DIFFERENCIÁLÁS, GRADIENS VEKTOR, HESSE MÁTRIX, LÁNCSZABÁLY, IMPLICIT FÜGGVÉNY TÉTEL

DR. NAGY TAMÁS
egyetemi docens

Miskolci Egyetem
Alkalmazott Matematikai Tanszék

„A bemutatott kutató munka a TÁMOP-4.2.1.B-10/2/KONV-2010-0001 jelű projekt részeként az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg"
„This research was carried out as part of the TAMOP-4.2.1.B-10/2/KONV-2010-0001 project with support by the European Union, co-financed by the European Social Fund."

Miskolc, 2012

Tartalomjegyzék

1. Bevezetés

Az optimalizálási feladatban szereplő többváltozós valós függvények (akár a célfüggvény, akár a feltételi függvények) nagyon sok esetben differenciálható függvények. Így a differenciálhatóság az optimalizálásban nagyon fontos szerepet játszik, ezért foglalkozunk vele külön tananyagban. A tananyag a többváltozós valós függvények differenciálhatóságának kérdéseit taglalja, többek között megismerkedünk az iránymenti derivált, a gradiens vektor, a Hesse mátrix fogalmával, majd az összetett függvény deriválási szabályával, az ún. láncszabállyal, illetve az ehhez szorosan kapcsolódó ún implicit függvény tétellel.

2. Differenciálhatóság egyváltozós függvény esetén

Először az egyváltozós függvény differenciálásával foglalkozunk nagyon röviden és azt fogjuk általánosítani a többváltozós függvényekre.

Az $f : R \to R$ egyváltozós valós függvény differenciálhatósága jól ismert fogalom, mi szerint akkor mondjuk, hogy az $f(x)$ függvény az $\bar{x} \in R$ pontban differenciálható, ha létezik az alábbi határérték és az véges

$$f'(\bar{x}) = \lim_{x \to \bar{x}} \frac{f(x) - f(\bar{x})}{x - \bar{x}},$$

vagy szokás az előzővel ekvivalens definíciót is használni

$$f'(\bar{x}) = \lim_{\lambda \to 0} \frac{f(\bar{x} + \lambda) - f(\bar{x})}{\lambda}.$$

A határértéket differenciálhányadosnak, vagy deriváltnak nevezzük és ahogy látjuk $f'(\bar{x})$ szimbólummal jelöljük. Szokásos az f' helyett a $\frac{df}{dx}$ szimbólummal való jelölés is.

Geometriai értelmet is adhatunk a deriváltnak. Az $\frac{f(x)-f(\bar{x})}{x-\bar{x}}$ differenciahányados az $f(x)$ függvény képének az x és $\bar{x}$ értékeknél a függvénypontokat összekötő egyenes, nevezetesen a szelő egyenes iránytangense (iránytényezője). A derivált pedig ennek a differenciahányadosnak a határértéke, amely geometriailag a függvény $\bar{x}$ pontbeli érintő egyenesének az iránytényezője. Iránytangens (iránytényező) alatt az egyenes és az x tengely pozitív félegyenese által bezárt szög tangensét értjük.

Megkülönböztetünk jobboldali és baloldali deriváltakat, aszerint, hogy jobbról, ill. balról tartunk az $\bar{x}$ ponthoz, vagy a második értelmezésben a λ pozitív, ill. negatív értékeken keresztül tart a zérushoz. Ez utóbbi esetben a határérték jelölésben a $\lambda \to 0+$, ill. a $\lambda \to 0-$ szimbólumokat használjuk. A jobboldali és baloldali deriváltakat nevezhetjük iránymenti deriváltaknak is, mivel más-más irányból képezzük a határértéket. Például az $f(x) = |x|$ függvény az $\bar{x} = 0$ pontban nem deriválható, viszont ebben a pontban létezik jobboldali és baloldali derivált, amely rendre $+1$ és -1.

Az egyváltozós függvény differenciálására adott második formulát egyszerűen általánosíthatjuk többváltozós függvényekre is, ezt a deriváltat iránymenti deriváltnak nevezzük és az alábbi pontban fogjuk precizen definiálni. Sajnos egyik formula sem alkalmas a többváltozós függvény differenciálására.

Amennyiben az első formulát átírjuk az

$$f(x) = f(\bar{x}) + f'(\bar{x})(x - \bar{x}) + |x - \bar{x}|\,\alpha(\bar{x}, x - \bar{x})$$

alakra, akkor az $f(x)$ függvény $\bar{x} \in R$ pontbeli differenciálhatósága az alábbiak szerint fogalmazható át.

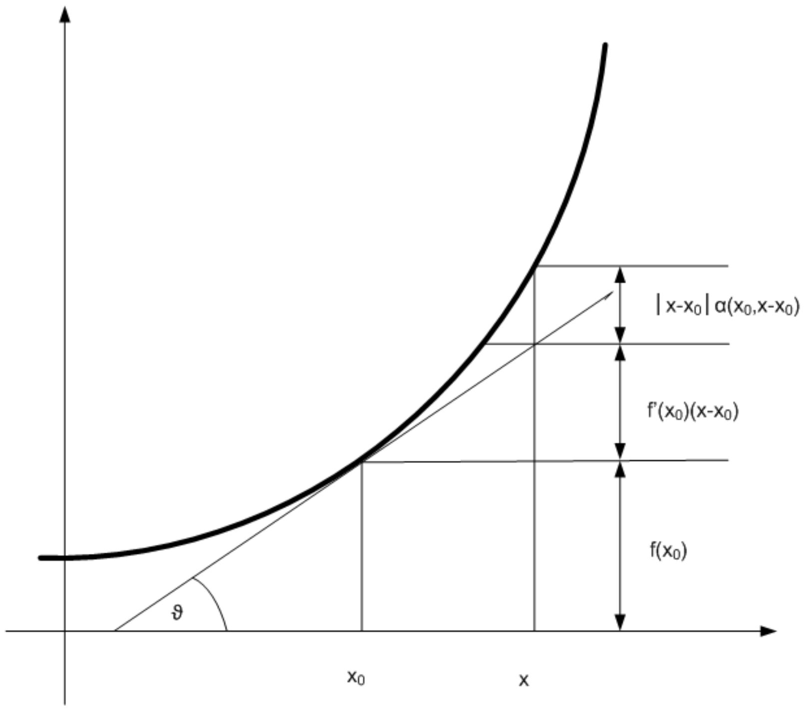
Definíció:

Az $f : R \to R$ függvény differenciálható az $\bar{x}$ pontban, ha létezik egy $f'(\bar{x}) \in R$ véges valós szám és egy $\alpha : R \to R$ függvény úgy, hogy

$$f(x) = f(\bar{x}) + f'(\bar{x})(x - \bar{x}) + |x - \bar{x}|\, \alpha(\bar{x}, x - \bar{x})$$

ahol $\lim_{x\to\bar{x}} \alpha(\bar{x}, x - \bar{x}) = 0$.

Az átfogalmazásra azért volt szükségünk, mert ez a formula már alkalmas többváltozós függvények differenciálhatóságának definiálására. Ezért vizsgáljuk meg közelebbről ezt a formulát, tekintsuk az alábbi ábrát. Az ábrán az $\bar{x}$ jelölésére az x_0 szimbólumot használtuk.



Az ábrán látható, hogy az x pontbeli függvényérték három mennyiség összegeként adódik. Az első mennyiség az $\bar{x} = x_0$ pontbeli függvényérték, a második mennyiség a deriválttal kapcsolatos, pontosabban az $\bar{x} = x_0$ pontbeli érintőegyenesen mért mennyiség, a harmadik mennyiség pedig az az érték, amely zérushoz tart, ahogy közelítünk a $\bar{x} = x_0$ ponthoz. Mint látni fogjuk a többváltozós függvények esetében is ez a három tag jelenik meg. A második tag viszont nem az érintőegyenessel kapcsolatos. Kétváltozós esetben még ábrázolhatók a viszonyok, de több változó esetében már nem. Kétváltozós esetben a második tag az ún. érintősíkon mért mennyiséget fejezi ki.

A formulából az $f'(\bar{x})$ derivált jelentése is kiolvasható. Amennyiben az x pont elegendően közel van az $\bar{x} = x_0$ ponthoz, akkor a harmadik tag elhanyagolhatóan kicsi, így az $f(x)$ és az $f(\bar{x})$ függvények különbsége ($\triangle f$) megközelítőleg a derivált és az x értékek különbségének a szorzata, képletben: $\triangle f \approx f'(\bar{x})(x - \bar{x})$.

Szokás azt is mondani, ha egy egységgel jobbra mozdulunk az $\bar{x} = x_0$ ponttól, akkor a függvény megváltozása megközelítőleg a deriválttal azonos. Itt azonban a "megközelítőleg" nem minden esetben igaz.

Gondoljunk az x^2 függvényre és legyen $\bar{x} = 3$. A derivált értéke itt $2\bar{x} = 6$. Egy egységgel jobbra mozdulva, a függvény megváltozása $\triangle f = 4^2 - 3^2 = 7$. Ez a megváltozás (7) pedig nem mondható, hogy megközelítőleg egyenlő a derivált értékével (6). Még rosszabb a helyzet magasabb kitevőjű hatványfüggvény esetén.

Gondoljunk az $\ln x$ függvényre és legyen $\bar{x} = 20$. A derivált értéke itt $\frac{1}{\bar{x}} = \frac{1}{20} = 0,05$. Egy egységgel jobbra mozdulva, a függvény megváltozása $\triangle f = \ln 21 - \ln 20 = 0.048797$. Ez a megváltozás már megközelítőleg egyenlő a derivált értékével. Ezért is szokás mondani a fentebb említetteket, mert bizonyos esetekben a derivált valóban közel van a függvényváltozáshoz. Ha például $\bar{x} = 1$, akkor már elég gyenge a közelítés.

Mint láttuk ez a felfogás nagyon hibás eredményre vezethet, ezért csak azt mondhatjuk, hogy **elegendően kicsi** $\lambda > 0$ esetén jobbra mozdulva a függvény megváltozása $\triangle f \approx f'(\bar{x}) \cdot \lambda$.

A derivált előjele viszont egyértelműen jelzi azt, hogy a függvény milyen irányban változik. Pozitív értékű derivált esetén a függvény jobbra növekszik, balra pedig csökken. Negatív értékű derivált esetén a függvény jobbra csökken, balra pedig növekszik.

Sokszor a második formula átírását használjuk, amely szerint az $f : R \to R$ függvény differenciálható az $\bar{x}$ pontban, ha létezik egy $f'(\bar{x}) \in R$ véges valós szám és egy $\alpha : R \to R$ függvény úgy, hogy

$$f(\bar{x} + \lambda) = f(\bar{x}) + \lambda f'(\bar{x}) + |\lambda|\, \alpha(\bar{x}, \lambda)$$

ahol $\lim_{\lambda \to 0} \alpha(\bar{x}, \lambda) = 0$.

3. Iránymenti derivált

Legyen $S \subseteq R^n$ egy nemüres halmaz és legyen $f : S \to R$ függvény. Legyen $\bar{\mathbf{x}} \in S$ és legyen $\mathbf{d} \in R^n, \mathbf{d} \neq \mathbf{0}$ olyan vektor, hogy $\bar{\mathbf{x}} + \lambda\mathbf{d} \in S,\ \lambda > 0$ elegendően kicsi valós számra. Az f függvénynek az $\bar{\mathbf{x}}$ pontban a $\mathbf{d}$ irány mentén vett iránymenti deriváltjának nevezzük és $f'(\bar{\mathbf{x}}, \mathbf{d})$-vel jelöljük az alábbi határértéket, ha létezik:

$$f'(\bar{\mathbf{x}}, \mathbf{d}) = \lim_{\lambda \to 0+} \frac{f(\bar{\mathbf{x}} + \lambda\mathbf{d}) - f(\bar{\mathbf{x}})}{\lambda}.$$

Az $f'(\bar{\mathbf{x}}, \mathbf{d})$ iránymenti derivált az f függvénynek a változását méri a $\mathbf{d}$ irányban. A $\mathbf{d}$ irány elvileg bármilyen vektor lehet, a változás mérése azonban csak akkor lesz hasonló az egyváltozós esethez, ha a $\mathbf{d}$ irányvektor hossza egy, azaz $\|\mathbf{d}\| = 1$. A fenti képlet számlálójában a függvény megváltozása ($\triangle f$) szerepel, így az iránymenti derivált azt fejezi ki, hogy elegendően kicsi $\lambda > 0$ esetén a függvény megváltozása a $\mathbf{d}$ irányban megközelítőleg az iránymenti derivált értékének és a λ számnak a szorzata, képletben: $\triangle f \approx f'(\bar{\mathbf{x}}, \mathbf{d}) \cdot \lambda$.

4. Differenciálhatóság többváltozós függvény esetén

4.1. Differenciálhatóság fogalma

Legyen $S \subseteq R^n$ egy nemüres halmaz, legyen $f : S \to R$ függvény. Az f függvény differenciálható az S halmaz egy $\bar{\mathbf{x}}$ belső pontjában, ha létezik egy $\boldsymbol{\nabla} f(\bar{\mathbf{x}}) \in R^n$ vektor és egy

$\alpha : R^n \to R$ függvény úgy, hogy

$$f(\mathbf{x}) = f(\bar{\mathbf{x}}) + \boldsymbol{\nabla} f(\bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}}) + \|\mathbf{x} - \bar{\mathbf{x}}\| \, \alpha(\bar{\mathbf{x}}, \mathbf{x} - \bar{\mathbf{x}})$$

minden $\mathbf{x} \in S$-re, ahol $\lim_{\mathbf{x} \to \bar{\mathbf{x}}} \alpha(\bar{\mathbf{x}}, \mathbf{x} - \bar{\mathbf{x}}) = 0$.

Megjegyzés:
A későbbiekben általában úgy használjuk a differenciálhatósági összefüggést, hogy az $\bar{\mathbf{x}}$ helyébe $\mathbf{x}$-et, $\mathbf{x}$ helyébe pedig $\mathbf{x} + \lambda\mathbf{d}$ $(\lambda \in R)$ vektort írunk, ekkor $\mathbf{x} - \bar{\mathbf{x}} = \lambda\mathbf{d}$. Ez esetben a differenciálhatósági formula az alábbi alakot ölti:

$$f(\mathbf{x} + \lambda\mathbf{d}) = f(\mathbf{x}) + \lambda \boldsymbol{\nabla} f(\mathbf{x})\mathbf{d} + |\lambda| \, \|\mathbf{d}\| \, \alpha(\mathbf{x}, \lambda\mathbf{d})$$

minden $\mathbf{d}$ vektorra és λ valós számra. A formulában szereplő határérték az alábbiak szerint alakul, $\lim_{\lambda \to 0} \alpha(\mathbf{x}, \lambda\mathbf{d}) = 0$.

Az alábbiakban a definícióban szereplő, nagyon fontos $\nabla f(\mathbf{x})$ vektort vizsgáljuk. A $\boldsymbol{\nabla} f(\mathbf{x})$ vektort gradiens vektornak nevezzük.

4.2. Gradiens vektor vizsgálata

TÉTEL (d és −d irányba vett iránymenti derivált kapcsolata)
Legyen $f : S \to R$ differenciálható függvény. Ekkor a $\mathbf{d}$ és a $-\mathbf{d}$ irányba vett iránymenti deriváltak abszolút értékben nem, csak előjelben különböznek egymástól, azaz

$$f'(\mathbf{x}, \mathbf{d}) = -f'(\mathbf{x}, -\mathbf{d}).$$

Bizonyítás
Induljunk ki a

$$f(\mathbf{x} + \lambda\mathbf{d}) = f(\mathbf{x}) + \lambda \boldsymbol{\nabla} f(\mathbf{x})\mathbf{d} + |\lambda| \, \|\mathbf{d}\| \, \alpha(\mathbf{x}, \lambda\mathbf{d})$$

differenciálhatósági összefüggésből. Átrendezés és határértékképzés után adódik, hogy

$$\lim_{\lambda \to 0+} \frac{f(\mathbf{x} + \lambda\mathbf{d}) - f(\mathbf{x})}{\lambda} = \boldsymbol{\nabla} f(\mathbf{x})\mathbf{d}.$$

Most induljunk ki a

$$f(\mathbf{x} + \lambda(-\mathbf{d})) = f(\mathbf{x}) + \lambda \boldsymbol{\nabla} f(\mathbf{x})(-\mathbf{d}) + |\lambda| \, \|-\mathbf{d}\| \, \alpha(\mathbf{x}, \lambda(-\mathbf{d}))$$

differenciálhatósági összefüggésből. Az átrendezés és határértékképzés után a következő adódik

$$\lim_{\lambda \to 0+} \frac{f(\mathbf{x} + \lambda(-\mathbf{d})) - f(\mathbf{x})}{\lambda} = -\boldsymbol{\nabla} f(\mathbf{x})\mathbf{d}.$$

A baloldal értéke az első esetben $f'(\mathbf{x}, \mathbf{d})$, a második esetben pedig $f'(\mathbf{x}, -\mathbf{d})$. A jobboldalak figyelembe vételével valóban a tételbeli $f'(\mathbf{x}, \mathbf{d}) = -f'(\mathbf{x}, -\mathbf{d})$ összefüggést kapjuk. **Q.e.d.**

A fenti tétel egy fontos megállapítást mond ki differenciálhatóság (létezik a $\boldsymbol{\nabla} f(\mathbf{x})$ vektor) esetére, nevezetesen, ha egy irányban növekszik (csökken) a függvény, akkor az ellenkező irányban ugyanolyan mértékben csökken (növekszik).

TÉTEL (az iránymenti derivált és a $\boldsymbol{\nabla} f(\mathbf{x})$ vektor kapcsolata)
Legyen $f : S \to R$ differenciálható függvény és legyen a $\mathbf{d}$ irányvektor egységhosszúságú, azaz $\|\mathbf{d}\| = 1$. Ekkor az iránymenti derivált megegyezik a $\boldsymbol{\nabla} f(\mathbf{x})$ vektor és az egységhosszúságú irányvektor skaláris szorzatával, azaz

$$f'(\mathbf{x}, \mathbf{d}) = \boldsymbol{\nabla} f(\mathbf{x})\mathbf{d}.$$

Bizonyítás
A fenti tétel bizonyításából kiolvasható az állítás, azaz

$$\lim_{\lambda \to 0+} \frac{f(\mathbf{x} + \lambda \mathbf{d}) - f(\mathbf{x})}{\lambda} = \boldsymbol{\nabla} f(\mathbf{x})\mathbf{d}.$$

Q.e.d.

TÉTEL (a $\boldsymbol{\nabla} f(\mathbf{x})$ vektor a parciális deriváltak vektora)
Legyen $f : S \to R$ differenciálható függvény. A $\boldsymbol{\nabla} f(\mathbf{x})$ vektor elemei az f függvény parciális deriváltjai, azaz

$$\boldsymbol{\nabla} f(\mathbf{x}) = \left(\frac{\partial f(\mathbf{x})}{\partial x_1}, \frac{\partial f(\mathbf{x})}{\partial x_2}, \dots, \frac{\partial f(\mathbf{x})}{\partial x_n} \right).$$

Bizonyítás
Induljunk ki a $\mathbf{d}$ és $-\mathbf{d}$ irányba vett iránymenti derivált kapcsolatát bemutató tétel bizonyításából, ahol

$$\lim_{\lambda \to 0+} \frac{f(\mathbf{x} + \lambda \mathbf{d}) - f(\mathbf{x})}{\lambda} = \boldsymbol{\nabla} f(\mathbf{x})\mathbf{d}.$$

$$\lim_{\lambda \to 0+} \frac{f(\mathbf{x} + \lambda(-\mathbf{d})) - f(\mathbf{x})}{\lambda} = -\boldsymbol{\nabla} f(\mathbf{x})\mathbf{d}.$$

Mindegyik összefüggésben írjuk a $\mathbf{d}$ vektor helyébe az i-edik ($i = 1, 2, \dots, n$) egységvektorokat, amelyeknek i-edik komponense 1, a többi pedig zérus.

$$\lim_{\lambda \to 0+} \frac{f(\mathbf{x} + \lambda \mathbf{e}_i) - f(\mathbf{x})}{\lambda} = \boldsymbol{\nabla} f(\mathbf{x})\mathbf{e}_i.$$

$$\lim_{\lambda \to 0+} \frac{f(\mathbf{x} + \lambda(-\mathbf{e}_i)) - f(\mathbf{x})}{\lambda} = -\boldsymbol{\nabla} f(\mathbf{x})\mathbf{e}_i.$$

A jobboldalakon szereplő $\boldsymbol{\nabla} f(\mathbf{x})\mathbf{e}_i$ mennyiség a $\boldsymbol{\nabla} f(\mathbf{x})$ vektor i-edik koordinátája, jelben $(\boldsymbol{\nabla} f(\mathbf{x}))_i$. A baloldalakon pedig a határértéket részletezve az alábbi adódik:

$$\lim_{\lambda \to 0+} \frac{f(\mathbf{x} + \lambda \mathbf{e}_i) - f(\mathbf{x})}{\lambda} = \lim_{\lambda \to 0+} \frac{f(x_1, x_2, \dots, x_{i-1}, x_i + \lambda, x_{i+1}, \dots, x_n) - f(\mathbf{x})}{\lambda} = (\boldsymbol{\nabla} f(\mathbf{x}))_i \,.$$

A második összefüggés baloldalát egy kis átalakítással írjuk fel

$$\begin{aligned}\lim_{\lambda\to 0+}\frac{f(\mathbf{x}+\lambda(-\mathbf{e}_i))-f(\mathbf{x})}{\lambda} &= -\lim_{\lambda\to 0+}\frac{f(x_1,x_2,\ldots,x_{i-1},x_i-\lambda,x_{i+1},\ldots,x_n)-f(\mathbf{x})}{-\lambda} = \\ &= -\lim_{\lambda\to 0-}\frac{f(x_1,x_2,\ldots,x_{i-1},x_i+\lambda,x_{i+1},\ldots,x_n)-f(\mathbf{x})}{\lambda} = \\ &= -\left(\boldsymbol{\nabla} f(\mathbf{x})\right)_i ,\end{aligned}$$

amelyből

$$\lim_{\lambda\to 0-}\frac{f(x_1,x_2,\ldots,x_{i-1},x_i+\lambda,x_{i+1},\ldots,x_n)-f(\mathbf{x})}{\lambda} = \left(\boldsymbol{\nabla} f(\mathbf{x})\right)_i .$$

A jobboldalak azonosak, a baloldali mennyiségeket pedig a közös határértékkel írhatjuk, azaz

$$\lim_{\lambda\to 0}\frac{f(x_1,x_2,\ldots,x_{i-1},x_i+\lambda,x_{i+1},\ldots,x_n)-f(\mathbf{x})}{\lambda}.$$

Ez utóbbi összefüggés az ismert definíció szerint az f függvény i-edik parciális deriváltja. A jobboldal pedig a $\boldsymbol{\nabla} f(\mathbf{x})$ vektor i-edik koordinátája. Ebből azonnal következik a tétel állítása. **Q.e.d.**

A gradiens képzés tulajdonságai:

Legyenek $f, g : R^n \to R$ differenciálható függvények; $\lambda, n \in R$ skalárok. Ekkor igazak az alábbi összefüggések

$$\begin{aligned}\boldsymbol{\nabla}\left[\lambda f(\mathbf{x})\right] &= \lambda\boldsymbol{\nabla} f(\mathbf{x}) \\ \boldsymbol{\nabla}\left[f(\mathbf{x})\pm g(\mathbf{x})\right] &= \boldsymbol{\nabla} f(\mathbf{x})\pm\boldsymbol{\nabla} g(\mathbf{x}) \\ \boldsymbol{\nabla}\left[f(\mathbf{x})\cdot g(\mathbf{x})\right] &= g(\mathbf{x})\cdot\boldsymbol{\nabla} f(\mathbf{x})+f(\mathbf{x})\cdot\boldsymbol{\nabla} g(\mathbf{x}) \\ \boldsymbol{\nabla}\left[\frac{f(\mathbf{x})}{g(\mathbf{x})}\right] &= \frac{g(\mathbf{x})\cdot\boldsymbol{\nabla} f(\mathbf{x})-f(\mathbf{x})\cdot\boldsymbol{\nabla} g(\mathbf{x})}{\left[g(\mathbf{x})\right]^2} \\ \boldsymbol{\nabla}\left[f(\mathbf{x})\right]^n &= n\left[f(\mathbf{x})\right]^{n-1}\boldsymbol{\nabla} f(\mathbf{x})\end{aligned}$$

4.3. Gradiens vektor tulajdonságai

TÉTEL (a gradiens vektor tulajdonságai)

Legyen $f : S \to R$ differenciálható függvény.

a) A $\boldsymbol{\nabla} f(\mathbf{x})$ vektor iránya az f függvény legnagyobb növekedési irányát jelöli ki,
b) a $\boldsymbol{\nabla} f(\mathbf{x})$ vektor hossza ($\|\boldsymbol{\nabla} f(\mathbf{x})\|$) az f függvény legnagyobb növekedésének mértékét adja,
c) a $\boldsymbol{\nabla} f(\mathbf{x})$ vektor az $\mathbf{x}$ ponton átmenő $f(\mathbf{x}) =$ *állandó* szintfelület érintősíkjának normálisa.

Bizonyítás

a) Az

$$f'(\mathbf{x},\mathbf{d}) = \boldsymbol{\nabla} f(\mathbf{x})\mathbf{d}$$

alapképletben a jobboldalon lévő skaláris szorzat akkor a legnagyobb, ha a $\mathbf{d}$ irányvektor párhuzamos a $\boldsymbol{\nabla} f(\mathbf{x})$ vektorral, tehát a gradiens vektor az f függvény legnagyobb növekedési irányába mutat.

b) Ha a $\mathbf{d}$ irányvektor a gradiens vektor irányába mutató egység hooszúságú vektor, azaz $\mathbf{d} = \frac{\nabla f(\mathbf{x})}{\|\nabla f(\mathbf{x})\|}$, akkor $f'(\mathbf{x}, \mathbf{d}) = \|\nabla f(\mathbf{x})\|$, amelyből nyilvánvaló az állítás.

c) Az alapképlet szerint, ha a gradiens vektor merőleges az irányvektorra, akkor a $\nabla f(\mathbf{x})\mathbf{d}$ skalárszorzat zérus, tehát az f függvény ebben az irányban nem változik. Ez az irány nem más, mint az $\mathbf{x}$ ponton áthaladó szintfelület (azonos függvényértékű pontok halmaza) normálisa. **Q.e.d.**

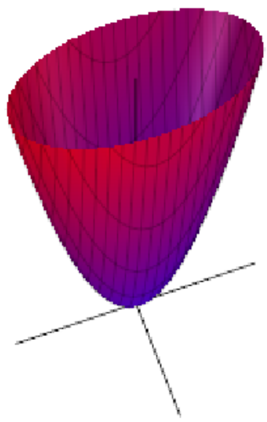
4.4. Példa gradiens vektorra

Példa:

Tekintsük az alábbi kétváltozós függvényt!

$$f(x_1, x_2) = x_1^2 + 4x_2^2$$

A kétváltozós függvény képe az R^3 térben egy felület, amelyet az alábbi ábra mutat.



Vizsgáljuk meg a gradiens vektort az $\bar{\mathbf{x}} = (3, 1)$ pontban.

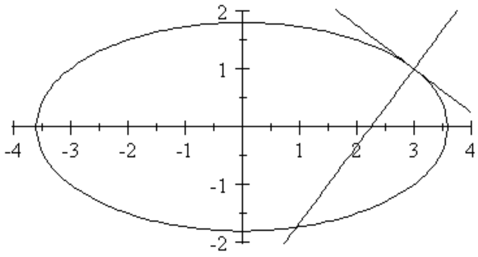
$$\nabla f(\mathbf{x}) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2} \right) = (2x_1, 8x_2).$$

Az adott pontban $f(\bar{\mathbf{x}}) = 13$, $\nabla f(\bar{\mathbf{x}}) = (6, 8)$, $\|\nabla f(\bar{\mathbf{x}})\| = \sqrt{6^2 + 8^2} = 10$. Ezek jelentése a következő. Az $\bar{\mathbf{x}} = (3, 1)$ pontban a függvény értéke 13 és ez a függvényérték a $\nabla f(\bar{\mathbf{x}}) = (6, 8)$ irányban történő elmozdulás esetén növekszik a legnagyobbat. A $\|\nabla f(\bar{\mathbf{x}})\| = 10$ pedig a növekedés mértékét jelenti. Ha a $\nabla f(\bar{\mathbf{x}}) = (6, 8)$ irányban elmozdulunk egy elegendően kicsit ($\lambda > 0$), akkor a függvény növekedése ($\triangle f$) megközelítőleg $10 \cdot \lambda$, a függvény értéke pedig megközelítőleg $13 + 10 \cdot \lambda$.

Az $\bar{\mathbf{x}} = (3, 1)$ pontban a függvényérték 13, így ehhez a ponthoz tartozó szintfelület az $f(x_1, x_2) = x_1^2 + 4x_2^2 = 13$ egyenletet kielégítő görbe. Az egyenlet átalakított alakja:

$$\frac{x_1^2}{\left(\sqrt{13}\right)^2} + \frac{x_2^2}{\left(\frac{\sqrt{13}}{2}\right)^2} = 1,$$

azaz a görbe esetünkben egy origó centrumú ellipszis. A ellipszis x_1 irányú féltengelyének hossza $\sqrt{13}$, x_2 irányú féltengelyének hossza pedig $\frac{\sqrt{13}}{2}$. Az alábbi ábra mutatja az R^2 síkon a szintgörbét és az $\bar{\mathbf{x}} = (3, 1)$ pontban a gradiens vektor irányát. Láthatjuk, hogy a gradiens vektor merőleges a szintgörbe $\bar{\mathbf{x}} = (3, 1)$ pontbeli érintőjére.



5. Kétszer differenciálhatóság többváltozós függvény esetén

5.1. Kétszer differenciálhatóság fogalma

Legyen $S \subseteq R^n$ egy nemüres halmaz, legyen $f : S \to R$ függvény. Az f függvényt kétszer differenciálhatónak mondunk az S halmaz egy $\bar{\mathbf{x}}$ belső pontjában, ha létezik egy $\boldsymbol{\nabla} f(\bar{\mathbf{x}}) \in R^n$ vektor, egy $\mathbf{H}(\bar{\mathbf{x}}) \in R^{n\times n}$ szimmetrikus mátrix és egy $\alpha : R^n \to R$ függvény úgy, hogy

$$f(\mathbf{x}) = f(\bar{\mathbf{x}}) + \boldsymbol{\nabla} f(\bar{\mathbf{x}})(\mathbf{x}-\bar{\mathbf{x}}) + \frac{1}{2}(\mathbf{x}-\bar{\mathbf{x}})\mathbf{H}(\bar{\mathbf{x}})(\mathbf{x}-\bar{\mathbf{x}}) + \|\mathbf{x}-\bar{\mathbf{x}}\|^2 \alpha(\bar{\mathbf{x}}, \mathbf{x}-\bar{\mathbf{x}})$$

minden $\mathbf{x} \in S$-re, ahol $\lim_{\mathbf{x}\to\bar{\mathbf{x}}} \alpha(\bar{\mathbf{x}}, \mathbf{x}-\bar{\mathbf{x}}) = 0$.

A $\mathbf{H}(\bar{\mathbf{x}})$ mátrixot Hesse mátrixnak nevezzük, elemei az f függvény másodrendű parciális deriváltjai, azaz

$$\mathbf{H}(\mathbf{x}) = \begin{bmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_1} & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_1} & \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_1} & \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_n} \end{bmatrix},$$

más szóvan úgy is mondhatjuk, hogy a Hesse mátrix i-edik sorvektora a $\boldsymbol{\nabla} f(\mathbf{x})$ gradiens vektor i-edik koordinátájának $\left(\frac{\partial f(\mathbf{x})}{\partial x_i}\right)$, mint valós függvénynek a gradiense. Éppen ezért a Hesse mátrixot nagyon gyakran $\boldsymbol{\nabla}^2 f(\mathbf{x})$ szimbólummal jelöljük. Mivel a vegyes másodrendű parciális deriváltak megegyeznek, így a Hesse mátrix szimmetrikus.

5.2. Példa Hesse mátrixra

Példa:

Tekintsük az alábbi háromváltozós függvényt!

$$f(x_1, x_2) = 2x_1^3 x_2^2 x_3 + \frac{x_2 x_3}{x_1^2}$$

Határozzuk meg a gradiens vektort és a Hesse mátrixot az $\bar{\mathbf{x}} = (1, 2, 3)$ pontban.

A Hesse mátrix könnyebb előállítása miatt a gradiens vektort oszlopvektorként adjuk meg.

$$\nabla f(\mathbf{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \frac{\partial f}{\partial x_3} \end{bmatrix} = \begin{bmatrix} 6x_1^2x_2^2x_3 - \frac{2x_2x_3}{x_1^3} \\ 4x_1^3x_2x_3 + \frac{x_3}{x_1^2} \\ 2x_1^3x_2^2 + \frac{x_2}{x_1^2} \end{bmatrix},$$

$$\begin{aligned} \mathbf{H}(\mathbf{x}) &= \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_1 \partial x_3} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_2} & \frac{\partial^2 f}{\partial x_2 \partial x_3} \\ \frac{\partial^2 f}{\partial x_3 \partial x_1} & \frac{\partial^2 f}{\partial x_3 \partial x_2} & \frac{\partial^2 f}{\partial x_3 \partial x_3} \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial x_1}(\frac{\partial f}{\partial x_1}) & \frac{\partial}{\partial x_2}(\frac{\partial f}{\partial x_1}) & \frac{\partial}{\partial x_3}(\frac{\partial f}{\partial x_1}) \\ \frac{\partial}{\partial x_1}(\frac{\partial f}{\partial x_2}) & \frac{\partial}{\partial x_2}(\frac{\partial f}{\partial x_2}) & \frac{\partial}{\partial x_3}(\frac{\partial f}{\partial x_2}) \\ \frac{\partial}{\partial x_1}(\frac{\partial f}{\partial x_3}) & \frac{\partial}{\partial x_2}(\frac{\partial f}{\partial x_3}) & \frac{\partial}{\partial x_3}(\frac{\partial f}{\partial x_3}) \end{bmatrix} = \\ &= \begin{bmatrix} \nabla(\frac{\partial f}{\partial x_1}) \\ \nabla(\frac{\partial f}{\partial x_2}) \\ \nabla(\frac{\partial f}{\partial x_3}) \end{bmatrix} = \begin{bmatrix} 12x_1x_2^2x_3 + \frac{6x_2x_3}{x_1^4} & 12x_1^2x_2x_3 - \frac{2x_3}{x_1^3} & 6x_1^2x_2^2 - \frac{2x_2}{x_1^3} \\ 12x_1^2x_2x_3 - \frac{2x_3}{x_1^3} & 4x_1^3x_3 & 4x_1^3x_2 + \frac{1}{x_1^2} \\ 6x_1^2x_2^2 - \frac{2x_2}{x_1^3} & 4x_1^3x_2 + \frac{1}{x_1^2} & 0 \end{bmatrix}. \end{aligned}$$

Az adott $\bar{\mathbf{x}} = (1, 2, 3)$ pontban a gradiens vektor és a Hesse mátrix értéke az alábbi:

$$\nabla f(\bar{\mathbf{x}}) = \begin{bmatrix} 60 \\ 27 \\ 10 \end{bmatrix}, \qquad \mathbf{H}(\bar{\mathbf{x}}) = \begin{bmatrix} 180 & 66 & 20 \\ 66 & 12 & 9 \\ 20 & 9 & 0 \end{bmatrix}.$$

Megjegyzés:
Hasonlóan a differenciálhatóságnál látottakhoz, a kétszer differenciálhatóság esetében is gyakran használatos az alábbi formula:

$$f(\mathbf{x} + \lambda\mathbf{d}) = f(\bar{\mathbf{x}}) + \lambda\nabla f(\mathbf{x})\mathbf{d} + \frac{1}{2}\lambda^2\mathbf{d}\mathbf{H}(\mathbf{x})\mathbf{d} + \lambda^2 \left\|\mathbf{d}\right\|^2 \alpha(\mathbf{x}, \lambda\mathbf{d})$$

minden $\mathbf{d}$ vektorra és λ valós számra. A formulában szereplő határérték az alábbiak szerint alakul, $\lim_{\lambda\to 0} \alpha(\bar{\mathbf{x}}, \lambda\mathbf{d}) = 0$.

A diifferenciálhatóságot mint láttuk egy belső pontban definiáltuk. Az f függvényt differenciálhatónak mondunk az $S \subseteq R^n$ halmazon, ha az f függvény az S halmaz minden belső pontjában differenciálható. Ez természetesen a kétszer differenciálhatóságra is vonatkozik.

5.3. Gyakori függvények gradiense és Hesse mátrixa

Néhány gyakran használt függvény gradiense és Hesse mátrixa:

	$f(\mathbf{x})$	$\nabla f(\mathbf{x})$	$\nabla^2 f(\mathbf{x})$
1.	$\mathbf{c}^T\mathbf{x}$	$\mathbf{c}$	$\mathbf{0}$
2.	$\mathbf{x}^2 = \mathbf{x}^T\mathbf{x}$	$2\mathbf{x}$	$2\mathbf{E}$
3.	$\mathbf{x}^T\mathbf{A}\mathbf{x}$	$(\mathbf{A} + \mathbf{A}^T)\mathbf{x}$	$\mathbf{A}^T + \mathbf{A}$
4.	$(\mathbf{A}\mathbf{x})^2 = (\mathbf{A}\mathbf{x})^T(\mathbf{A}\mathbf{x}) = \mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x}$	$2\mathbf{A}^T\mathbf{A}\mathbf{x}$	$2\mathbf{A}^T\mathbf{A}$
5.	$\mathbf{c}^T\mathbf{A}\mathbf{x}$	$\mathbf{A}^T\mathbf{c}$	0

Megjegyezzük, hogy a 3. függvénynél, ha az $\mathbf{A}$ mátrix szimmetrikus, ekkor a gradiens $2\mathbf{A}\mathbf{x}$, a Hesse mátrix pedig $2\mathbf{A}$. A Hesse mátrix meghatározásánál nem valósértékű függvény deriváltját kell meghatározni, hanem egy vektorértékű függvényét. Gyakran használt függvény például az $f(\mathbf{x}) = \mathbf{A}\mathbf{x}$ vektorértékű függvény, a példában a 2., 3., 4. gradiens vektor ilyen függvény. Ennek a komponensei valósértékű lineáris függvények, amelynek a gradiensét viszont már könnyű számítani. Egy kis meggondolás után kapjuk, hogy $\boldsymbol{\nabla}(\mathbf{A}\mathbf{x}) = \mathbf{A}^T$. Ebből már a példabeli Hesse mátrixok adódnak.

6. Középérték tétel, Taylor tétel

Az alábbi tétel az egyszer ill. a kétszer differenciálható függvényekre ad meg egy-egy fontos összefüggést. A tétel középérték tételként ill. Taylor tételként ismert.

TÉTEL (Középérték tétel, Taylor tétel)
Legyen $S \subseteq R^n$ egy nemüres, nyílt, konvex halmaz, legyen $f : S \to R$ függvény.
Ha az f függvény differenciálható az S halmazon, akkor minden $\mathbf{x} \in S$ és $\bar{\mathbf{x}} \in S$ vektorok esetén létezik olyan $\hat{\mathbf{x}} \in S$ vektor, hogy

$$f(\mathbf{x}) = f(\bar{\mathbf{x}}) + \boldsymbol{\nabla} f(\hat{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}}),$$

ahol $\hat{\mathbf{x}}$ vektor az $\mathbf{x}$ és $\bar{\mathbf{x}}$ vektorokat összekötő szakasz valamely belső pontja.
Ha az f függvény kétszer differenciálható az S halmazon, akkor minden $\mathbf{x} \in S$ és $\bar{\mathbf{x}} \in S$ vektorok esetén létezik olyan $\hat{\mathbf{x}} \in S$ vektor, hogy

$$f(\mathbf{x}) = f(\bar{\mathbf{x}}) + \boldsymbol{\nabla} f(\bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}}) + \frac{1}{2}(\mathbf{x} - \bar{\mathbf{x}})\mathbf{H}(\hat{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})$$

ahol $\hat{\mathbf{x}}$ vektor az $\mathbf{x}$ és $\bar{\mathbf{x}}$ vektorokat összekötő szakasz valamely belső pontja.

Az alábbi két fejezetben két fontos összefüggést ismertetünk a diifferenciálással kapcsolatban, az egyik az összetett függvény deriválási szabálya, az ún. láncszabály, a másik pedig ezzel szorosan kapcsolatos az ún. implicit függvény tétel. A láncszabályt is és az implicit függvény tételt is több alkalmazási környezetben mondjuk ki, elősegítve ezzel a könnyebb alkalmazását.

7. Láncszabály

TÉTEL (Láncszabály 1)
Legyen $f : R^n \to R$ differenciálható függvény. Legyen továbbá $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in R^n$ vektor és $p \in R$ skalár. Használjuk a függvényre az $f(x_1, x_2, \ldots, x_n)$ jelölést. Tegyük fel, hogy az $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in R^n$ vektor a p skalár valamilyen differenciálható függvénye, azaz

$$\begin{aligned} x_1 &= x_1(p), \\ x_2 &= x_2(p), \\ &\vdots \\ x_n &= x_n(p). \end{aligned}$$

Ekkor az f függvénynek a p skalár szerinti deriváltja, amelyet f'_p szimbólummal jelölünk, a következőképpen számítható:

$$f'_p = \frac{\partial f}{\partial x_1}\frac{dx_1}{dp} + \frac{\partial f}{\partial x_2}\frac{dx_2}{dp} + \ldots + \frac{\partial f}{\partial x_n}\frac{dx_n}{dp}.$$

Amennyiben a $\frac{\partial f}{\partial x_j}(j = 1, \ldots, n)$ mennyiségeket a $\nabla f \in R^n$ vektorba, a $\frac{\partial x_j}{\partial p}$ $(j = 1, \ldots, n)$ mennyiségeket az $\mathbf{x}'_p \in R^n$ vektorba foglaljuk, akkor a fenti összefüggés az alábbi egyszerűbb formában írható fel

$$f'_p = \nabla f \cdot \mathbf{x}'_p.$$

A láncszabályban megfogalmazott állítás szavakban megfogalmazva a következőt jelenti. Az f függvény az x_j mennyiségeken keresztül **közvetve** függ a p skalártól. Mint tudjuk, a derivált a p skalár elemi megváltozásához tartozó függvényváltozást jelenti. Amikor a p megváltozik, akkor megváltoznak az x_j mennyiségek is, amelyek az f függvénykapcsolat alapján hatást gyakorolnak az f függvényre. A láncszabály szerint a teljes hatás az x_j mennyiségek megváltozásain keresztül gyakorolt **közvetett hatások** összegeként adódik, tehát az egyes hatások szuperponálódnak.

TÉTEL (Láncszabály 2)

Legyen $f : R^{n+1} \to R$ differenciálható függvény. Legyen továbbá $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in R^n$ vektor és $p \in R$ skalár. Használjuk a függvényre az $f(x_1, x_2, \ldots, x_n, p)$ jelölést. Tegyük fel, hogy az $\mathbf{x}$ vektor a p skalár valamilyen differenciálható függvénye, azaz

$$\begin{aligned} x_1 &= x_1(p), \\ x_2 &= x_2(p), \\ &\vdots \\ x_n &= x_n(p). \end{aligned}$$

Ekkor az f függvénynek a p skalár szerinti deriváltja a következőképpen számítható:

$$f'_p = \frac{\partial f}{\partial p} + \frac{\partial f}{\partial x_1}\frac{dx_1}{dp} + \frac{\partial f}{\partial x_2}\frac{dx_2}{dp} + \ldots + \frac{\partial f}{\partial x_n}\frac{dx_n}{dp}.$$

Amennyiben a $\frac{\partial f}{\partial x_j}(j = 1, \ldots, n)$ mennyiségeket a $\nabla_{\mathbf{x}} f \in R^n$ vektorba, a $\frac{\partial x_j}{\partial p}$ $(j = 1, \ldots, n)$ mennyiségeket az $\mathbf{x}'_p \in R^n$ vektorba foglaljuk, akkor a fenti összefüggés az alábbi egyszerűbb formában írható fel

$$f'_p = \frac{\partial f}{\partial p} + \nabla_{\mathbf{x}} f \cdot \mathbf{x}'_p.$$

Ebben az esetben az f függvény nemcsak az x_j mennyiségeken keresztül **közvetve** függ a p skalártól, hanem **közvetlenül is**. A láncszabály most a közvetlen és a közvetett hatásokat fejezi ki. A teljes hatás közvetlen és a közvetett hatások összege. A fenti összefüggés egyszerűen belátható, ha az $x_{n+1} = x_{n+1}(p) = p$ választással élünk és alkalmazzuk a láncszabály első formuláját.

Megjegyezzük, hogy az $\frac{\partial f}{\partial x_j}$ mennyiségek vektorára azért használtuk a $\nabla_{\mathbf{x}} f$ jelölést, mert azt akartuk kihangsúlyozni, hogy az f függvény gradiens vektorának csak az első n eleme szerepel, azaz csak az x típusú változók szerinti parciális deriváltakat tartalmazza.

TÉTEL (Láncszabály 3, általános formula)

Legyen $f : R^{n+m} \to R$ differenciálható függvény. Legyen továbbá $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in R^n$ vektor és legyen $\mathbf{p} = (p_1, p_2, \ldots, p_m) \in R^m$ vektor. Az $f(x_1, x_2, \ldots, x_n, p_1, p_2, \ldots, p_m)$ jelölést használjuk a függvényre. Tegyük fel, hogy az $\mathbf{x}$ vektor a $p_1, p_2, \ldots, p_m$ skalárok valamilyen differenciálható függvénye, azaz

$$\begin{aligned} x_1 &= x_1(p_1, p_2, \ldots, p_m), \\ x_2 &= x_2(p_1, p_2, \ldots, p_m), \\ &\vdots \\ x_n &= x_n(p_1, p_2, \ldots, p_m). \end{aligned}$$

Ekkor az f függvénynek a p_i $(i = 1, 2, \ldots, m)$ skalár szerinti deriváltja a következőképpen számítható:

$$f'_{p_i} = \frac{\partial f}{\partial p_i} + \frac{\partial f}{\partial x_1}\frac{\partial x_1}{\partial p_i} + \frac{\partial f}{\partial x_2}\frac{\partial x_2}{\partial p_i} + \ldots + \frac{\partial f}{\partial x_n}\frac{\partial x_n}{\partial p_i} \qquad \text{minden } i = 1, 2, \ldots, m \text{ esetén.}$$

Foglaljuk az f'_{p_i} $(i = 1, \ldots, m)$ mennyiségeket az $\mathbf{f}'_{\mathbf{p}} \in R^m$ vektorba, a $\frac{\partial f}{\partial p_i}$ $(i = 1, \ldots, m)$ mennyiségeket a $\nabla_{\mathbf{p}} f \in R^m$ vektorba, a $\frac{\partial f}{\partial x_j}(j = 1, \ldots, n)$ mennyiségeket a $\nabla_{\mathbf{x}} f \in R^n$ vektorba, valamint a $\frac{\partial x_j}{\partial p_i}$ $(i = 1, \ldots, m,\ j = 1, \ldots, n)$ mennyiségeket a $\mathbf{J}_{\mathbf{x},\mathbf{p}} \in R^{m \times n}$ ún. Jacobi-mátrixba. A Jacobi-mátrix tehát az alábbi

$$\mathbf{J}_{\mathbf{x},\mathbf{p}} = \begin{bmatrix} \frac{\partial x_1}{\partial p_1} & \frac{\partial x_2}{\partial p_1} & \cdots & \frac{\partial x_n}{\partial p_1} \\ \frac{\partial x_1}{\partial p_2} & \frac{\partial x_2}{\partial p_2} & \cdots & \frac{\partial x_n}{\partial p_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_1}{\partial p_m} & \frac{\partial x_2}{\partial p_m} & \cdots & \frac{\partial x_n}{\partial p_m} \end{bmatrix}.$$

Ekkor a fenti összefüggés az alábbi egyszerűbb formában írható fel:

$$\mathbf{f}'_{\mathbf{p}} = \nabla_{\mathbf{p}} f + \mathbf{J}_{\mathbf{x},\mathbf{p}} \cdot \nabla_{\mathbf{x}} f.$$

Példa:
Legyen $n = 2$, $m = 1$. Legyen

$$\begin{aligned} f(x_1, x_2, p) &= x_1^2 x_2^3 p^2 \\ x_1 &= p + p^2 \\ x_2 &= \sqrt{p} \end{aligned}$$

Példánkban a Lácszabály 2 formulát kell alkalmazni. A számított értékek:

$$\begin{aligned} \frac{\partial f}{\partial p} &= 2x_1^2 x_2^3 p \\ \nabla_{\mathbf{x}} f &= \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 2x_1 x_2^3 p^2 \\ 3x_1^2 x_2^2 p^2 \end{bmatrix} \\ \mathbf{x}'_p &= \begin{bmatrix} \frac{\partial x_1}{\partial p} \\ \frac{\partial x_2}{\partial p} \end{bmatrix} = \begin{bmatrix} 1 + 2p \\ \frac{1}{2\sqrt{p}} \end{bmatrix} \end{aligned}$$

A keresett p szerinti derivált:

$$f'_p = \frac{\partial f}{\partial p} + \nabla_{\mathbf{x}} f \cdot \mathbf{x}'_p = 2x_1^2 x_2^3 p + 2x_1 x_2^3 p^2(1+2p) + 3x_1^2 x_2^2 p^2 \frac{1}{2\sqrt{p}}.$$

Számítsuk ki egy adott p értékre a deriváltat, legyen $p = 1$, ekkor $x_1 = 2$, $x_2 = 1$, a derivált pedig

$$f'_p = 8 + 12 + 6 = 26.$$

A közvetlen hatás 8, a közvetett hatás együttesen 18, amelyben az x_1 megváltozásának hatása 12, az x_2 megváltozásának hatása pedig 6. Az olvasó elvégezhet egy ellenőrzést úgy, hogy az f függvénybe behelyettesíti az x_1, x_2 függvényeket és ezt a függvényt, amely már csak p-től függ, deriválja. Ekkor is megkapja a deriváltra a 26 értéket, de azt nem tudja megmondani, hogy mik voltak az egyes hatások.

A teljes hatás, azaz a 26 értékű derivált jelentése az alábbi: Ha a p értéket 1-ről elegendően kicsi λ-val megváltoztatjuk, akkor a függvény értékének megváltozása megközelítőleg $26 \cdot \lambda$, ebből a közvetlen hatás eredménye $8 \cdot \lambda$, a közvetetté pedig $18 \cdot \lambda$, amelynek megoszlása $12 \cdot \lambda$ ill. $6 \cdot \lambda$.

8. Implicit függvény tétel

TÉTEL (Implicit függvény tétel 1)

Legyenek $f_1, f_2, \ldots, f_n : R^{n+1} \to R$ folytonosan differenciálható függvények. Legyen továbbá $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in R^n$ vektor és $p \in R$ skalár. Használjuk az $f_i(x_1, x_2, \ldots, x_n, p)$ $(i = 1, 2, \ldots, n)$ jelölést. Legyen az $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in R^n$ vektor és a $p \in R$ skalár között definiálva az alábbi egyenletrendszer

$$\begin{aligned} f_1(x_1, x_2, \ldots, x_n, p) &= 0 \\ f_2(x_1, x_2, \ldots, x_n, p) &= 0 \\ &\vdots \\ f_n(x_1, x_2, \ldots, x_n, p) &= 0 \end{aligned}$$

Definiáljuk a $\mathbf{J_x} \in R^{n \times n}$ Jacobi-mátrixot, amely i-edik sorának j-edik eleme $\frac{\partial f_i}{\partial x_j}$, úgy is fogalmazhatnánk, hogy az i-edik sor az f_i függvény gradiens vektorának első n eleme, tehát csak az x típusú változók szerinti parciális deriváltakat tartalmazza, ezért szerepel az $\mathbf{x}$ szimbólum a $\mathbf{J_x}$ jelölésben. A Jacobi-mátrix tehát

$$\mathbf{J_x} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix}.$$

Tegyük fel, hogy egy $(\bar{\mathbf{x}}, \bar{p})$ pont kielégíti az egyenletrendszer egyenleteit, azaz $f_i(\bar{\mathbf{x}}, \bar{p}) = 0$ $(i = 1, 2, \ldots, n)$. Az alábbi két állítást fogalmazhatjuk meg.

a) Ha az $(\bar{\mathbf{x}}, \bar{p})$ pontban a $\mathbf{J_x}$ Jacobi-mátrix determinánsa nem zérus ($\mathbf{J_x}$ invertálható), akkor az $(\bar{\mathbf{x}}, \bar{p})$ pont környezetében léteznek az alábbi folytonosan differenciálható függvények

$$\begin{aligned} x_1 &= x_1(p) \\ x_2 &= x_2(p) \\ &\vdots \\ x_n &= x_n(p) \end{aligned}$$

és az $(\bar{\mathbf{x}}, \bar{p})$ pont környezetében igaz, hogy

$$\begin{aligned} f_1(x_1(p), x_2(p), \ldots, x_n(p), p) &= 0 \\ f_2(x_1(p), x_2(p), \ldots, x_n(p), p) &= 0 \\ &\vdots \\ f_n(x_1(p), x_2(p), \ldots, x_n(p), p) &= 0 \end{aligned}$$

b) Az $(\bar{\mathbf{x}}, \bar{p})$ pontban az $x_i(p)$ $(i = 1, 2, \ldots, n)$ függvények p szerinti deriváltjára igaz, hogy

$$\begin{bmatrix} \frac{\partial x_1}{\partial p} \\ \frac{\partial x_2}{\partial p} \\ \vdots \\ \frac{\partial x_n}{\partial p} \end{bmatrix} = - \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial f_1}{\partial p} \\ \frac{\partial f_2}{\partial p} \\ \vdots \\ \frac{\partial f_n}{\partial p} \end{bmatrix}.$$

Amennyiben a $\frac{\partial x_i}{\partial p}$ $(i = 1, \ldots, n)$ mennyiségeket az $\mathbf{x}'_p \in R^n$ vektorba, a $\frac{\partial f_i}{\partial p}$ $(i = 1, \ldots, n)$ mennyiségeket az $\mathbf{f}'_p \in R^n$ vektorba, a $\frac{\partial f_i}{\partial x_j}(i = 1, \ldots, n,\ j = 1, \ldots, n)$ mennyiségeket a már ismertetett $\mathbf{J_x} \in R^{n \times n}$ Jacobi-mátrixba foglaljuk, akkor a fenti összefüggés az alábbi egyszerűbb formában írható fel

$$\mathbf{x}'_p = -\mathbf{J_x}^{-1} \cdot \mathbf{f}'_p.$$

Bizonyítás:

A b) részt bizonyítjuk, hogy lássuk a láncszabály egyik alkalmazását. Vegyük mindegyik egyenletnek a p szerinti parciális deriváltját. A baloldal deriváltja a Láncszabály 2 formula szerint írható fel, a jobboldal pedig zérus, mivel konstans deriváltja zérus

$$\frac{\partial f_i}{\partial p} + \frac{\partial f_i}{\partial x_1}\frac{dx_1}{dp} + \frac{\partial f_i}{\partial x_2}\frac{dx_2}{dp} + \ldots + \frac{\partial f_i}{\partial x_n}\frac{dx_n}{dp} = 0, \qquad (i = 1, 2, \ldots, n)$$

A fenti egyenleteket az alábbi egyszerűbb formában írhatjuk:

$$\begin{bmatrix} \frac{\partial f_1}{\partial p} \\ \frac{\partial f_2}{\partial p} \\ \vdots \\ \frac{\partial f_n}{\partial p} \end{bmatrix} + \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix} \begin{bmatrix} \frac{\partial x_1}{\partial p} \\ \frac{\partial x_2}{\partial p} \\ \vdots \\ \frac{\partial x_n}{\partial p} \end{bmatrix} = \mathbf{0},$$

amely a bevezetett mátrix-vektor jelölésekkel még egyszerűbben írható, azaz

$$\mathbf{f}'_p + \mathbf{J_x} \cdot \mathbf{x}'_p = \mathbf{0}.$$

Mivel a Jacobi-mátrix invertálható, így azonnal adódik a tételben szereplő formula.

TÉTEL (Implicit függvény tétel 2, általános formula)

Legyenek $f_1, f_2, \ldots, f_n : R^{n+m} \to R$ folytonosan differenciálható függvények. Legyen továbbá az $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in R^n$ vektor és a $\mathbf{p} = (p_1, p_2, \ldots, p_m) \in R^m$ vektor. Használjuk az $f_i(x_1, x_2, \ldots, x_n, p_1, p_2, \ldots, p_m)$ $(i = 1, 2, \ldots, n)$ jelölést. Legyen az $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in R^n$ és a $\mathbf{p} = (p_1, p_2, \ldots, p_m) \in R^m$ vektorok között definiálva az alábbi egyenletrendszer

$$\begin{aligned} f_1(x_1, x_2, \ldots, x_n, p_1, p_2, \ldots, p_m) &= 0 \\ f_2(x_1, x_2, \ldots, x_n, p_1, p_2, \ldots, p_m) &= 0 \\ &\vdots \\ f_n(x_1, x_2, \ldots, x_n, p_1, p_2, \ldots, p_m) &= 0 \end{aligned}$$

Definiáljuk a $\mathbf{J_x} \in R^{n\times n}$ Jacobi-mátrixot, amely i-edik sorának j-edik eleme $\frac{\partial f_i}{\partial x_j}$, azaz

$$\mathbf{J_x} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix}.$$

Tegyük fel, hogy egy $(\bar{\mathbf{x}}, \bar{\mathbf{p}})$ pont kielégíti az egyenletrendszer egyenleteit, azaz $f_i(\bar{\mathbf{x}}, \bar{\mathbf{p}}) = 0$ $(i = 1, 2, \ldots, n)$. Az alábbi két állítást fogalmazhatjuk meg.

a) Ha az $(\bar{\mathbf{x}}, \bar{\mathbf{p}})$ pontban a $\mathbf{J_x}$ Jacobi-mátrix determinánsa nem zérus ($\mathbf{J_x}$ invertálható) , akkor az $(\bar{\mathbf{x}}, \bar{\mathbf{p}})$ pont környezetében léteznek az alábbi folytonosan differenciálható függvények

$$\begin{aligned} x_1 &= x_1(p_1, p_2, \ldots, p_m) \\ x_2 &= x_2(p_1, p_2, \ldots, p_m) \\ &\vdots \\ x_n &= x_n(p_1, p_2, \ldots, p_m) \end{aligned}$$

és az $(\bar{\mathbf{x}}, \bar{\mathbf{p}})$ pont környezetében igaz, hogy

$$\begin{aligned} f_1(x_1(\mathbf{p}), x_2(\mathbf{p}), \ldots, x_n(\mathbf{p}), \mathbf{p}) &= 0 \\ f_2(x_1(\mathbf{p}), x_2(\mathbf{p}), \ldots, x_n(\mathbf{p}), \mathbf{p}) &= 0 \\ &\vdots \\ f_n(x_1(\mathbf{p}), x_2(\mathbf{p}), \ldots, x_n(\mathbf{p}), \mathbf{p}) &= 0 \end{aligned}$$

b) Az $(\bar{\mathbf{x}}, \bar{\mathbf{p}})$ pontban az $x_i(\mathbf{p})$ $(i = 1, 2, \ldots, n)$ függvények p_k $(k = 1, 2, \ldots, m)$ szerinti deriváltjaira igaz, hogy

$$\begin{bmatrix} \frac{\partial x_1}{\partial p_k} \\ \frac{\partial x_2}{\partial p_k} \\ \vdots \\ \frac{\partial x_n}{\partial p_k} \end{bmatrix} = - \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial f_1}{\partial p_k} \\ \frac{\partial f_2}{\partial p_k} \\ \vdots \\ \frac{\partial f_n}{\partial p_k} \end{bmatrix}$$

Amennyiben a $\frac{\partial x_i}{\partial p_k}$ $(i = 1, \ldots, n)$ mennyiségeket az $\mathbf{x}'_{p_k} \in R^n$ vektorba, a $\frac{\partial f_i}{\partial p_k}$ $(i = 1, \ldots, n)$ mennyiségeket az $\mathbf{f}'_{p_k} \in R^n$ vektorba, a $\frac{\partial f_i}{\partial x_j}(i = 1, \ldots, n,\ j = 1, \ldots, n)$ mennyiségeket a már ismertetett $\mathbf{J_x} \in R^{n\times n}$ Jacobi-mátrixba foglaljuk, akkor a fenti összefüggés az alábbi egyszerűbb formában írható fel

$$\mathbf{x}'_{p_k} = -\mathbf{J_x}^{-1} \cdot \mathbf{f}'_{p_k} \qquad (k = 1, 2, \ldots, m).$$

Megjegyzés:

A fenti formulát írhatjuk mátrixos alakban is. Az $\mathbf{x}'_{p_k}$ $(k = 1, 2, \ldots, m)$ vektorokat foglaljuk egy $\mathbf{X'_p}$ mátrixba úgy, hogy ezek a vektorok a mátrix oszlopvektorai legyenek, hasonlóan módon foglaljuk az $\mathbf{f}'_{p_k}$ $(k = 1, 2, \ldots, m)$ vektorokat egy $\mathbf{F'_p}$ mátrixba. Az $\mathbf{X'_p}$ ill. $\mathbf{F'_p}$ mátrixok i-edik sorának k-edik eleme tehát $\frac{\partial x_i}{\partial p_k}$ ill. $\frac{\partial f_i}{\partial p_k}$. Ekkor a fenti formulából az alábbi formulát kapjuk

$$\mathbf{X'_p} = -\mathbf{J_x}^{-1} \cdot \mathbf{F'_p},$$

amely részletezve

$$\begin{bmatrix} \frac{\partial x_1}{\partial p_1} & \frac{\partial x_1}{\partial p_2} & \cdots & \frac{\partial x_1}{\partial p_m} \\ \frac{\partial x_2}{\partial p_1} & \frac{\partial x_2}{\partial p_2} & \cdots & \frac{\partial x_2}{\partial p_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial p_1} & \frac{\partial x_n}{\partial p_2} & \cdots & \frac{\partial x_n}{\partial p_m} \end{bmatrix} = - \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial f_1}{\partial p_1} & \frac{\partial f_1}{\partial p_2} & \cdots & \frac{\partial f_1}{\partial p_m} \\ \frac{\partial f_2}{\partial p_1} & \frac{\partial f_2}{\partial p_2} & \cdots & \frac{\partial f_2}{\partial p_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial p_1} & \frac{\partial f_n}{\partial p_2} & \cdots & \frac{\partial f_n}{\partial p_m} \end{bmatrix}.$$

Példa:

Legyen $n = 2$, $m = 2$. Legyen

$$\begin{aligned} f_1(x_1, x_2, p_1, p_2) &= x_1 - p_1^2 x_1 - p_2 - 5 = 0 \\ f_2(x_1, x_2, p_1, p_2) &= p_1 x_2 \sqrt{p_2} - 2 = 0 \end{aligned}$$

Legyen az adott pont az $x_1 = -2, x_2 = 1, p_1 = 2, p_2 = 1$. Először határozzuk meg a Jacobi-mátrixot

$$\mathbf{J_x} = \begin{bmatrix} 1 - p_1^2 & 0 \\ 0 & p_1\sqrt{p_2} \end{bmatrix}$$

A Jacobi-mátrix determinánsa az adott pontban -6, tehát nem zérus, így léteznek a pont környezetében az $x_1 = x_1(p_1, p_2)$ és $x_2 = x_2(p_1, p_2)$ függvények. A tétel b) része szerint, anélkül, hogy ezeket a függvényeket meghatároznánk, kiszámíthatók ezek p_1, p_2 szerinti parciális deriváltjai. A további szükséges számított mennyiségek:

$$\begin{aligned} \mathbf{J_x}^{-1} &= \begin{bmatrix} \frac{1}{1-p_1^2} & 0 \\ 0 & \frac{1}{p_1\sqrt{p_2}} \end{bmatrix} \\ \mathbf{f}'_{p_1} &= \begin{bmatrix} \frac{\partial f_1}{\partial p_1} \\ \frac{\partial f_2}{\partial p_1} \end{bmatrix} = \begin{bmatrix} -2p_1x_1 \\ x_2\sqrt{p_2} \end{bmatrix} \\ \mathbf{f}'_{p_2} &= \begin{bmatrix} \frac{\partial f_1}{\partial p_2} \\ \frac{\partial f_2}{\partial p_2} \end{bmatrix} = \begin{bmatrix} -1 \\ \frac{p_1x_2}{2\sqrt{p_2}} \end{bmatrix} \end{aligned}$$

Az adott pontban az értékek:

$$\begin{aligned} \mathbf{J_x}^{-1} &= \begin{bmatrix} -\frac{1}{3} & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \\ \mathbf{f}'_{p_1} &= \begin{bmatrix} 8 \\ 1 \end{bmatrix} \\ \mathbf{f}'_{p_2} &= \begin{bmatrix} -1 \\ 1 \end{bmatrix} \end{aligned}$$

Az eredmények értékelése:

Ha a p_1 értéket 2-ről elegendően kicsi λ-val megváltoztatjuk, akkor az f_1 függvény értékének megváltozása megközelítőleg $8 \cdot \lambda$, az f_2 függvény értékének megváltozása pedig megközelítőleg $1 \cdot \lambda$.

Ha a p_2 értéket 1-ről elegendően kicsi λ-val megváltoztatjuk, akkor az f_1 függvény értékének megváltozása megközelítőleg $(-1) \cdot \lambda$, az f_2 függvény értékének megváltozása pedig megközelítőleg $1 \cdot \lambda$.

A létező $x_1 = x_1(p_1, p_2)$, $x_2 = x_2(p_1, p_2)$ függvények p_1 ill. p_2 szerinti parciális deriváltjai az adott pontban

$$\begin{aligned}
\mathbf{x}'_{p_1} &= \begin{bmatrix} \frac{\partial x_1}{\partial p_1} \\ \frac{\partial x_2}{\partial p_1} \end{bmatrix} = -\mathbf{J}_x^{-1}\mathbf{f}'_{p_1} = \begin{bmatrix} \frac{8}{3} \\ -\frac{1}{2} \end{bmatrix} \\
\mathbf{x}'_{p_2} &= \begin{bmatrix} \frac{\partial x_1}{\partial p_2} \\ \frac{\partial x_2}{\partial p_2} \end{bmatrix} = -\mathbf{J}_x^{-1}\mathbf{f}'_{p_2} = \begin{bmatrix} -\frac{1}{3} \\ -\frac{1}{2} \end{bmatrix}
\end{aligned}$$

Az eredmények értékelése:

Az $\mathbf{x}'_{p_1}$ vektor elemeinek jelentése:

Ha a p_1 értéket 2-ről elegendően kicsi λ-val megváltoztatjuk, akkor az $x_1(p_1, p_2)$ függvény értékének megváltozása megközelítőleg $\frac{8}{3} \cdot \lambda$, az $x_2(p_1, p_2)$ függvény értékének megváltozása pedig megközelítőleg $-\frac{1}{2} \cdot \lambda$.

Ha a p_2 értéket 1-ről elegendően kicsi λ-val megváltoztatjuk, akkor az $x_1(p_1, p_2)$ függvény értékének megváltozása megközelítőleg $-\frac{1}{3} \cdot \lambda$, az $x_2(p_1, p_2)$ függvény értékének megváltozása pedig megközelítőleg $-\frac{1}{2} \cdot \lambda$.

Megjegyzés:

Általánosan is felírhatjuk a formulákat, mégpedig minden olyan esetben, amelyben a Jacobi-mátrix determinánsa nem zérus. Könnyen meggyőződhetünk róla, hogy $p_1 \neq 0, p_1 \neq 1, p_1 \neq -1, p_2 > 0$ esetekben a Jacobi-mátrix invertálható, az inverz

$$\mathbf{J}_\mathbf{x}^{-1} = \begin{bmatrix} \frac{1}{1-p_1^2} & 0 \\ 0 & \frac{1}{p_1\sqrt{p_2}} \end{bmatrix},$$

ekkor a kérdéses deriváltak általánosan

$$\begin{aligned}
\mathbf{x}'_{p_1} &= -\mathbf{J}_x^{-1}\mathbf{f}'_{p_1} = \begin{bmatrix} \frac{2p_1x_1}{1-p_1^2} \\ -\frac{x_2}{p_1} \end{bmatrix} \\
\mathbf{x}'_{p_2} &= -\mathbf{J}_x^{-1}\mathbf{f}'_{p_2} = \begin{bmatrix} \frac{1}{1-p_1^2} \\ -\frac{x_2}{2p_2} \end{bmatrix}
\end{aligned}$$

Tapasztalhattuk a példa kapcsán, hogy az $x_1 = x_1(p_1, p_2)$, $x_2 = x_2(p_1, p_2)$ függvények konkrét ismerete nélkül is meg tudtuk határozni a deriváltjaikat. Gyakorlásképpen meghatározhatjuk ezeket a függvényeket, az egyenletrendszert kell megoldani x_1, x_2 változókra, amelyek a következők:

$$\begin{aligned}
x_1 &= x_1(p_1, p_2) = \frac{p_2 + 5}{1 - p_1^2} \\
x_2 &= x_2(p_1, p_2) = \frac{2}{p_1\sqrt{p_2}}
\end{aligned}$$

Feladat:

Végezzen el egy ellenőrzést, a fenti függvényeket deriválja p_1, p_2 szerint! Ekkor a deriváltakra a már ismert eredményeket fogja megkapni.