pmatrix} x_1 - a_1 \\ \vdots \\ x_k - a_k \end{pmatrix} + \vec{H}(\underline{x}).$$

Ebbe a definícióba beleférnek az eddigiek.

Tényleg? De akkor a derivált az mi? Nos, általános esetben egy lineáris transzformáció; és például az egyváltozós esetben ez egy szám, hiszen két egydimenziós vektortér között a legáltalánosabb transzformáció a számmal való szorzás.

Könnyű belátni, hogy a $\mathbf{Q}$ mátrix elemeire $q_{ij} = (F_i)'_{x_j}(\underline{a})$ teljesül.

És a magasabbrendű deriváltak?

Két speciális esetet említünk meg.

Az $\mathbb{R} \to \mathbb{R}^3$ típusú függvény: $\underline{r}(t) = \big(x(t), y(t), z(t)\big)$ egy 3 dimenziós görbevonal. A lehetséges deriváltak:

$$\underline{v} := \underline{r}'(t) = \big(x'(t), y'(t), z'(t)\big), \quad \underline{w} := \underline{r}''(t) = \big(x''(t), y''(t), z''(t)\big), \quad \ldots$$

tehát a deriváltak szintén 3-dimenziós vektorok.

Az $\mathbb{R}^k \to \mathbb{R}$ típusú függvény: $f(\underline{x}) = f(x_1, \ldots, x_k)$ a szokásos k-változós függvény. Egyszer deriválva,

$$\big(f'_{x_1}, \ldots, f'_{x_k}\big) =: \operatorname{grad} f$$

a gradiensvektor. Ezt (mint vektorfüggvényt) deriválva kapjuk a

$$\begin{pmatrix} f''_{x_1x_1} & \cdots & f''_{x_1x_k} \\ \vdots & \ddots & \vdots \\ f''_{x_kx_1} & \cdots & f''_{x_kx_k} \end{pmatrix}$$

szimmetrikus mátrixot, az ún. Hesse-féle mátrixot.

De mit jelenthetnek ezek?

Az egydimenziós esetben az első derivált az érintő meredekségét adta meg, a második derivált a konvexitást jellemezte. Gondoljunk most az (egydimenziós) konvexitásra úgy, hogy a konvex fv. lokálisan az óramutató járásával ellentétes irányba csavarodó, fölfelé nyíló, „mosolygó” ív; a konkáv pedig az óramutató járásának megfelelően csavarodó, lefelé nyíló, „szomorú” ív.

Az $\mathbb{R} \to \mathbb{R}^3$ típusú térgörbe esetében a fenti $\underline{v}$ vektor érintőirányú; amennyiben t az ívhossz-paraméter, érintőirányú egységvektor. A kétszeri deriválással kapott $\underline{w}$ vektor pedig most is azt jellemzi, hogy merre csavarodik a görbe; a görbe ugyanis lokálisan „jó közelítéssel” egy körív, és ha t az ívhosszparaméter, $\underline{w}$ a „símulókör” középpontja felé mutat és hosszának reciproka a kör sugara (de erről bővebben a differenciálgeo./alkalmazott geo. tárgyban tanulnak).

Az $\mathbb{R}^k \to \mathbb{R}$ típusú többváltozós függvény esetében a Hesse-mátrix, pontosabban a hozzá tartozó kvadratikus forma, ha definit, meghatározza, hogy a függvény képe lokálisan mosolygós vagy szomorkás ...

Tehát a függvény lokálisan, mivel folytonos, nulladik közelítésben $f(\underline{a})$-val egyenlő; mivel differenciálható, első közelítésben egy sík („kicsiben lineáris”); mivel mégegyszer differenciálható, második közelítésben mosolygós gödör vagy szomorú hegycsúcs. Bővebben lásd a Taylor-formulánál, illetve a szélsőértékek vizsgálatánál.