Nemlineáris optimalizálás

Rapcsák Tamás

2007.

Előszó

A *Nemlineáris optimalizálás* című anyag a gazdaságmatematikai elemző közgazdász hallgatók számára készült és egyrészt a matematikai alapozó kurzusokra (Dancs és Puskás, *Vektorterek*, 2001; Dancs, Magyarkúti, Medvegyev és Tallos, *Bevezetés a matematikai analízisbe*, 2003) épít, másrészt Stahl, *Optimumszámítás* című jegyzetére. A hallgatók a nemlineáris optimalizálás alapjaival az *Optimumszámítás* című tantárgy keretében ismerkednek meg, majd bővebb tárgyalásra az *Operációkutatás* szakirány *Nemlineáris optimalizálás* című tantárgyában kerül sor.

A jegyzet szakít azzal az általános gyakorlattal, hogy az anyag tárgyalása módszertani szempontok alapján történik, hanem inkább a gyakorlati alkalmazások lehetőségét szem előtt tartva, a modellezésre helyezi a fő hangsúlyt. Ebből következően az analízist és az algebra eszköztárát magasabb szinten használjuk.

A nemlineáris optimalizálás kifejlődéséhez a hazai hozzájárulás kiemelkedő. Itt elsősorban Farkas (mechanikai egyensúly, *Farkas tétel*) és Egerváry (mátrix elmélet, rangszámcsökkentés) eredményeire gondolunk. Az újabb munkák közül Forgó (1988), Martos (1975), Mayer (1998), Pintér (1996), Prékopa (1995), Rapcsák (1997) és Roos, Terlaky és Vial (1997) monográfiáit említjük.

Köszönetet szeretnék mondani Fiala Tibornak, Forgó Ferencnek és Komlósi Sándornak az anyag gondos átolvasásáért, a hasznos észrevételekért és tanácsokért, a TeX-file készítéséért pedig Móczár Károlynak.

Külön köszönettel tartozom Fülöp Jánosnak, aki vállalta a jegyzet lektorálását.

BEVEZETÉS

A nemlineáris optimalizálás mind elméleti érdekességénél fogva, mind a gyakorlati alkalmazásokat tekintve az optimalizáláselmélet rendkívül gyorsan fejlődő ága. A nemlineáris optimalizálási kutatások alig több mint ötven éves múltra tekintenek vissza, jóllehet már jóval korábban több matematikai és fizikai probléma vezetett ilyen jellegű feladatra. Azonban a nemlineáris optimalizálási feladatoknak az igazi jelentőségét a széles körű gyakorlati alkalmazhatóságuk és az alkalmazások fontossága adta meg. Mindezt a számítógépek elterjedése tette lehetővé, ami lényeges szemléleti változással is járt. Míg korábban csak a feladatok (elméleti) megoldása volt a cél, addig napjainkban a megoldó algoritmusok és a szoftverek előnyös tulajdonságainak a megléte is nagyon lényeges szempont (pl. minél kisebb számítási időigény és memória kapacitás, a mérethatárok növelése, könnyen kezelhető és változtatható programok). Ez a magyarázata annak, hogy a nemlineáris optimalizáláson belül a kutatások három irányban ágaztak el: a feladatok és a megoldó algoritmusok matematikai vizsgálata, a megoldó algoritmusok számítógépes implementálása és az experimentálás, valamint a nemlineáris optimalizálás gyakorlati alkalmazása irányában. Mivel a nemlineáris optimalizálás ilyen méretű fejlődését a gyakorlati alkalmazások és az egyre nagyobb teljesítményű számítógépek segítették elő, ezért érthető, hogy elsősorban az algoritmusokkal való számítógépes kísérletek és az alkalmazások területén nagy az előrelépés. Azonban elméleti vonatkozásban is komoly eredmények születtek, és a nemlineáris optimalizálási feladatok matematikai tulajdonságainak mélyrehatóbb elemzése során felhasználásra vagy továbbfejlesztésre kerültek a klasszikus matematikai diszciplínák eredményei is (pl. geometria, funkcionál és numerikus analízis, differenciálegyenletek, mértékelmélet, statisztika, valószínűségelmélet).

Néhány matematikai és fizikai példa nemlineáris optimalizálási feladatra. Az elméleti matematikán belül az 1637-től 1996-ig megoldatlan, híres Fermat sejtés és van der Waerden 1926-ban permanensekre megfogalmazott és 1981-ben megoldott sejtése is nemlineáris optimalizálási problémára vezet (lásd 2. fejezet). Statisztikán belül a regressziószámítás nemlineáris optimalizálási feladat megoldását jelenti (lásd pl., Hunyadi és Vita, 2002). Lagrange 1788-ban közölte a függvények egyenlőség feltételek melletti szélsőértékeinek meghatározására vonatkozó multiplikátoros módszerét, a *Mécanique Analytique* című könyve első kötetében (77-79. oldal). Farkas a mechanikai egyensúly szükséges feltételeinek levezetésére dolgozta ki a homogén, lineáris egyenlőtlenségekre vonatkozó híres tételét, ami a nemlineáris optimalizálási szakirodalomban egyike a leggyakrabban idézett dolgozatoknak (lásd 1. és 8. fejezet).

A nemlineáris optimalizálás gyakorlati alkalmazásai közül először néhány hazai, mérnöki tervezési példáról lesz szó. A rúdszerkezetek méretezésekor adott külső terhelés esetén több, a funkcionális követelményeknek jól megfelelő szerkezet közül választhatunk. Ezért valamilyen gazdaságossági szempont alapján érdemes kiválasztani a legmegfelelőbbet. Az IKARUS buszok oldalfalainak méretezése során ez a szerkezet súlya volt [24, 25]. Az IKARUS gyár megrendelésére készült el a mechanikus sebességváltóval rendelkező autóbuszok erőátviteli láncának optimális méretezése. Ezt a feladatot négyfokozatú váltó esetén, *12* változót és *58* feltételt tartalmazó, míg hatfokozatú váltó esetén, *16* változót és *82* feltételt tartalmazó nemlineáris optimalizálási probléma megoldására vezettük vissza [52, 55]. A gyakorlati alkalmazások során kiemelt jelentősége van a lineáris optimalizálásra visszavezethető nemlineáris modelleknek. Erre példa egy új létesítmény megvalósítása során a tereprendezési feladat megoldása, ami időigényes, sok fáradságot igénylő feladat, mivel nagy volumenű földmennyiség megmozgatását teszi szükségessé [53, 54]. Prékopa vezette be az együttes valószínűségekre korlátot adó sztochasztikus optimalizálási feladatokat, amelyek nemlineáris optimalizálási feladatok megoldására vezetnek. Ezek részletes kifejtése megtalálható a könyvében [49], illetve a [10] munkában. Együttműködő víztározók sztochasztikus programozással történő méretezését ismertetik a [50, 51] cikkek.

A közgazdaságtanban a matematikai közgazdaságtan és a mikroökonómia az általános közgazdász képzés standard tananyagává vált. A matematikai közgazdaságtant – amit mint önálló tudományterületet 1930 óta ismerünk – a matematikai formanyelv és eszközök segítségével kifejtett közgazdasági elméletek és modellek összességeként lehet röviden meghatározni. Szoros rokonságban áll az ökonometriával, az operációkutatással és azon belül a nemlineáris optimalizálással. Erre példa a mikroökonómia, ahol az alapvető eszköztár ma is a termelési és hasznossági függvények, illetve optimumra törekvő gazdasági döntéshozók feltételezése alapján, nemlineáris optimalizálási modellek felhasználásával levezetett keresleti és kínálati függvények, valamint egyensúlyi árak. A matematikai közgazdaságtan részletesebb tárgyalását tartalmazza Zalai (2000) könyve.

1. fejezet

A NEMLINEÁRIS OPTIMALIZÁLÁS KIALAKULÁSA

A nemlineáris optimalizálás elnevezés az 1950-ben publikált Kuhn-Tucker cikkből származik, amelyben a szerzők az optimalitás szükséges feltételeit vezették le. Jóllehet Karush ugyanezeket az összefüggéseket már 1939-ben megkapta, és - mint Prékopa rámutat az optimalizáláselmélet kialakulásáról szóló cikkeiben [47, 48] - Lagrange, Bernoulli, Fourier, Cournot, Gauss, Osztrogradszkij eredményeinek felhasználásával lényegében ugyanezt az állítást bizonyította Farkas is a mechanikai egyensúly problémáját vizsgálva, mégis a nemlineáris optimalizálás gyors fejlődése csak a Kuhn-Tucker cikk megjelenése után indult meg. Ugyanis, kialakulására és jelentőségének felismerésére döntő hatással volt az elektronikus számítógépek megjelenése (az első példányt a második világháború idején fejlesztették ki az Egyesült Államokban, és 1946 II. 15-én állították üzembe), továbbá a lineáris optimalizálás és a szimplex módszer megalkotása (Kantorovics 1939, Dantzig 1947). (A lineáris optimalizálással hasonló volt a helyzet, mint a nemlineáris optimalizálással, mivel Kantorovics orosz matematikus már 1939-ben tárgyalta a feladatot, de akkor még nem ismerték fel a téma fontosságát.) Mindkét felfedezés döntő, szemléleti változást hozott nemcsak a matematikában, hanem más tudományokban és számos gyakorlati területen is, mert segítségükkel lehetővé vált nagyméretű és bonyolult problémák elfogadható időn belül történő megoldása. Ennek hatására az operációkutatáson és az alkalmazott matematikán belül újabb és újabb ágak születtek (pl. nemlineáris (ezen belül kvadratikus és geometriai), diszkrét és sztochasztikus optimalizálás, irányításelmélet), amelyek már - jóllehet sok közös elem is volt bennük - minőségileg is különböztek a klasszikus matematikai diszciplínáktól.

Az operációkutatásban és az alkalmazott matematikában ugyanis az elméleti vizsgálatokon túlmenően a cél mindig a megoldás kiszámítása, képletek helyett zömében algoritmusok alkalmazásával, ahol sok egyéb szempontot is figyelembe kell venni (pl. milyen információtechnológia áll rendelkezésre, mely adatok ismertek, milyen típusú a modell, milyen körülmények között kerül sor az alkalmazásra). Látható tehát, hogy itt inkább az algoritmusok és nem a tételek dominálnak, továbbá a deduktív módszer keveredik induktív elemekkel (pl. egy megoldási módszer hatékonyságát elsősorban a tapasztalatra támaszkodva ítéljük meg).

A nemlineáris optimalizálás történetében az első komoly eredményt Lagrange érte el, aki 1788-ban publikálta a függvények egyenlőség feltételek melletti szélsőértékeinek meghatározására vonatkozó multiplikátoros módszerét, a *Mécanique Analytique* című könyve első kötetében (77-79. oldal). A módszer érvényességét algebrai úton bizonyította.

Ezután Farkas munkásságát kell kiemelni, akinek a *Crelle Journal*ban 1901-ben publikált híres dolgozata egyike lett a leggyakrabban említett dolgozatoknak a matematikai és a nemlineáris optimalizálási szakirodalomban. Ezt a dolgozatát elsősorban a homogén, lineáris egyenlőtlenségekre vonatkozó tétele miatt idézik, amelyre *Farkas-tétel* néven hivatkoznak, s amelyet a nemlineáris optimalizálásban az optimalitás szükséges feltételeinek a levezetésére használnak. Azonban Farkas jól meghatározott cél érdekében fejlesztette ki a lineáris egyenlőtlenségek elméletét. Az elméleti fizika professzora volt a Kolozsvári Egyetemen és az eredményeit a mechanikai egyensúly problémájára, a Fourier-féle elvre vonatkozóan alkalmazta. Mivel a legismertebb cikkében erről nem tesz említést, „*Emiatt munkásságának ez a vonatkozása nem vált nemzetközileg ismertté. Ennek oka az is, hogy az analitikus mechanikában nyert eredmény optimalizáláselméleti interpretálása akkor nem történt meg, márpedig úgy tűnik, hogy ilyen irányú jelentősége fontosabb, mint a* *mechanikai*” [47]. Ezt az interpretációt Prékopa [47, 48] elvégzi a dolgozataiban és megmutatja, hogy a Fourier-féle mechanikai elv duális alakja, amit Cournot írt fel és Farkas bizonyított be először, lényegében azonos az optimalitás nemlineáris optimalizálásbeli szükséges feltételeivel. Rámutat, hogy a nemlineáris optimalizálás kialakulásának történetében feltétlenül meg kell említeni Fourier 1798-ban írt dolgozatát, amelyben a róla elnevezett egyenlőtlenségi elvet mondja ki. Később Gauss és Osztrogradszkij újból kimondta az egyenlőtlenségi elvet. Ennek alapján Cournot és később Osztrogradszkij felírta a szükséges feltételeket sejtés formájában, Farkas pedig bizonyította e feltételek érvényességét, miközben a bizonyítás első felét illetően Fourier munkájára hagyatkozott, amelyből hiányzott a regularitási feltétel (*constraint qualification*). A regularitási feltétel mind az optimalizáláselmélet, mind pedig a mechanika számára alapvető feltétel. Egyenlőség feltételekkel korlátozott feladatok esetén Lagrange (1788) óta ismert ilyen feltétel, egyenlőtlenségi feltételek esetén viszont először Hamel 1927-ben megjelent dolgozatában található, amelyben a klasszikus mechanika axiomatikus felépítését kísérli meg.

Az egyenlőség feltételekkel megadott feladatokat vizsgálta Carathéodory 1935-ben, majd részletesebben Bliss 1938-ban, aki ebben az időben a Chicagói Egyetemen működő variációszámítási iskola vezetője volt. Ott dolgozott, többek között, Valentine, aki az egyenlőtlenség feltételekkel korlátozott variációszámítási problémával foglalkozott. Valószínűleg ennek hatására vetődött fel az egyenlőtlenség feltételekkel korlátozott nemlineáris optimalizálási feladat mint a variációszámítási probléma véges dimenziós változata. Graves ajánlotta a témát, akinek a vezetése alatt Karush (1939) ebből írta a „*master's thesis*”-t. A szerző az eredményeket nem publikálta, ezért azok sokáig ismeretlenek maradtak.

Karush munkájának elkészülte után, de Kuhnt és Tuckert megelőzve, John is vizsgálta az egyenlőtlenségi feltételekkel adott nemlineáris optimalizálási problémát. Ő nem használt regularitási feltételt, kivéve azt, hogy minden függvény folytonosan differenciálható. Az eredménye viszont gyengébb, mint Karushé. Ebben az időben John a konvex halmazokkal és a velük kapcsolatos geometriai jellegű egyenlőtlenségekkel foglalkozott. Az általa kidolgozott tételre a Sylvester probléma[1] egyik általánosításának megoldásához volt szüksége.

A nemlineáris programozás elnevezés Kuhn és Tucker 1950-ben megjelent cikkében szerepelt először, amelyben az egyenlőtlenség feltételekkel korlátozott feladat optimalitásának szükséges feltételeit vezették le. Eredményükhöz a lineáris programozás dualitás tételének általánosításával jutottak el. E cikk megjelenése után indult meg a nemlineáris optimalizálás rohamos fejlődése. Érdekes megemlíteni, hogy jóllehet a háttér különböző volt, Karush, illetve Kuhn és Tucker ugyanazt a tételt bizonyították be és ugyanazt a regularitási feltételt használták.

Az előzőekben láttuk, hogy a nemlineáris optimalizálás alapvető fontosságú eredményeihez, az optimalitási feltételekhez a legkülönbözőbb területeken dolgozó matematikusok és fizikusok, sokszor egymástól függetlenül jutottak el. A megfelelő problémák a mechanikai egyensúllyal, variációszámítással, geometriai egyenlőtlenségekkel, játékelmélettel, hálózatelmélettel, dualitás elmélettel és a lineáris programozással voltak kapcsolatosak.

Az optimalitási feltételek ismeretében sok szerző foglalkozott a különböző regularitási feltételekkel és a közöttük levő kapcsolatokkal. Az elért eredmények jól áttekinthető összefoglalása található Bazaraa és Shetty (1976, 1979) könyveiben. Az optimalitással kapcsolatban, a függvények általánosított konvexitási tulajdonságairól is érdekes eredmények születtek. Ezekről részletesebben lehet olvasni Mangasarian (1969), Martos (1975) és Avriel et al. (1988) könyveiben. Az utóbbi időben a nemdifferenciálható függvényekkel képzett nemlineáris optimalizálási feladatok optimalitási kérdéseinek van nagy irodalma. A nemlineáris optimalizálás történetéről részletesebb ismertetés található Rapcsák (1997) könyvében.

[1] lásd pl. Handbook of convex geometry, eds.: P.M. Gruber and J.M. Wills, North-Holland, 1993.

2. fejezet

NEMLINEÁRIS OPTIMALIZÁLÁSI FELADAT

Az optimalizálási problémákat a következőképpen lehet megfogalmazni:

legyen az f skalár értékű függvény tetszőleges A halmazon értelmezve és keressük az A halmaznak azt az $\mathbf{x}^*$ pontját, amelyre

$$f(\mathbf{x}^*) = \min\{f(\mathbf{x}) \mid \mathbf{x} \in A\}, \tag{2.1}$$

ha a minimum létezik. Ha a minimum nem létezik, de az infimum igen, akkor a probléma olyan A-beli $\hat{\mathbf{x}}$ pontot vagy pontokat találni, amely(ek)re az $f(\hat{\mathbf{x}})$ érték „közel" van az infimum értékhez. Ha se minimum, se infimum nem létezik, vagy nem tudjuk, hogy léteznek-e vagy sem, akkor olyan A halmazhoz tartozó pont, vagy más szóval megengedett megoldás megkeresése a cél, ahol a célfüggvény érték jobb, mint az induló pontban. Maximalizálási problémákat hasonlóan lehet megfogalmazni. A (2.1) probléma neve többszempontú optimalizálási probléma, ha f vektorértékű függvény.

A nemlineáris optimalizálási, vagy nemlineáris programozási problémák (rövidítve NLO vagy NLP) definiciója nem egyértelmű az optimalizáláselmélet irodalmában. A „nemlineáris" jelző is félrevezető, mivel minden optimalizálási feladatot magában foglal, amiben nemlineáris függvények szerepelnek. Az NLO gyakorlati alkalmazásait alapul véve, akkor nevezünk egy (2.1) optimalizálási problémát NLO-nak, ha a következő három tulajdonság teljesül:

1. $A \subseteq R^n$ vagy $A \subseteq H$, ahol R^n jelöli az n-dimenziós Euklideszi teret és H egy Hilbert teret;
2. az A halmazt véges vagy végtelen számú egyenlőség és/vagy egyenlőtlenség határozza meg, és
3. az A halmaz összefüggő[2].

A klasszikus NLO a következő formában adható meg:

$$\min \ f(\mathbf{x})$$

$$g_i(\mathbf{x}) - b_i = y_i, \qquad i = 1, \ldots, m, \tag{2.2}$$

$$\mathbf{x} \in R^n, \qquad \mathbf{y} \in R^m,$$

ahol az f, g_i, $i = 1, \ldots, m$, függvények az R^n-ben vagy az R^n egy részhalmazán vannak értelmezve, a b_i, $i = 1, \ldots, m$, értékek állandók, az $\mathbf{x}$ n-dimenziós és az $\mathbf{y}$ m-dimenziós változók, amelyek közül bármely változó csoportra nemnegativitási feltételek lehetnek érvényesek. Ha az f célfüggvény helyett a $-f$ célfüggvényt tekintjük a (2.2) feladatban, akkor minimalizálás helyett maximalizálás a feladat. Ezért a minimalizálási és maximalizálási feladat ekvivalens. Az alábbi példák mutatják, hogy a (2.2) NLO sok ismert optimalizálási problémát tartalmaz.

Ha a (2.2) problémában szereplő célfüggvény és a feltételi függvények lineárisak, az $\mathbf{x}$ és $\mathbf{y}$ vektor változók nemnegatívak, akkor a (2.2) probléma a következő formára hozható:

$$\min \ \mathbf{c}^T\mathbf{x}$$

$$\mathbf{A}\mathbf{x} \geq \mathbf{b}, \qquad \mathbf{x} \geq 0, \tag{2.3}$$

$$\mathbf{x} \in R^n,$$

[2]Egy halmaz összefüggő, ha nem adható meg két, nem üres, nyílt és diszjunkt halmaz uniójaként.

ahol $\mathbf{c}$ R^n-beli vektor és $\mathbf{A}$ $m \times n$-es mátrix.

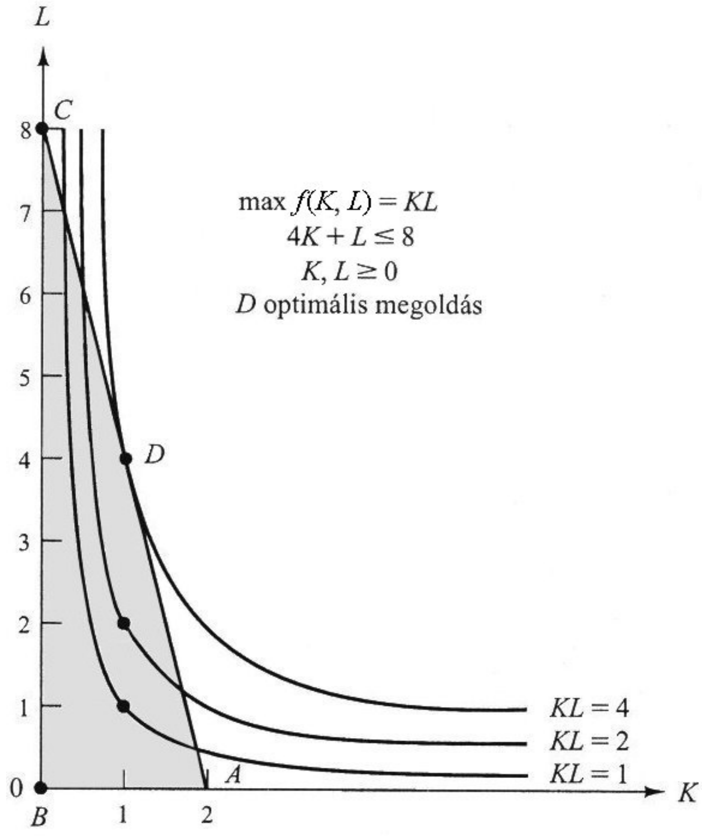
Ha a (2.2) problémában az előbbi feltételek mellett $\mathbf{y} = \mathbf{0}$, akkor a feladat a következő alakkal ekvivalens:

$$\min\ \mathbf{c}^T\mathbf{x}$$

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \qquad \mathbf{x} \geq \mathbf{0}, \tag{2.4}$$

$$\mathbf{x} \in R^n.$$

Ebből látható, hogy az NLO a lineáris optimalizálási probléma (LO) általánosítása. Az 1. ábra olyan NLO-t mutat, aminek az optimális megoldása nem extremális pont.



1. ábra

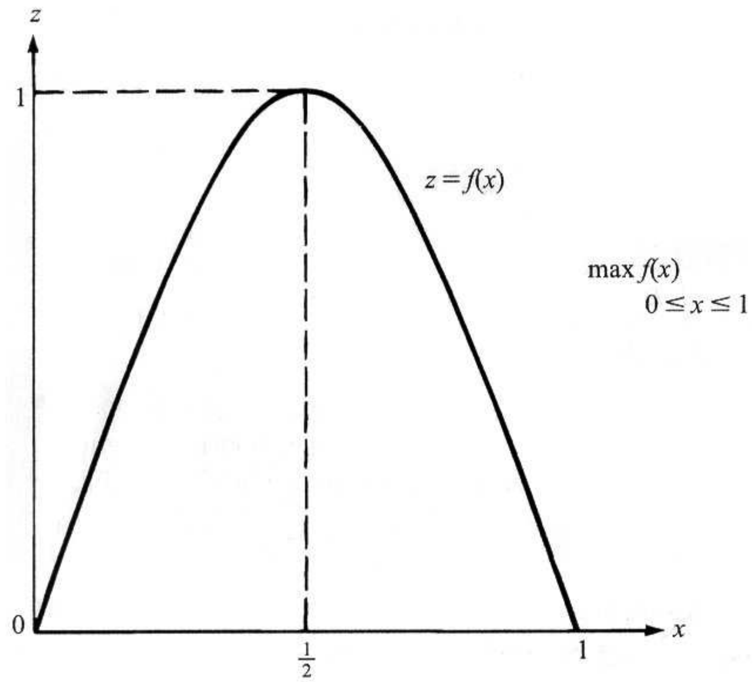
Egy NLO, aminek az optimális megoldása nem extremális pont

Ha a (2.2) problémában $\mathbf{y} = \mathbf{0}$ és $\mathbf{b} = \mathbf{0}$, akkor az először Lagrange által vizsgált, egyenlőség feltételekkel korlátozott NLO-t kapjuk.

Ha $A \subseteq R^n$ tetszőleges halmaz és g_1 a halmaz *karakterisztikus függvénye* ($g_1(\mathbf{x}) = 1$, $\mathbf{x} \in A$; $g_1(\mathbf{x}) = 0$, $\mathbf{x} \notin A$), továbbá $m = 1$, $b_1 = 1$ és $y_1 = 0$, akkor (2.2) a következő problémává alakul:

$$\begin{aligned} &\min\ f(\mathbf{x}) \\ &\mathbf{x} \in A \subseteq R^n. \end{aligned} \tag{2.5}$$

Nem biztos, hogy az NLO optimális megoldása a megengedett tartomány határán található, lásd 2. ábra.



2. ábra

Egy NLO, aminek az optimális megoldása nem a megengedett tartomány határán található

Ha a (2.2) problémában m feltétel helyett $p + m$-et tekintünk, és a $(p + m)$-dimenziós $\mathbf{y}$ vektor utolsó m komponense nemnegatív, az első p pedig nulla, a $(p+m)$-

dimenziós **b** vektor a nulla vektor, akkor a klasszikus NLO-t kapjuk:

$$
\begin{aligned}
&\min\ f(\mathbf{x}) \\
&h_j(\mathbf{x}) = 0, \qquad j = 1, \dots, p, \\
&g_i(\mathbf{x}) \geq 0, \qquad i = 1, \dots, m, \\
&\mathbf{x} \in R^n.
\end{aligned}
\tag{2.6}
$$

Vezessük be a következő jelölést:

$$
M[\mathbf{h}, \mathbf{g}] = \{\mathbf{x} \in R^n |\ h_j(\mathbf{x}) = 0, j = 1, \dots, p, \quad g_i(\mathbf{x}) \geq 0, i = 1, \dots, m\}. \tag{2.7}
$$

Egy $\mathbf{x}_0$ pont a (2.6) NLO (szigorú) lokális minimuma, ha van olyan $U(\mathbf{x}_0, \delta)$ környezet, hogy $\mathbf{x}_0 \in M[\mathbf{h}, \mathbf{g}]$ és

$$
f(\mathbf{x}) \geq (>) f(\mathbf{x}_0) \quad \text{minden} \qquad \mathbf{x} \in U(\mathbf{x}_0, \delta) \cap M[\mathbf{h}, \mathbf{g}] \quad \text{esetén}, \tag{2.8}
$$

ahol

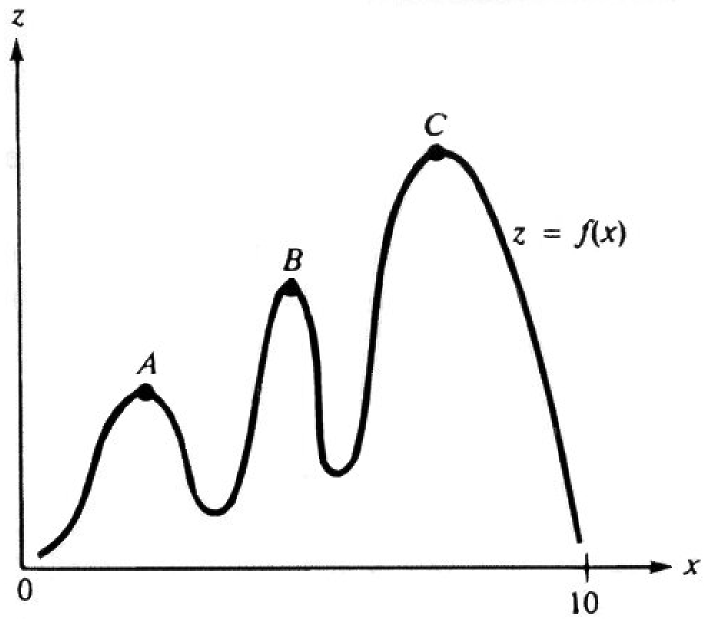
$$
\begin{gathered}
U(\mathbf{x}_0, \delta) = \{\mathbf{x} \in R^n | \quad \|\mathbf{x} - \mathbf{x}_0\| \leq \delta\}, \\
\text{(az } \mathbf{x}_0 \text{ pont } \delta \text{ sugarú környezete).}
\end{gathered}
\tag{2.9}
$$

Egy lokális optimum nem feltétlenül az NLO optimális megoldása, lásd 3. ábra.

Az $M[\mathbf{h}, \mathbf{g}]$ a (2.6) probléma megengedett pontjainak halmaza. Ha $\mathbf{x}_0 \in M[\mathbf{h}, \mathbf{g}]$ és

$$
f(\mathbf{x}) \geq f(\mathbf{x}_0) \qquad \text{minden} \qquad \mathbf{x} \in M[\mathbf{h}, \mathbf{g}] \qquad \text{esetén}, \tag{2.10}
$$

akkor az $\mathbf{x}_0$ pont a (2.6) probléma globális minimum pontja. Ha a (2.10) egyenlőtlenségben a „≥ 0” helyett „$\leq$” szerepel, akkor az $\mathbf{x}_0$ pont a (2.6) globális maximum pontja.



3. ábra

Egy lokális maximum nem feltétlenül az NLO optimális megoldása

Speciális eseteket kivéve nagyon nehéz feladat megtalálni egy NLO globális minimumát vagy azt ellenőrizni, hogy adott megengedett megoldás globális minimum-e. Lássunk rá egy híres példát!

Tekintsük az 1637-től 1996-ig megoldatlan, híres Fermat sejtést, ami szerint az $x^n + y^n - z^n = 0$ egyenletnek nincs egész számokból álló megoldása az $x \geq 1$, $y \geq 1$, $z \geq 1$, $n \geq 3$, egyenlőtlenségekkel megadott tartományban.

Tekintsük a következő NLO-t:

$$\min\ (x^n + y^n - z^n)^2 + r\Bigg[\Big(-1+\cos(2\pi x)\Big)^2 + \Big(-1+\cos(2\pi y)\Big)^2 + \Big(-1+\cos(2\pi z)\Big)^2 + \Big(-1+\cos(2\pi n)\Big)^2\Bigg] \tag{2.11}$$

$$x \geq 1, \quad y \geq 1, \quad z \geq 1, \quad n \geq 3,$$

ahol r pozitív paraméter, π irracionális szám és egyenlő az R^2-beli egységsugarú kör kerületének a felével, $\cos\alpha$ pedig a radiánban mért α szög koszinuszát megadó függvény.

Látható, hogy (2.11) lineáris egyenlőtlenségekkel korlátozott 4 változós NLO. Bizonyítható, hogy Fermat sejtése akkor és csak akkor nem teljesül, ha (2.11) optimum értéke 0 és ezt az optimum értéket valamely megengedett pontban felveszi a célfüggvény, mivel a (2.11) NLO bármely (x, y, z, n) globális optimum pontja ellenpéldát szolgáltatna Fermat sejtésére. Megjegyezzük, hogy (2.11) minden egész számokból álló megengedett megoldása a feladat egy lokális minimuma.

Egy NLO-ban a lokális minimumok száma nagyon nagy lehet. Példaként tekintsük a következő feladatot:

$$\min\ -\sum_{j=1}^{n}(x_j - 1/2)^2, \quad 0 \leq x_j \leq 1, \quad j = 1, \ldots, n. \tag{2.12}$$

Ebben a feladatban a megengedett halmaz mind a 2^n számú csúcspontja lokális minimum. Sajnos, jelenleg nincs más kidolgozott technika a lokális minimumok számának meghatározására, mint a teljes leszámlálás, azaz minden szóba jöhető pont esetén megvizsgálni, hogy az adott pont lokális minimum-e.

Egy másik híres matematikai feladat, ami egy NLO globális optimumának a meghatározására vezet, van der Waerden (1926) nevéhez fűződik.

Ha $\mathbf{C} = (c_{ij})$ egy n-ed rendű négyzetes mátrix, akkor a $\mathbf{C}$ mátrix *permanense*, ami

$$f(\mathbf{C}) = \sum_{(p_1,\ldots,p_n)} c_{1p_1} \ldots c_{np_n}$$

alakban adható meg, egyenlő az $1, \ldots, n$ számok $n!$ számú $(p_1, \ldots, p_n)$ permutációihoz tartozó mátrixelemek szorzatának összegével. Egy n-ed rendű, négyzetes mátrix *kétszeresen sztochasztikus*, ha az elemei nemnegatív számok és minden sor és oszlop összege 1-gyel egyenlő.

Az optimalizálási feladat olyan négyzetes $\mathbf{C} = (c_{ij})$ mátrix meghatározása, ami megoldása a következő NLO-nak:

$$
\begin{aligned}
&\min\ f(\mathbf{C})\\
&\sum_{j=1}^{n} c_{ij} = 1, \quad i = 1, \ldots, n,\\
&\sum_{i=1}^{n} c_{ij} = 1, \quad j = 1, \ldots, n,\\
&c_{ij} \geq 0, \quad c_{ij} \in R, \quad i, j = 1, \ldots n.
\end{aligned}
\tag{2.13}
$$

1926-ban van der Waerden azt sejtette, hogy ennek a feladatnak a globális minimuma az a kétszeresen sztochasztikus mátrix, amelynek elemeire az teljesül, hogy

$$c_{ij} = 1/n, \qquad i, j = 1, \ldots, n,$$

és a célfüggvény értéke ebben a pontban $n!/n^n$. Ez a sejtés hosszú ideig ellenállt a matematikusok rohamainak, míg végül 1981-ben Egorychev és Falikman is bebizonyította.

Nézzünk meg néhány egyszerű nemlineáris optimalizálási feladatot!

2.1. Példa. *Egy cégnek c forintba kerül egy egységnyi termék előállítása. Ha a cég egységenként x forintért kínálná értékesítésre a terméket, akkor $F(x)$ egységre lenne fogyasztói igény. Milyen árat kell a cégnek megállapítania profitja maximalizálásához?*

Megoldás: A cég döntési változója x. Mivel a cég profitja $(x - c)F(x)$, a cég a következő maximalizálási feladatot akarja megoldani: $\max f(x) = (x - c)F(x)$, $x > 0$.

2.2. Példa. *Ha egy cég K egységnyi tőkét és L egységnyi munkaerőt használ fel, akkor KL egységnyi terméket tud előállítani. A tőke egy egysége 4\$, a munkaerő egy egysége pedig 1\$ áron szerezhető be. Tőkére és munkaerőre összesen 8\$ áll rendelkezésre. Hogyan tudja a cég maximalizálni az előállítandó termék mennyiségét?*

Megoldás: Jelölje K és L, hogy hány egységnyi tőkét, illetve hány egységnyi munkaerőt használ fel a cég. Ekkor K és L nyilván eleget tesz a $4K + L \leq 8$, $K \geq 0$ és $L \geq 0$ feltételeknek. Tehát, a cég a következő feltételes maximalizálási feladatot akarja megoldani:

$$\max f(K, L) = KL$$

$$4K + L \leq 8,$$

$$K, L \geq 0.$$

2.3. Példa. *Ismeretesek a különféle termelési függvények, amik a valamiképpen mért hozamot az ugyancsak valamiképpen mért ráfordítások függvényében írják le. Legyenek ezen utóbbiak a K tőke és az L munkaerő, és tekintsük az*

$$f(K, L) = cK^{\alpha}L^{1-\alpha}, \quad K > 0, \quad L > 0, \quad c > 0, \quad 0 < \alpha < 1,$$

Cobb-Douglas termelési függvényt, ahol c és α adott állandó. Ha ilyen kifejezést szeretnénk valamilyen a tőkére és a munkaerőre vonatkozó feltételek mellett maximalizálni, vagy ilyen alakú kifejezések összegét akarjuk maximalizálni, akkor ez a probléma már nem modellezhető mint LO, hanem NLO kezelését igényli.

3. fejezet

OPTIMALITÁSI FELTÉTELEK

Ebben a részben a

$$
\begin{aligned}
&\min \; f(\mathbf{x}) \\
&h_j(\mathbf{x}) = 0, \qquad j = 1, \dots, p, \\
&g_i(\mathbf{x}) \geq 0, \qquad i = 1, \dots, m, \\
&\mathbf{x} \in R^n,
\end{aligned}
\tag{3.1}
$$

alakú NLO szükséges és elégséges, lokális optimalitási feltételei találhatók, amit a szakirodalomban gyakran a *Lagrange szorzók módszer*ének neveznek. Ezek a feltételek szolgáltatják az alapot az elméleti és módszertani vizsgálatokhoz, valamint a megállási kritériumokat a számítógépen elvégzendő kísérletekhez. A nemlineáris optimalizálásnak hatalmas irodalma van, és az optimalitási feltételek szinte mindegyikben megtalálhatók. Itt csak néhány ismert munkára hivatkozunk, pl., Fiacco és McCormick (1968), Mangasarian (1969), Luenberger (1973), Martos (1975), Bazaraa és Shetty (1976, 1979).

Ha a (3.1) NLO célfüggvénye és feltételi függvényei differenciálhatók, akkor a lokális optimalitás az elsőrendű feltételekkel jellemezhető. Vezessük be a következő jelölést:

$$L(\mathbf{x},\boldsymbol{\mu},\boldsymbol{\lambda}) = f(\mathbf{x}) + \sum_{j=1}^{p}\mu_j h_j(\mathbf{x}) - \sum_{i=1}^{m}\lambda_i g_i(\mathbf{x}),$$
$$\mathbf{x}\in R^n,\quad \boldsymbol{\mu}\in R^p,\quad \boldsymbol{\lambda}\in R^m,\quad \boldsymbol{\lambda}\geq \mathbf{0}. \tag{3.2}$$

Ezt a függvényt a (3.1) NLO *Lagrange függvény*ének nevezzük. A következő állítás kimondásához szükség van regularitási feltételre, ami a vizsgált pont egy környezetében jelent megszorítást a megengedett pontok halmazának analitikus leírására. Az optimalizáláselmélet számos regularitási feltételt ismer, pl., Fiacco és McCormick (1968), Mangasarian (1969), Bazaraa és Shetty (1976, 1979).

3.1. Definíció. *Tegyük fel, hogy a (3.1) NLO célfüggvénye és feltételi függvényei folytonosan differenciálhatók. A LICQ (linearly independent constraint qualification) regularitási feltétel teljesül az* $\mathbf{x}_0\in M[\mathbf{h},\mathbf{g}]$ *pontban, ha a*

$$\nabla h_j(\mathbf{x}_0),\quad j=1,\dots,p,$$

$$\nabla g_i(\mathbf{x}_0),\quad i\in I(\mathbf{x}_0)=\{i|\ g_i(\mathbf{x_0})=0,\quad i=1,\dots,m\}$$

vektorok lineárisan függetlenek.

Az $I(\mathbf{x}_0)$ indexhalmaz jelöli az *aktív egyenlőtlenség* feltételeket. A továbbiakban egy függvény gradiense mindig sorvektor.

3.1. Példa. *Tekintsük az* R^2 *2-dimenziós Euklideszi síkot és legyen* $h(x_1,x_2)=x_1$, $(x_1,x_2)\in R^2$. *A* $h(x_1,x_2)=0$, $(x_1,x_2)\in R^2$, *egyenlőség meghatározza a* $(0,x_2)$ *koordináta tengelyt, és ezen a tengelyen minden pontban teljesül a LICQ regularitási feltétel, mivel* $\nabla h(x_1,x_2)=(1,0)$.

Ha a $h(x_1,x_2)=x_1^2$, $(x_1,x_2)\in R^2$, *függvényt tekintjük, akkor a* $h(x_1,x_2)=0$, $(x_1,x_2)\in R^2$, *egyenlőség ugyanazt a koordináta tengelyt határozza meg, de a koordináta tengely egyetlen pontjában sem teljesül a LICQ regularitási feltétel, mivel* $\nabla h(\mathbf{x})=(2x_1,0)$.

Tekintsük a (3.1) optimalizálási feladatot és az $M[\mathbf{h},\mathbf{g}]$ megengedett pontok halmazát. Egy $\mathbf{x}(t):[a,b]\to M[\mathbf{h},\mathbf{g}]$, $a,b\in R$, folytonos leképzést az $M[\mathbf{h},\mathbf{g}]$ megengedett tartományban haladó *görbé*nek nevezünk. A *görbe* (*folytonosan*) *differenciál-*

ható, ha az $\mathbf{x}(t)$, $t \in [a,b]$, vektor értékű függvény minden komponense (folytonosan) differenciálható, és *kétszer* (*folytonosan*) *differenciálható*, ha minden komponense kétszer (folytonosan) differenciálható. Az első és a második differenciálhányadosokat (deriváltakat) az $\mathbf{x}'(t) = \dfrac{d\mathbf{x}(\mathbf{t})}{dt}$, $t \in [a,b]$, és az $\mathbf{x}''(t) = \dfrac{d^2\mathbf{x}(\mathbf{t})}{dt^2}$, $t \in [a,b]$, szimbólumok jelölik. Azt mondjuk, hogy az $\mathbf{x}(t)$, $t \in [a,b]$, *görbe átmegy az* $\mathbf{x}_0 \in M[\mathbf{h},\mathbf{g}]$ *ponton*, ha valamely $t_0 \in [a,b]$ értékre $\mathbf{x}(t_0) = \mathbf{x}_0$.

Tekintsünk most minden, az $\mathbf{x}_0$ ponton átmenő és az $M[\mathbf{h},\mathbf{g}]$ halmazban haladó, folytonosan differenciálható görbét, valamint a görbék *első deriváltjai*t az $\mathbf{x_0}$ pontban, amelyek az R^n n-dimenziós Euklideszi tér vektorai. Ha az összes, $\mathbf{x}_0$ ponton átmenő és az $M[\mathbf{h},\mathbf{g}]$ halmazban haladó görbe *első deriváltjai* az R^n egy alterét határozzák meg, akkor azt az $M[\mathbf{h},\mathbf{g}]$ halmaz $\mathbf{x}_0$ pontbeli *érintősíkjá*nak nevezzük és a $TM[\mathbf{h},\mathbf{g}]_{\mathbf{x_0}}$ szimbólummal jelöljük.

Vezessük be a következő jelölést:

$$\tilde{M}[\mathbf{h},\mathbf{g}] = \{\mathbf{x} \in R^n | h_j(\mathbf{x}) = 0,\ j = 1,\dots,p,\ g_i(\mathbf{x}) = 0, \quad i \in I(\mathbf{x}_0)\}. \tag{3.3}$$

A nemlineáris optimalizálásban alapvető fontosságú az $\tilde{M}[\mathbf{h},\mathbf{g}]$ halmaz érintősíkjainak explicit megadása.

3.1. Lemma. *Ha az* $\mathbf{x}_0 \in \tilde{M}[\mathbf{h},\mathbf{g}]$ *pontban teljesül a LICQ regularitási feltétel, akkor az* $\tilde{M}[\mathbf{h},\mathbf{g}]$ *halmaz* $\mathbf{x}_0$ *pontbeli érintősíkja létezik és a következő formában adható meg:*

$$T\tilde{M}[\mathbf{h},\mathbf{g}]_{\mathbf{x}_0} = \{\mathbf{v} \in R^n |\ \nabla h_j(\mathbf{x}_0)\mathbf{v} = 0, \quad j = 1,\dots,p, \quad \nabla g_i(\mathbf{x}_0)\mathbf{v} = 0, \quad i \in I(\mathbf{x}_0)\}. \tag{3.4}$$

Bizonyítás. Ha tetszőleges, az $\mathbf{x}_0 = \mathbf{x}(t_0)$ ponton átmenő $\mathbf{x}(t)$, $t \in [a,b]$, görbe esetén valamely $j = 1,\dots,p$, indexre $\nabla h_j(\mathbf{x}_0)\mathbf{x}'(t_0) \neq 0$, vagy valamely $i \in I(\mathbf{x}_0)$ indexre $\nabla g_i(\mathbf{x}_0)\mathbf{x}'(t_0) \neq 0$, akkor biztos, hogy a görbe kilép az $\tilde{M}[\mathbf{h},\mathbf{g}]$ halmazból. Ebből következik, hogy az $\mathbf{x}_0$ ponton átmenő és az $\tilde{M}[\mathbf{h},\mathbf{g}]$ halmazban haladó görbék első deriváltjai benne vannak a $T\tilde{M}[\mathbf{h},\mathbf{g}]_{\mathbf{x}_0}$ halmazban.

Mivel alterek metszete is altér, a $T\tilde{M}[\mathbf{h},\mathbf{g}]_{\mathbf{x}_0}$ halmaz R^n-beli altér. Ezért azt kell belátni, hogy minden, a $T\tilde{M}[\mathbf{h},\mathbf{g}]_{\mathbf{x}_0}$ halmazba tartozó $\mathbf{v}$ vektor esetén létezik olyan, az $\mathbf{x}_0$ ponton átmenő és az $\tilde{M}[\mathbf{h},\mathbf{g}]$ halmazban haladó görbe, amelynek első deriváltja a $\mathbf{v}$ vektor. Ebből következik majd, hogy a fenti alakú $T\tilde{M}[\mathbf{h},\mathbf{g}]_{\mathbf{x}_0}$ éppen az $\tilde{M}[\mathbf{h},\mathbf{g}]_{\mathbf{x}_0}$ érintősíkja.

Tekintsük a következő egyenleteket:

$$\begin{aligned} h_j\Bigg(\mathbf{x}_0 + t\mathbf{v} + \sum_{l=1}^{p}\nabla h_l(\mathbf{x}_0)^T u_l(t) + \sum_{k\in I(\mathbf{x}_0)}\nabla g_k(\mathbf{x}_0)^T\tilde{u}_k(t)\Bigg) &= 0, \quad j = 1,\ldots,p, \\ g_i\Bigg(\mathbf{x}_0 + t\mathbf{v} + \sum_{l=1}^{p}\nabla h_l(\mathbf{x}_0)^T u_l(t) + \sum_{k\in I(\mathbf{x}_0)}\nabla g_k(\mathbf{x}_0)^T\tilde{u}_k(t)\Bigg) &= 0, \quad i \in I(\mathbf{x}_0), \end{aligned} \tag{3.5}$$

ahol tetszőlegesen rögzített t értékre $u_l(t),\ l = 1,\ldots,p,\ \ \tilde{u}_k(t), k \in I(\mathbf{x}_0)$, a változók. Így $p+|\,I(\mathbf{x}_0)|$ egyenletből álló és $p+|\,I(\mathbf{x}_0)|$ változót tartalmazó nemlineáris egyenletrendszert kapunk, ahol $|\,I(\mathbf{x}_0)|$ jelenti az aktív egyenlőtlenség feltételek számát.

Tekintsük a $t = 0$ pontban a (3.5) egyenleteket, amiknek az $u_l(0) = 0$, $l = 1,\ldots,p$, $\tilde{u}_k(0) = 0$, $k \in I(\mathbf{x}_0)$, értékek egy megoldását adják. A (3.5) rendszer $u_l,\ l = 1,\ldots,p,\ \tilde{u}_k,\ k \in I(\mathbf{x}_0)$, változók szerint képzett Jacobi mátrixa a $t = 0$ pontban

$$\begin{bmatrix} J\mathbf{h}(\mathbf{x}_0) \\ J\mathbf{g}(\mathbf{x}_0) \end{bmatrix} \left[J\mathbf{h}(\mathbf{x}_0)^T,\ J\mathbf{g}(\mathbf{x}_0)^T\right] = \begin{bmatrix} J\mathbf{h}(\mathbf{x}_0)J\mathbf{h}(\mathbf{x}_0)^T & J\mathbf{h}(\mathbf{x}_0)J\mathbf{g}(\mathbf{x}_0)^T \\ J\mathbf{g}(\mathbf{x}_0)J\mathbf{h}(\mathbf{x}_0)^T & J\mathbf{g}(\mathbf{x}_0)J\mathbf{g}(\mathbf{x}_0)^T \end{bmatrix}, \tag{3.6}$$

ami a LICQ regularitási feltétel miatt nem szinguláris. Ezért alkalmazni tudjuk az implicit függvény tételt, ami szerint léteznek (3.5)-öt kielégítő $u_l(t),\ l = 1,\ldots,p$, és $\tilde{u}_k(t),\ \ k \in I(\mathbf{x}_0),\ t \in (-a,a)$ folytonos függvények. Az így nyert

$$\mathbf{x}(t) = \mathbf{x}_0 + t\mathbf{v} + \sum_{j=1}^{p}\nabla h_j(\mathbf{x}_0)^T u_j(t) + \sum_{i\in I(\mathbf{x}_0)}\nabla g_i(\mathbf{x}_0)^T\tilde{u}_i(t), \quad t \in (-a,a) \tag{3.7}$$

görbék a konstrukció miatt az $\tilde{M}[\mathbf{h},\mathbf{g}]$, halmazban haladnak. Differenciáljuk a (3.5) egyenleteket a t változó szerint a $(-a,a)$ tartományban és tekintsük az eredményt a $t = 0$ pontban:

$$
\begin{aligned}
0 = \frac{d}{dt} h_j(\mathbf{x}(t))_{|t=0} &= \nabla h_j(\mathbf{x}_0)\mathbf{v} + \sum_{l=1}^{p} \nabla h_j(\mathbf{x}_0)\nabla h_l(\mathbf{x}_0)^T u_l'(0) + \\
&\sum_{i\in I(\mathbf{x}_0)} \nabla h_j(\mathbf{x}_0)\nabla g_i(\mathbf{x}_0)^T \tilde{u}_i'(0), \quad j = 1, \dots, p, \\
0 = \frac{d}{dt} g_i(\mathbf{x}(t))_{|t=0} &= \nabla g_i(\mathbf{x}_0)\mathbf{v} + \sum_{j=1}^{p} \nabla g_i(\mathbf{x}_0)\nabla h_j(\mathbf{x}_0)^T u_j'(0) + \\
&\sum_{l\in I(\mathbf{x}_0)} \nabla g_i(\mathbf{x}_0)\nabla g_l(\mathbf{x}_0)^T \tilde{u}_l'(0), \quad i \in I(\mathbf{x}_0).
\end{aligned}
\tag{3.8}
$$

A $\mathbf{v}$ vektor definíciója miatt

$$
\nabla h_j(\mathbf{x}_0)\mathbf{v} = 0, \quad j = 1, \dots, p, \quad \nabla g_i(\mathbf{x}_0)\mathbf{v} = 0, \quad i \in I(\mathbf{x}_0).
$$

A $\begin{bmatrix} J\mathbf{h}(\mathbf{x}_0) \\ J\mathbf{g}(\mathbf{x}_0) \end{bmatrix} \left[J\mathbf{h}(\mathbf{x}_0)^T,\ J\mathbf{g}(\mathbf{x}_0)^T \right]$ mátrix nemszingularitása miatt a (3.8) egyenletrendszer egyedüli megoldása

$$
u_j'(0) = 0, \quad j = 1, \dots, p, \quad \tilde{u}_i'(0) = 0, \quad i \in I(\mathbf{x}_0),
$$

ezért

$$
\mathbf{x}'(0) = \mathbf{v},
$$

tehát, a vizsgált pontban a (3.7) képlettel adott görbe érintője $\mathbf{v}$, amiből következik az állítás. ■

Megemlítjük, hogy a LICQ regularitási feltétel nem a megengedett halmazra, hanem a megengedett halmaz reprezentációjára vonatkozó feltétel. A *3.1. Példá*ban, ahol a LICQ regularitási feltétel nem teljesül a (0,0) pontban, $TM[h]_{\mathbf{0}} = R^2$, jóllehet a geometriai szemléletünk alapján ez a koordináta tengely lenne.

3.2. Lemma. *Ha az* $\mathbf{x}_0 \in \tilde{M}[\mathbf{h}, \mathbf{g}]$ *pontban teljesül a LICQ regularitási feltétel és az* $f : \tilde{M}[\mathbf{h}, \mathbf{g}] \to R$ *függvénynek az* $\mathbf{x}_0$ *pontban lokális minimuma van, akkor*

$$
\nabla f(\mathbf{x}_0)\mathbf{v} = 0, \quad \mathbf{v} \in T\tilde{M}[\mathbf{h}, \mathbf{g}]_{\mathbf{x}_0}. \tag{3.9}
$$

Bizonyítás. Legyen $\mathbf{v} \in T\tilde{M}[\mathbf{h}, \mathbf{g}]_{\mathbf{x}_0}$, és legyen $\mathbf{x}(t)$, $t \in (-a, a)$, egy, az $\mathbf{x}_0$

ponton átmenő és az $\tilde{M}[\mathbf{h},\mathbf{g}]$ halmazban haladó, folytonosan differenciálható görbe, aminek az érintője $\mathbf{v}$. Mivel az f függvénynek az $\mathbf{x}_0$ pont lokális minimuma az $\tilde{M}[\mathbf{h},\mathbf{g}]$ halmazon, ezért

$$\frac{d}{dt} f\big(\mathbf{x}(t)\big)_{|t=0} = \nabla f(\mathbf{x}_0)\mathbf{v} = 0, \tag{3.10}$$

ami az állítás. ■

A *3.2. Lemma* állítása geometriailag azt jelenti, hogy a $\nabla f(\mathbf{x}_0)$ gradiens vektor a lokális minimum pontban merőleges az érintőtérre.

3.1. Tétel. *Ha $\mathbf{x}_0$ lokális optimuma a (3.1) problémának és a LICQ regularitási feltétel teljesül ebben a pontban, akkor léteznek $\boldsymbol{\mu} \in R^p$ és $\boldsymbol{\lambda} \in R^m$ Lagrange multiplikátorok, amikre a*

$$\begin{aligned} \nabla_{\mathbf{x}} L(\mathbf{x}_0, \boldsymbol{\mu}, \boldsymbol{\lambda}) &= \mathbf{0}, \\ \boldsymbol{\lambda} &\geq \mathbf{0}, \\ \lambda_i g_i(\mathbf{x}_0) &= 0, \quad i = 1, \dots, m, \end{aligned} \tag{3.11}$$

feltételek teljesülnek.

Bizonyítás. Ha $\lambda_i \geq 0$ és $g_i(\mathbf{x}_0) \geq 0$, $i = 1, \dots, m$, akkor a $\lambda_i g_i(\mathbf{x}_0) = 0$, $i = 1, \dots, m$, egyenlőségek teljesüléséhez szükséges, hogy azon $i \in \{1, \dots, m\}$ indexekre, amelyekre $g_i(\mathbf{x}_0) > 0$ a $\lambda_i = 0$ egyenlőségek teljesüljenek.

Mivel $\mathbf{x}_0$ lokális minimum az $M[\mathbf{h},\mathbf{g}]$ halmazon és $\tilde{M}[\mathbf{h},\mathbf{g}] \subseteq M[\mathbf{h},\mathbf{g}]$ lokálisan, ezért $\mathbf{x}_0$ az f függvény lokális minimuma az $\tilde{M}[\mathbf{h},\mathbf{g}]$ halmazon. A *3.2. Lemma* miatt a

$$\max \nabla f(\mathbf{x}_0)\mathbf{v}, \quad \mathbf{v} \in T\tilde{M}[\mathbf{h},\mathbf{g}]_{\mathbf{x}_0},$$

lineáris optimalizálási feladat optimum értéke nulla, ezért a lineáris optimalizálás dualitás tétele miatt a duális problémának van megengedett megoldása, azaz léteznek $\boldsymbol{\mu} \in R^p$ és $\boldsymbol{\lambda} \in R^{|I(\mathbf{x}_0)|}$ Lagrange multiplikátorok, amikre a

$$\nabla f(\mathbf{x}_0) + \sum_{j=1}^{p} \mu_j \nabla h_j(\mathbf{x}_0) - \sum_{i \in I(\mathbf{x}_0)} \lambda_i \nabla g_i(\mathbf{x}_0) = 0$$

feltétel teljesül.

Azt kell még belátni, hogy $\lambda_i \geq 0$, $i \in I(\mathbf{x}_0)$. Tegyük fel, hogy valamilyen $\tilde{\imath}$ indexre $\lambda_{\tilde{\imath}} < 0$. A LICQ regularitási feltétel miatt van olyan $\mathbf{v}$ vektor, amelyre

$$\nabla h_j(\mathbf{x}_0)\mathbf{v} = 0, \quad j = 1, \dots, p; \quad \nabla g_i(\mathbf{x}_0)\mathbf{v} = 0, \quad i \in I(\mathbf{x}_0)\backslash\{\tilde{\imath}\}, \quad \nabla g_{\tilde{\imath}}(\mathbf{x}_0)\mathbf{v} < 0.$$

A *3.1. Lemma* miatt létezik az $\mathbf{x}_0 = \mathbf{x}(0)$ ponton áthaladó olyan $\mathbf{x}(t)$, $t \in (-a, a)$, görbe, amely a

$$h_j(\mathbf{x}) = 0, \quad j = 1, \dots, p; \quad g_i(\mathbf{x}) = 0, \quad i \in I(\mathbf{x}_0)\backslash\{\tilde{\imath}\},$$

halmazban halad és az $\mathbf{x}_0$ pontbeli érintője a $\mathbf{v}$ vektor. Belátható, hogy kicsiny $t \geq 0$ értékekre a görbe pontjai a megengedett tartományhoz tartoznak és

$$\frac{df\big(\mathbf{x}(t)\big)}{dt}_{|t=0} = \nabla f(\mathbf{x}_0)\mathbf{v} < 0,$$

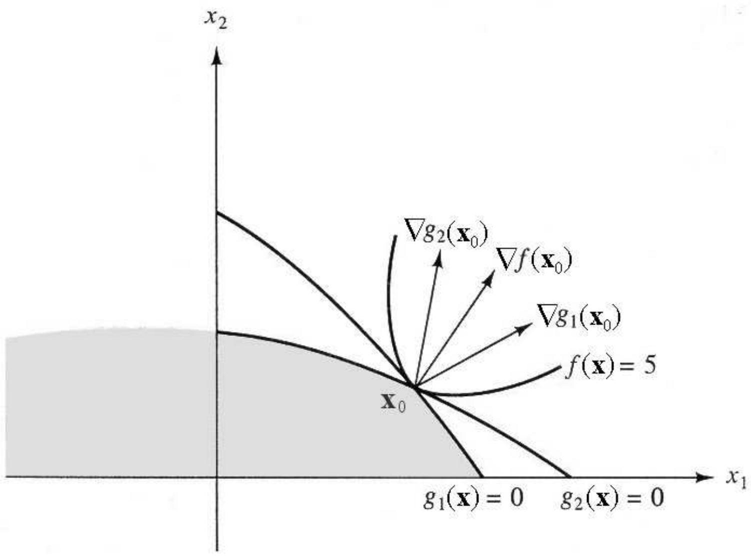
ami ellentmond az $\mathbf{x}_0$ pont lokális minimalitásának. ∎

A (3.11) elsőrendű optimalitási feltételek neve *Karush-Kuhn-Tucker feltételek*, amikből a harmadikat nevezik *komplementaritási feltétel*nek. Egy megengedett pont akkor *stacionárius*, ha a (3.11) feltételek teljesülnek. Megjegyezük, hogy a

$$\nabla_{\boldsymbol{\mu}} L(\mathbf{x}_0, \boldsymbol{\mu}, \boldsymbol{\lambda}) = \mathbf{h}(\mathbf{x}_0) = \mathbf{0},$$
$$\nabla_{\boldsymbol{\lambda}} L(\mathbf{x}_0, \boldsymbol{\mu}, \boldsymbol{\lambda}) = \mathbf{g}(\mathbf{x}_0) \geq \mathbf{0},$$

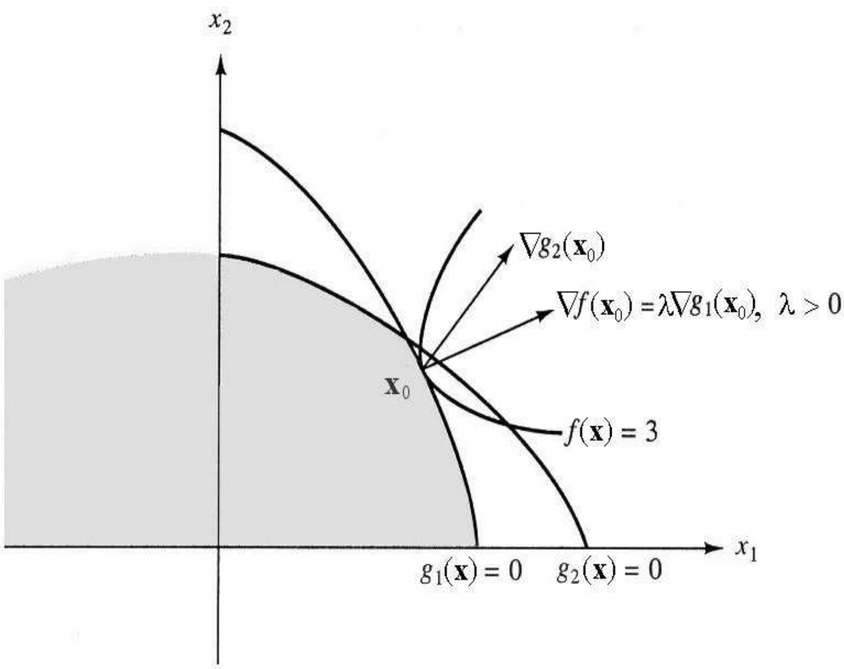
feltételek az $\mathbf{x}_0$ pont megengedettségét biztosítják.

A 4. és 5. ábra példát ad a *Karush-Kuhn-Tucker* feltételekre.



4. ábra

Példa a Karush-Kuhn-Tucker feltételekre: mind a két feltétel aktív



5. ábra

Példa a Karush-Kuhn-Tucker feltételekre: az egyik feltétel aktív, a másik nem

3.2. Példa. *Tekintsük a*

$$\min\ x_1x_2 + x_2x_3 + x_1x_3$$

$$x_1 + x_2 + x_3 = 3, \quad (x_1, x_2, x_3) \in R^3,$$

nemlineáris optimalizálási problémát. Az elsőrendű szükséges feltételek és a feltételi egyenlőség:

$$\begin{array}{ccccccccc} & & x_2 & + & x_3 & + & \mu & = & 0, \\ x_1 & + & & & x_3 & + & \mu & = & 0, \\ x_1 & + & x_2 & & & + & \mu & = & 0, \\ x_1 & + & x_2 & + & x_3 & & & = & 3. \end{array}$$

Ez négy egyenletből és négy ismeretlenből álló lineáris egyenletrendszer, aminek a megoldása

$$x_1 = x_2 = x_3 = 1, \quad \mu = -2.$$

Ha a (3.1) NLO célfüggvénye és feltételi függvényei kétszer folytonosan differenciálhatóak (vagy röviden C^2 függvények), akkor a lokális optimalitás jellemzése a másodrendű szükséges és elegendő feltételekkel történik.

3.2. Tétel. *Tegyük fel, hogy a (3.1) NLO célfüggvénye és feltételi függvényei kétszer folytonosan differenciálhatók. Ha $\mathbf{x}_0$ a (3.1) NLO lokális optimuma és a LICQ regularitási feltétel teljesül ebben a pontban, akkor léteznek olyan $\boldsymbol{\mu} \in R^p$ és $\boldsymbol{\lambda} \in R^m$ Lagrange multiplikátor vektorok, amelyekre a*

$$\begin{aligned} \nabla_{\mathbf{x}} L(\mathbf{x}_0, \boldsymbol{\mu}, \boldsymbol{\lambda}) &= \mathbf{0}, \\ \boldsymbol{\lambda} &\geq \mathbf{0}, \\ \lambda_i g_i(\mathbf{x}_0) &= 0, \quad i = 1, \ldots, m, \end{aligned} \tag{3.12}$$

$$\mathbf{v}^T HL(\mathbf{x}_0, \boldsymbol{\mu}, \boldsymbol{\lambda})\mathbf{v} \geq 0, \qquad \mathbf{v} \in T\tilde{M}[\mathbf{h}, \mathbf{g}]_{\mathbf{x_0}} \tag{3.13}$$

feltételek teljesülnek, ahol

$$T\tilde{M}[\mathbf{h},\mathbf{g}]_{\mathbf{x_0}} = \{\mathbf{v} \in R^n |\ \nabla h_j(\mathbf{x}_0)\mathbf{v} = 0, \quad j = 1,\dots,p, \quad \nabla g_i(\mathbf{x}_0)\mathbf{v} = 0, \quad i \in I(\mathbf{x}_0)\},$$

és a tételben szereplő $HL(\mathbf{x}_0, \boldsymbol{\mu}, \boldsymbol{\lambda})$ *kifejezés a Lagrange függvény* $\mathbf{x}$ *változó szerint képzett Hesse mátrixát jelenti az* $\mathbf{x}_0$ *pontban.*

Bizonyítás. A *3.1. Tétel* miatt a (3.11) elsőrendű feltételek teljesülnek. Az $\mathbf{x}_0$ pont lokális minimalitása miatt, az $\mathbf{x}_0$ ponton átmenő és az $\tilde{M}[\mathbf{h},\mathbf{g}]$ halmazban haladó, kétszer folytonosan differenciálható görbékre teljesül a

$$\frac{d^2 f\big(\mathbf{x}(t)\big)}{dt^2}_{|t=0} = \mathbf{x}'(0)^T Hf(\mathbf{x}_0)\mathbf{x}'(0) + \nabla f(\mathbf{x}_0)\mathbf{x}''(0) \geq 0 \tag{3.14}$$

egyenlőtlenség. A

$$\sum_{j=1}^{p} \mu_j h_j\big(\mathbf{x}(t)\big) - \sum_{i\in I(\mathbf{x}_0)} \lambda_i g_i\big(\mathbf{x}(t)\big) = 0, \quad t \in (-a, a),$$

egyenlőséget kétszer differenciálva azt kapjuk, hogy

$$\begin{gathered} \sum_{j=1}^{p} \mu_j \mathbf{x}'(0)^T Hh_j(\mathbf{x}_0)\mathbf{x}'(0) + \sum_{j=1}^{p} \mu_j \nabla h_j(\mathbf{x}_0)\mathbf{x}''(0) - \\ \sum_{i\in I(\mathbf{x}_0)} \lambda_i \mathbf{x}'(0)^T Hg_i(\mathbf{x}_0)\mathbf{x}'(0) - \sum_{i\in I(\mathbf{x}_0)} \lambda_i \nabla g_i(\mathbf{x}_0)\mathbf{x}''(0) = 0. \end{gathered} \tag{3.15}$$

A (3.14) és a (3.15) egyenlőségek összeadása után adódik, hogy

$$\frac{d^2 f\big(\mathbf{x}(t)\big)}{dt^2}_{|t=0} = \mathbf{x}'(0)^T HL(\mathbf{x}_0, \boldsymbol{\mu}, \boldsymbol{\lambda})\mathbf{x}'(0) \geq 0.$$

Mivel $\mathbf{x}'(0) \in T\tilde{M}[\mathbf{h},\mathbf{g}]_{\mathbf{x}_0}$ tetszőleges, ezért az állítást bebizonyítottuk. ■

A (3.12) és (3.13) feltételek az optimalitás első- és másodrendű szükséges feltételei. Az optimalitás másodrendű elegendő feltételeit a következő állításban fogalmazzuk meg:

3.3. Tétel. T*együk fel, hogy a (3.1) NLO célfüggvénye és feltételi függvényei kétszer folytonosan differenciálhatók. Ha* $\mathbf{x}_0$ *a (3.1) NLO egy megengedett megoldása, amire a LICQ regularitási feltétel teljesül és léteznek olyan* $\boldsymbol{\mu} \in R^p$ *és* $\boldsymbol{\lambda} \in R^m$ *vektorok, amelyekre a (3.11) feltételek és a*

$$\mathbf{v}^T HL(\mathbf{x}_0, \boldsymbol{\mu}, \boldsymbol{\lambda})\mathbf{v} > 0, \qquad \mathbf{v} \in T\hat{M}[\mathbf{h}, \mathbf{g}]_{\mathbf{x_0}}, \qquad \mathbf{v} \neq \mathbf{0}, \tag{3.16}$$

egyenlőtlenségek teljesülnek, ahol

$$T\hat{M}[\mathbf{h}, \mathbf{g}]_{\mathbf{x}_0} = \{\mathbf{v} \in R^n | \nabla h_j(\mathbf{x}_0)\mathbf{v} = 0, \quad j = 1, \ldots, p,$$
$$\nabla g_i(\mathbf{x}_0)\mathbf{v} = 0, \quad i \in I(\mathbf{x}_0) \cap \{i | \lambda_i > 0\}\},$$

akkor $\mathbf{x}_0$ *a (3.1) NLO szigorú lokális minimuma.*

Bizonyítás. A tétel állítását indirekt módon bizonyítjuk. Tegyük fel, hogy az $\mathbf{x}_0$ pont nem az NLO szigorú lokális minimuma. Emiatt létezik olyan $\mathbf{x}_k \in M[\mathbf{h}, \mathbf{g}], \mathbf{x}_k \neq \mathbf{x}_0$, $k = 1, 2, \ldots,$ sorozat, amely tart az $\mathbf{x}_0$ ponthoz és amelyre $f(\mathbf{x}_k) \leq f(\mathbf{x}_0)$, $\forall k$, esetén. Írjuk az $\mathbf{x}_k$ sorozatot az

$$\mathbf{x}_k = \mathbf{x}_0 + \delta_k \mathbf{s}_k, \quad \mathbf{s}_k \in R^n, \quad \|\mathbf{s_k}\| = \mathbf{1}, \quad \delta_k > 0, \quad k = 1, 2, \ldots,$$

formába. Mivel $\lim_{k \to \infty} \delta_k \to 0$ és az $\mathbf{s}_k$ sorozat korlátos, ezért van konvergens részszorozata. Az általánosság megszorítása nélkül feltehetjük, hogy $\lim_{k \to \infty} \mathbf{s}_k \to \mathbf{s}_0$. Mivel $\mathbf{h}(\mathbf{x}_k) - \mathbf{h}(\mathbf{x}_0) = 0$, ezért

$$\lim_{k \to \infty} \frac{\mathbf{h}(\mathbf{x}_k) - \mathbf{h}(\mathbf{x}_0)}{\delta_k} = J\mathbf{h}(\mathbf{x}_0)\mathbf{s}_0 = \mathbf{0}.$$

Minden aktív egyenlőtlenség feltételre igaz az, hogy

$$g_i(\mathbf{x}_k) - g_i(\mathbf{x}_0) \geq 0, \quad i \in I(\mathbf{x}_0),$$

amiből következik, hogy

$$\nabla g_i(\mathbf{x}_0)\mathbf{s}_0 \geq 0, \quad i \in I(\mathbf{x}_0).$$

Először tegyük fel, hogy

$$\nabla g_i(\mathbf{x}_0)\mathbf{s}_0 = 0, \quad i \in I(\mathbf{x}_0) \cap \{i \mid \lambda_i > 0\}.$$

A *Taylor tétel*t alkalmazva,

$$0 = h_j(\mathbf{x}_k) = h_j(\mathbf{x}_0) + \delta_k \nabla h_j(\mathbf{x}_0)\mathbf{s}_k + \frac{\delta_k^2}{2}\mathbf{s}_k^T Hh_j(\boldsymbol{\eta}_j)\mathbf{s}_k, \quad j = 1, \ldots, p, \tag{3.17}$$

$$0 \leq g_i(\mathbf{x}_k) = g_i(\mathbf{x}_0) + \delta_k \nabla g_i(\mathbf{x}_0)\mathbf{s}_k + \frac{\delta_k^2}{2}\mathbf{s}_k^T Hg_i(\tilde{\boldsymbol{\eta}}_i)\mathbf{s}_k, \quad i \in I(\mathbf{x}_0), \tag{3.18}$$

$$0 \geq f(\mathbf{x}_k) - f(\mathbf{x}_0) = \delta_k \nabla f(\mathbf{x}_0)\mathbf{s}_k + \frac{\delta_k^2}{2}\mathbf{s}_k^T Hf(\boldsymbol{\eta}_0)\mathbf{s}_k, \tag{3.19}$$

ahol minden $\boldsymbol{\eta}_j, \tilde{\boldsymbol{\eta}}_i, \boldsymbol{\eta}_0$ egy-egy pontot jelent az $[\mathbf{x}_0, \mathbf{x_k}]$ szakaszon. A (3.17) egyenlőségeket μ_j-vel, a (3.18) egyenlőségeket $-\lambda_i$-vel megszorozva, majd a (3.17) és (3.18) egyenlőségeket a (3.19) egyenlőtlenséghez hozzáadva azt kapjuk, hogy

$$0 \geq \frac{\delta_k^2}{2}\mathbf{s}_k^T\left(Hf(\boldsymbol{\eta}_0) + \sum_{j=1}^{p}\mu_j Hh_j(\boldsymbol{\eta}_j) - \sum_{i\in I(\mathbf{x}_0)}\lambda_i Hg_i(\tilde{\boldsymbol{\eta}}_i)\right)\mathbf{s}_k,$$

ami a $k \to \infty$ határátmenet elvégzése után ellentmond a (3.16) feltételnek.

Ha létezik olyan $\tilde{\imath} \in I(\mathbf{x}_0) \cap \{i \mid \lambda_i > 0\}$ index, amelyre

$$\nabla g_{\tilde{\imath}}(\mathbf{x}_0)\mathbf{s}_0 > 0,$$

akkor az optimalitás elsőrendű feltételeiből adódik, hogy

$$0 \geq \nabla f(\mathbf{x}_0)\mathbf{s}_0 = -\sum_{j=1}^{p}\mu_j \nabla h_j(\mathbf{x}_0)\mathbf{s}_0 + \sum_{i\in I(\mathbf{x}_0)}\lambda_i \nabla g_i(\mathbf{x}_0)\mathbf{s}_0 > 0,$$

ami ellentmondás. Ezzel bebizonyítottuk az állítást. ■

3.3. Példa. *Tekintsük a*

$$\max\ x_1x_2 + x_2x_3 + x_1x_3$$
$$x_1 + x_2 + x_3 = 3, \quad (x_1, x_2, x_3) \in R^3,$$

nemlineáris optimalizálási problémát. A 3.2. Példában kiszámoltuk, hogy az elsőrendű szükséges feltételeket az

$$x_1 = x_2 = x_3 = 1, \quad \lambda = -2,$$

értékek elégítik ki. A

$$HL\big((1,1,1),-2\big) = Hf(1,1,1) = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

mátrix se nem pozitív, se nem negatív definit mátrix.

Ha a fenti mátrixot a

$$TM[h]_{(1,1,1)} = \{\mathbf{v} \in R^3 |\, v_1 + v_2 + v_3 = 0\}$$

altéren vizsgáljuk, akkor a

$$(v_1, v_2, -v_1 - v_2) \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ -v_1 - v_2 \end{pmatrix} = -v_1^2 - v_2^2 - (v_1 + v_2)^2 \leq 0$$

egyenlőtlenség miatt megállapítható, hogy a Lagrange függvény Hesse mátrixa a vizsgált pontban a $TM[h]_{(1,1,1)}$ altérre megszorítva negatív definit, tehát az (1,1,1) pont szigorú lokális maximum.

3.4. Példa *Legyen*

$$f(x) = 2 - (x-1)^2, \qquad ha \quad 0 \leq x < 3,$$
$$f(x) = -3 + (x-4)^2, \quad ha \quad 3 \leq x \leq 6.$$

Oldjuk meg a

$$\max f(x)$$
$$0 \leq x \leq 6$$

feladatot. Az f függvény a 6. ábrán látható.

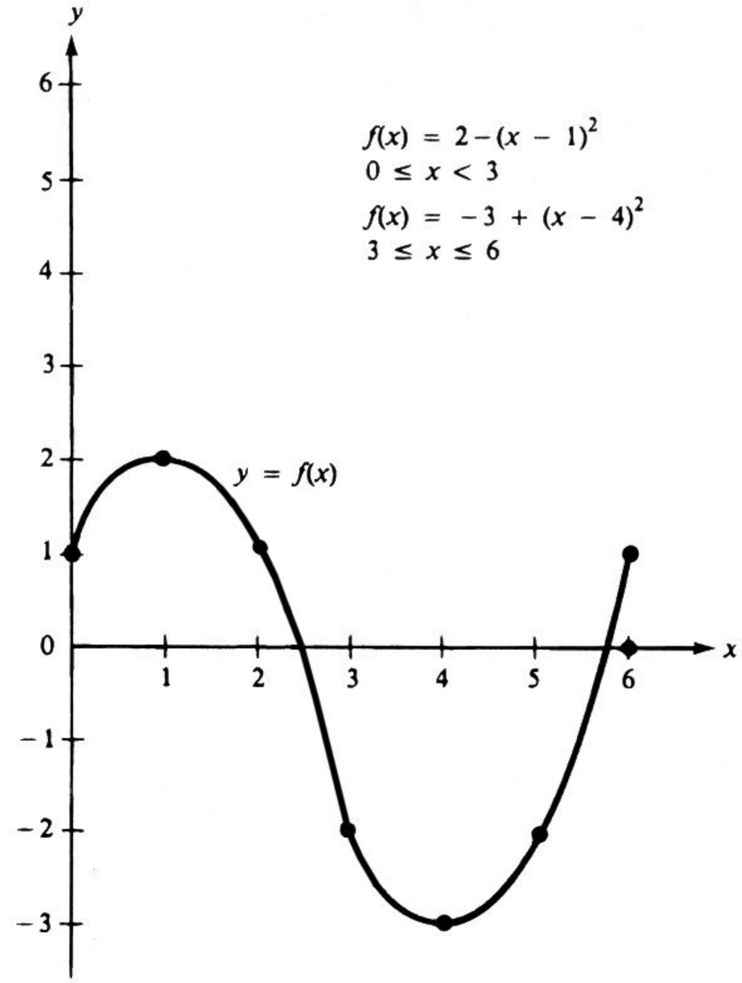
Megoldás:

1. eset: Ha $0 \leq x < 3$, akkor $f'(x) = -2(x-1)$ és $f''(x) = -2$. Ha $3 < x \leq 6$, akkor $f'(x) = 2(x-4)$ és $f''(x) = 2$. Ezért $f'(1) = f'(4) = 0$. Mivel $f''(1) < 0$, az $x = 1$ lokális maximum. Mivel $f''(4) > 0$, az $x = 4$ lokális minimum.

2. eset: Egyszerűen belátható, hogy $f(x)$ nem deriválható az $x = 3$ pontban (ha x kicsit kisebb, mint 3, akkor $f'(x)$ a -4 értékhez közelít, és ha x kicsit nagyobb, mint 3, akkor $f'(x)$ a -2 értékhez közelít). Mivel $f(2.9) = -1.61, f(3) = -2$ és $f(3.1) = -2.19$, és az f függvény a $[2.9,\ 3.1]$ intervallumban szigorúan monoton csökkenő, ezért az $x = 3$ nem lokális szélsőértékhely.

3. eset: Mivel $f'(0) = 2 > 0$, az $x = 0$ lokális minimum. Mivel $f'(6) = 4 > 0$, az $x = 6$ lokális maximum.

Tehát, a $[0, 6]$ intervallumban az $f(x)$ függvénynek az $x = 1$ és az $x = 6$ pontban van szigorú lokális maximuma. Mivel $f(1) = 2$ és $f(6) = 1$, azt kapjuk, hogy az NLO optimális megoldása az $x = 1$ pontban van.



6. ábra

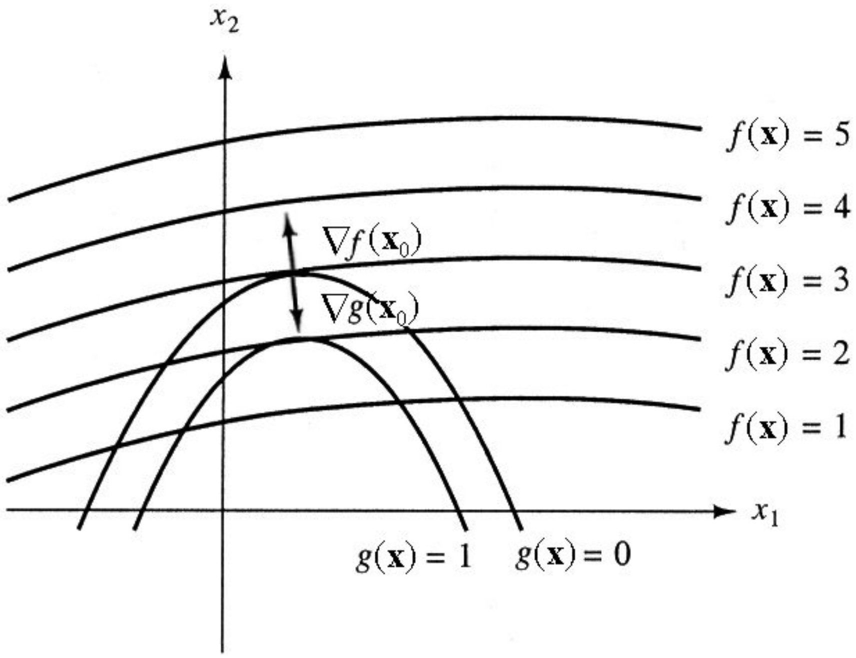
A 3.4. Példa függvénye

3.5. Példa. *Tegyük fel, hogy meg akarjuk határozni az* $x_1 + 3x_2 = 5$ *egyenletű egyenes origóhoz legközelebbi pontját. Minthogy a minimum helye szempontjából mindegy, hogy a távolságot, vagy a távolság négyzetét minimalizáljuk, azért feladatunk az* $x_1^2 + x_2^2$ *függvény minimalizálása az* $x_1 + 3x_2 = 5$ *feltétel mellett. Ez nagyon egyszerűen megoldható, mivel az* $x_1 + 3x_2 = 5$ *feltételből kifejezzük, mondjuk* x_1*-et és az*

$$x_1^2 + x_2^2 = (5 - 3x_2)^2 + x_2^2, \qquad x_2 \in R,$$

függvény minimalizálása az optimalitási feltételek felhasználásával elvégezhető. Minimumhelynek $x_2 = 3/2$ *és* $x_1 = 5 - 3 \cdot 3/2 = 1/2$ *adódik.*

A 7. ábra azt mutatja, hogy egy-feltételes NOP esetén a stacionárius pont(ok)ban a célfüggvény és feltétel gradiense egy egyenesbe esik.



7. ábra

Egy-feltételes példa

4. fejezet

KONVEX OPTIMALIZÁLÁS

A nemlineáris optimalizálás optimalitási feltételeire vonatkozó tételek a vizsgált pont egy környezetében igazak. Egy fontos feladatosztály, a konvex optimalizálási feladatok (KO) esetében a lokális információk globálisak. Pontosabban fogalmazva, a következők igazak:

1. a KO lokális optimuma globális optimum;
2. az optimalitás elsőrendű szükséges feltétele már elegendő az optimalitáshoz, azaz a *Karush-Kuhn-Tucker feltételek*, vagy más néven a *Lagrange multiplikátor szabály* az optimalitást jellemzi;
3. jól használható dualitás elmélet van kidolgozva;
4. a KO megoldása könnyen visszavezethető konvex függvények egy sorozatának feltétel nélküli minimalizálására;
5. a KO megoldása során a lokális optimalizáló algoritmusok lépéseinek végrehajtása után a célfüggvény értéke monoton csökken.

A most következő részben a konvex halmazokkal és függvényekkel kapcsolatos néhány fogalmat és állítást ismertetünk Rockafellar (1970), valamint Roberts és Varberg (1973) klasszikus könyveinek a tárgyalását követve.

4.1. Definíció. *Egy $A \subseteq R^n$ részhalmazt konvexnek nevezünk, ha bármely $\mathbf{x}_1, \mathbf{x}_2 \in A$ és* $0 \leq \lambda \leq 1$ esetén

$$(1-\lambda)\mathbf{x}_1 + \lambda\mathbf{x}_2 \in A. \tag{4.1}$$

Ez a definíció geometriailag azt fejezi ki, hogy két tetszőleges A halmazbeli pontot összekötő zárt szakasz is a halmazhoz tartozik. Így, minden affin halmaz, pl. az üres halmaz és az R^n n-dimenziós Euklideszi tér is konvex.

4.1. Tétel. *Konvex halmazok tetszőleges összességének a metszete is konvex.*

Bizonyítás. A konvex halmazok a tetszőleges két pontjukat összekötő egyenes szakaszokat is tartalmazzák, ezért azokat a metszet is tartalmazza. ■

Adott R^n-beli részhalmaz konvex burka az őt tartalmazó összes konvex halmaz metszete. Ez az adott részhalmazt tartalmazó legkisebb konvex halmaz. Egy konvex halmaz lezártja és (relatív) belseje szintén konvex. Egy nem üres konvex halmaz relatív belseje nem üres [62]. Általában, konvex halmazok uniója nem konvex. Két konvex halmaz esetén a mindkét halmazban található legnagyobb konvex halmaz egyértelműen meghatározott, mivel az a metszetük. A félterek nagyon fontos szerepet játszanak a konvex halmazok körében. Bármely nemnulla $\mathbf{c} \in R^n$ vektorra és $\alpha \in R$ valós számra a hozzá tartozó hipersíkot, ami $(n-1)$-dimenziós affin halmaz, a következő alakban adjuk meg:

$$\{\mathbf{x} \in R^n |\ \mathbf{c}^T\mathbf{x} = \alpha\}.$$

Bármely R^n-beli hipersík két zárt (nyílt) félteret ad meg:

$$\{\mathbf{x} \in R^n |\ \mathbf{c}^T\mathbf{x} \geq (>)\alpha\}, \quad \{\mathbf{x} \in R^n |\ \mathbf{c}^T\mathbf{x} \leq (<)\alpha\}.$$

Mind a négy fenti halmaz nemüres és konvex. Megjegyezzük, hogy ugyanazokat a féltereket kapjuk, ha $\mathbf{c}$ és α helyett az $\lambda\mathbf{c}$ vektort és a $\lambda\alpha$ számot tekintjük valamilyen $\lambda \neq 0 \in R$ értékre. Egy féltér homogén, ha az α értéke nulla. A konvexitási tulajdonság megőrződik számos algebrai művelet esetén.

4.2. Tétel. *Ha $A \subseteq R^n$ konvex halmaz, akkor bármely $A + \mathbf{a}$ eltolja és bármely λA skalárszorosa is konvex, ahol $\mathbf{a} \in R^n$, $A+\mathbf{a} = \{\mathbf{x}+\mathbf{a} \mid \mathbf{x} \in A\}$ és $\lambda A = \{\lambda\mathbf{x} \mid \mathbf{x} \in A\}$.*

Ha $A_1 \in R^n$ és $A_2 \in R^n$ konvex halmazok, akkor az $A_1 \pm A_2$ összegük és különbségük is konvex, ahol

$$A_1 \pm A_2 = \{\mathbf{x_1} \pm \mathbf{x_2} | \, \mathbf{x_1} \in A_1, \mathbf{x_2} \in A_2\}. \tag{4.2}$$

Ha $A_1, \ldots, A_m \in R^n$ konvex halmazok, akkor a $\lambda_1 A_1 + \cdots + \lambda_m A_m$ lineáris kombináció is konvex halmaz, ahol $\lambda_i \in R$, $i = 1, \ldots, m$.

Bizonyítás. Az állítások a definíciókból közvetlenül következnek. ■

Geometriailag, ha $\lambda > 0$, akkor λA az A halmaz λ-szoros nagyítása vagy kicsinyítése. A halmazalgebra egy a konvexitáshoz kötődő eredménye a következő:

4.3. Tétel. *Ha A konvex halmaz és $\lambda_1 \geq 0$, $\lambda_2 \geq 0$, akkor*

$$(\lambda_1 + \lambda_2)A = \lambda_1 A + \lambda_2 A. \tag{4.3}$$

Továbbá, A akkor és csak akkor konvex, ha a (4.3) egyenlőség bármely $\lambda_1 \geq 0$, $\lambda_2 \geq 0$ esetén teljesül.

Bizonyítás. Tetszőleges A halmaz esetén igaz az, hogy

$$(\lambda_1 + \lambda_2)A \subseteq \lambda_1 A + \lambda_2 A.$$

Az ellentétes irányú tartalmazási reláció a konvexitásból következik feltéve, hogy $\lambda_1 + \lambda_2 > 0$, mivel

$$A \supseteq \big(\lambda_1/(\lambda_1 + \lambda_2)\big)A + \big(\lambda_2/(\lambda_1 + \lambda_2)\big)A.$$

Ha $\lambda_1 = 0$ vagy $\lambda_2 = 0$, az állítás igaz.

A fordított állítás bizonyításához tegyük fel, hogy a (4.3) disztributív szabály teljesül egy adott $A \subseteq R^n$ halmazon. Ha $\lambda_1 = \lambda$, $\lambda_2 = (1 - \lambda)$, $0 \leq \lambda \leq 1$, akkor azt kapjuk, hogy

$$A = \lambda A + (1-\lambda)A,$$

tehát az A halmaz konvex. ■

Legyen $\mathbf{B} : R^n \to R^m$ tetszőleges lineáris leképezés és

$$\begin{aligned} \mathbf{B}A &= \{\mathbf{B}\mathbf{x} | \mathbf{x} \in A \subseteq R^n\}, \\ \mathbf{B}^{-1}A_1 &= \{\mathbf{x} | \mathbf{B}\mathbf{x} \in A_1 \subseteq R^m\}. \end{aligned} \tag{4.4}$$

$\mathbf{B}A$ az A halmaz $\mathbf{B}$ transzformációval nyert képe és $\mathbf{B}^{-1}A_1$ az A_1 halmaz $\mathbf{B}^{-1}$ inverzével nyert képe. A következő állítás azt mutatja meg, hogy a konvexitási tulajdonság ilyen transzformációkat alkalmazva megőrződik.

4.4. Tétel. *Legyen* $\mathbf{B}$ *egy* R^n*-ről* R^m*-re ható lineáris leképezés. Akkor a* $\mathbf{B}A$ *halmaz* R^m*-beli konvex halmaz bármely* $A \subseteq R^n$ *konvex halmaz esetén, és a* $\mathbf{B}^{-1}A_1$ *halmaz* R^n*-beli konvex halmaz bármely* $A_1 \subseteq R^m$ *konvex halmaz esetén.*

Bizonyítás. Az állítás a definíciókból közvetlenül következik. ■

A tétel következménye, hogy konvex halmaz altérre történő merőleges vetülete is konvex, mivel a merőleges vetítés lineáris leképezés.

4.5. Tétel. *Ha* $A_1 \subseteq R^n$ *és* $A_2 \subseteq R^m$ *konvex halmazok, akkor az*

$$A_1 \times A_2 = \{(\mathbf{x}_1, \mathbf{x}_2) | \, \mathbf{x}_1 \in A_1, \mathbf{x}_2 \in A_2\} \tag{4.5}$$

direkt összegük R^{n+m}*-beli konvex halmaz.*

Bizonyítás. Az állítás a definíciókból közvetlenül következik. ■

Másik fontos konvexitási fogalom a függvény konvexitás.

4.1. Definíció. *Ha* $A \subseteq R^n$ *konvex halmaz és* $f : A \to R$ *függvény, akkor* f *konvex, ha az*

$$f((1-\lambda)\mathbf{x}_1 + \lambda\mathbf{x}_2) \leq (1-\lambda)f(\mathbf{x}_1) + \lambda f(\mathbf{x}_2) \tag{4.6}$$

egyenlőtlenség teljesül bármely $\mathbf{x}_1, \mathbf{x}_2 \in A$ és $0 \leq \lambda \leq 1$ *esetén.*

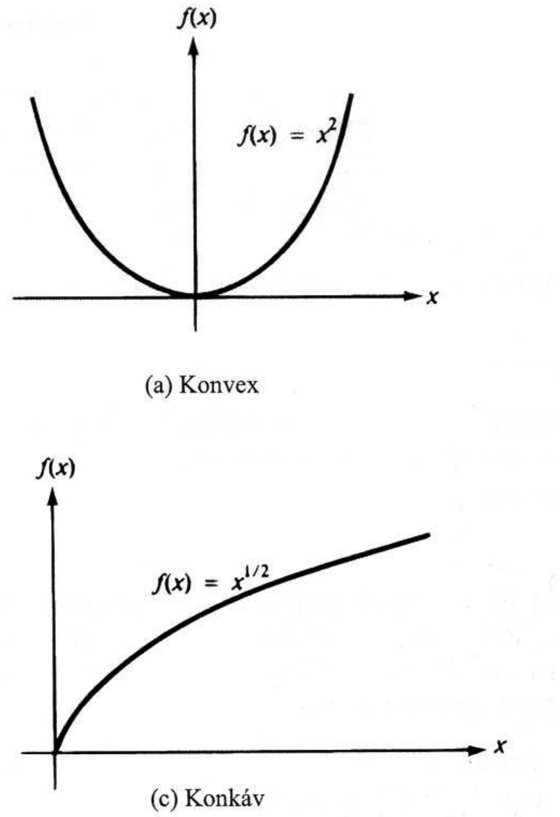
Az f függvény szigorúan konvex, ha a (4.6) egyenlőtlenség szigorúan teljesül $0 < \lambda < 1$ és $\mathbf{x}_1 \neq \mathbf{x}_2$ esetén. Konkáv függvények esetén a (4.6) egyenlőtlenségben a "$\geq$" szerepel, affin függvények esetén pedig egyenlőség. Nyílt és konvex halmazon értelmezett konvex függvények fontos tulajdonsága a folytonosság.

Az $A \subseteq R^n$ konvex halmazon értelmezett konvex függvényekre teljesül a *Jensen egyenlőtlenség*:

$$f\left(\sum_{i=1}^{n} \lambda_i \mathbf{x}_i\right) \leq \sum_{i=1}^{n} \lambda_i f(\mathbf{x}_i),$$

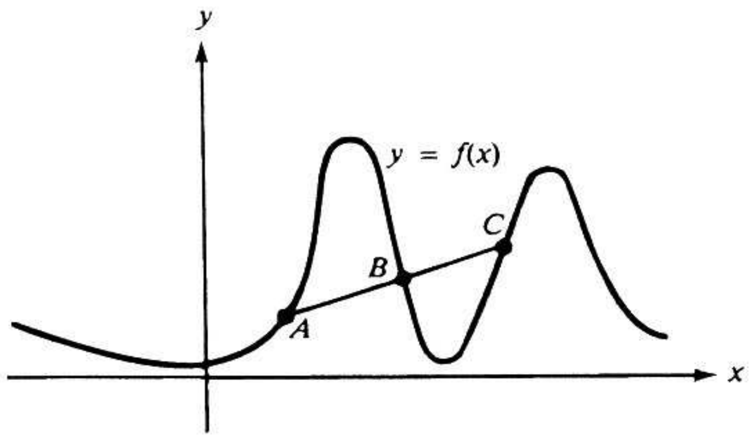
ahol $\mathbf{x}_i \in A$, $i = 1, \ldots, n$, $\lambda_i \geq 0$, $i = 1, \ldots, n$, és $\sum_{i=1}^{n} \lambda_i = 1$. (Lásd pl. Dancs, Magyarkúti, Medvegyev, Puskás és Tallos, 2003, 214. old.)

A 8. ábra példát ad konvex és konkáv függvényekre, a 9. ábra pedig egy függvényre, amely se nem konvex, se nem konkáv.



8. ábra

Példák konvex és konkáv függvényekre



9. ábra

Egy függvény, amely se nem konvex, se nem konkáv

A következő tételben a differenciálható konvex függvényeket jellemezzük.

4.6. Tétel. *Legyen $A \subseteq R^n$ nyílt konvex halmaz és $f : A \to R$ differenciálható függvény. Az f függvény akkor és csak akkor konvex az A halmazon, ha az*

$$f(\mathbf{x}_2) - f(\mathbf{x}_1) \geq \nabla f(\mathbf{x}_1)(\mathbf{x}_2 - \mathbf{x}_1) \tag{4.7}$$

egyenlőtlenség teljesül minden $\mathbf{x_1}, \mathbf{x_2} \in A$ esetén.

Bizonyítás.

I. Ha az f függvény konvex az A halmazon, akkor a (4.6) egyenlőtlenség teljesül bármely $\mathbf{x}_1, \mathbf{x}_2 \in A$, $\mathbf{x}_1 \neq \mathbf{x}_2$ és $0 \leq \lambda \leq 1$ esetén, amiből, $\lambda > 0$ mellett, átrendezéssel azt kapjuk, hogy

$$f(\mathbf{x}_2) - f(\mathbf{x}_1) \geq \frac{f\big(\mathbf{x}_1 + \lambda(\mathbf{x}_2 - \mathbf{x}_1)\big) - f(\mathbf{x}_1)}{\lambda}.$$

Ha a λ értékével tartunk nullához, akkor pontosan a (4.7) egyenlőtlenséget kapjuk. Ha $\mathbf{x}_1 = \mathbf{x}_2$, akkor (4.7) triviálisan igaz.

II. Legyen $\mathbf{x}_1, \mathbf{x}_2 \in A$ és $0 \leq \lambda \leq 1$. Mivel az A halmaz konvex, $(1-\lambda)\mathbf{x}_1 + \lambda\mathbf{x}_2 \in A$. Mivel a (4.7) egyenlőtlenség teljesül minden $\mathbf{x}_1, \mathbf{x}_2 \in A$ esetén, ezért

$$f(\mathbf{x}_1) - f\big((1-\lambda)\mathbf{x}_1 + \lambda\mathbf{x}_2\big) \geq \lambda \nabla f\big((1-\lambda)\mathbf{x}_1 + \lambda\mathbf{x}_2\big)(\mathbf{x}_1 - \mathbf{x}_2),$$
$$f(\mathbf{x}_2) - f\big((1-\lambda)\mathbf{x}_1 + \lambda\mathbf{x}_2\big) \geq -(1-\lambda) \nabla f\big((1-\lambda)\mathbf{x}_1 + \lambda\mathbf{x}_2\big)(\mathbf{x}_1 - \mathbf{x}_2).$$

Szorozzuk be az első egyenlőtlenséget $(1-\lambda)$-val, a másodikat λ-val, majd a két egyenlőtlenséget összeadva azt kapjuk, hogy

$$(1-\lambda)f(\mathbf{x}_1) + \lambda f(\mathbf{x}_2) \geq f\big((1-\lambda)\mathbf{x}_1 + \lambda\mathbf{x}_2\big),$$

ami éppen az állítás. ■

A *4.6. Tétel* geometriai jelentése az, hogy differenciálható konvex függvény tetszőleges pontban vett lineáris közelítése (érintősík) mindig a függvény gráfja alatt marad.

Ha az f függvény konvex az $A \subseteq R^n$ nyílt konvex halmazon, akkor mindig létezik egyik oldali iránymenti derivált. A *4.6. Tétel* nem pontos megfelelője az egyváltozós függvényekre bizonyított állításnak, ami szerint f akkor és csak akkor konvex, ha a differenciálhányados függvény monoton növekvő. Többváltozós függvényekre hasonló állítás fogalmazható meg a monotonitási tulajdonság felhasználásával.

4.7. Tétel. *Legyen $A \subseteq R$ nyílt konvex halmaz és $f : A \to R^n$ differenciálható függvény. Az f függvény akkor és csak akkor konvex az A halmazon, ha a*

$$(\nabla f(\mathbf{x}_2) - \nabla f(\mathbf{x}_1))(\mathbf{x}_2 - \mathbf{x}_1) \geq 0 \tag{4.8}$$

egyenlőtlenség teljesül minden $\mathbf{x}_1, \mathbf{x}_2 \in A$ esetén.

Bizonyítás.

I. Tegyük fel, hogy az f függvény konvex az A halmazon és $\mathbf{x}_1, \mathbf{x}_2 \in A$. A *4.6. Tétel* alkalmazásával azt kapjuk, hogy

$$f(\mathbf{x}_2) - f(\mathbf{x}_1) - \nabla f(\mathbf{x}_1)(\mathbf{x}_2 - \mathbf{x}_1) \geq 0,$$

és

$$f(\mathbf{x}_1) - f(\mathbf{x}_2) - \nabla f(\mathbf{x}_2)(\mathbf{x}_1 - \mathbf{x}_2) \geq 0.$$

A két egyenlőtlenséget összeadva kapjuk az állítást:

$$\big(\nabla f(\mathbf{x}_2) - \nabla f(\mathbf{x}_1)\big)(\mathbf{x}_2 - \mathbf{x}_1) \geq 0.$$

II. Ha $\mathbf{x}_1, \mathbf{x}_2 \in A$, akkor $0 \leq \lambda \leq 1$ esetén $(1-\lambda)\mathbf{x}_1 + \lambda\mathbf{x}_2 \in A$. A *Taylor tétel* alkalmazásával azt kapjuk, hogy

$$f(\mathbf{x}_2) - f(\mathbf{x}_1) = \nabla f\big(\mathbf{x}_1 + \tilde{\lambda}(\mathbf{x}_2 - \mathbf{x}_1)\big)(\mathbf{x}_2 - \mathbf{x}_1)$$

valamely $0 < \tilde{\lambda} < 1$ értékre.

Mivel a (4.8) egyenlőtlenség teljesül minden $\mathbf{x}_1, \mathbf{x}_2 \in A$ esetén, ezért

$$\Big(\nabla f\big(\mathbf{x}_1 + \tilde{\lambda}(\mathbf{x}_2 - \mathbf{x}_1)\big) - \nabla f(\mathbf{x}_1)\Big)\tilde{\lambda}(\mathbf{x}_2 - \mathbf{x}_1) \geq 0,$$

amiből az következik, hogy

$$\nabla f\big(\mathbf{x}_1 + \tilde{\lambda}(\mathbf{x}_2 - \mathbf{x}_1)\big)(\mathbf{x}_2 - \mathbf{x}_1) \geq \nabla f(\mathbf{x}_1)(\mathbf{x}_2 - \mathbf{x}_1),$$

és

$$f(\mathbf{x}_2) - f(\mathbf{x}_1) \geq \nabla f(\mathbf{x}_1)(\mathbf{x}_2 - \mathbf{x}_1).$$

Így a *4.6. Tétel* alapján kapjuk, hogy az f függvény konvex az A halmazon.

■

4.8. Tétel. *Legyen $A \subseteq R$ nyílt konvex halmaz és $f : A \rightarrow R$ kétszer folytonosan differenciálható függvény. Az f függvény akkor és csak akkor konvex az A halmazon, ha az f függvény Hf Hesse mátrixa pozitív szemidefinit az A halmaz minden pontjában.*

Bizonyítás.

I. Ha az f függvény az A halmazon konvex, akkor a *4.6. Tétel* miatt a (4.7) egyenlőtlenségek teljesülnek. Ha az $\mathbf{x}_1, \mathbf{x}_2 \in A$ és az $\mathbf{x}_1 + \lambda(\mathbf{x}_2 - \mathbf{x}_1)$, $0 < \lambda < 1$, pontokat tekintjük, akkor azt kapjuk, hogy

$$\frac{f\big(\mathbf{x}_1 + \lambda(\mathbf{x}_2 - \mathbf{x}_1)\big) - f(\mathbf{x}_1) - \lambda \nabla f(\mathbf{x}_1)(\mathbf{x}_2 - \mathbf{x}_1)}{\lambda^2} \geq 0,$$

amiből a *Taylor tétel* alapján és $\lambda \rightarrow 0$ határátmenettel adódik az állítás.

II. A *Taylor tétel* miatt $\mathbf{x}_1, \mathbf{x}_2 \in A$ esetén

$$f(\mathbf{x}_2) = f(\mathbf{x}_1) + \nabla f(\mathbf{x}_1)(\mathbf{x}_2 - \mathbf{x}_1) +$$
$$\frac{1}{2}(\mathbf{x}_2 - \mathbf{x}_1)^T Hf\big(\theta \mathbf{x}_1 + (1-\theta)\mathbf{x}_2\big)(\mathbf{x}_2 - \mathbf{x}_1)$$

valamely $0 \leq \theta \leq 1$ értékre. Mivel a feltételezés szerint az f függvény Hf Hesse mátrixa pozitív szemidefinit az A halmaz tetszőleges pontjában, ezért az

$$f(\mathbf{x}_2) - f(\mathbf{x}_1) \geq \nabla f(\mathbf{x}_1)(\mathbf{x}_2 - \mathbf{x}_1)$$

egyenlőtlenség teljesül minden $\mathbf{x}_1, \mathbf{x}_2 \in A$ esetén, és a *4.6. Tételt* alkalmazva kapjuk, hogy az f függvény konvex az A halmazon. ■

Legyen $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} + \mathbf{b}^T\mathbf{x} + \lambda$, ahol $\lambda \in R$, $\mathbf{x}, \mathbf{b} \in R^n$ és $\mathbf{Q}$ szimmetrikus $(n \times n)$-es mátrix. A *4.8 Tétel*ből következik, hogy az f kvadratikus függvény akkor és csak akkor konvex az egész téren, ha a $\mathbf{Q}$ mátrix pozitív szemidefinit. Az Euklideszi tér egyik legfontosabb függvénye, az euklideszi norma, aminek képlete

$$\parallel \mathbf{x} \parallel = \left(\sum_{i=1}^{n} x_i^2 \right)^{1/2} = \left(\mathbf{x}^T \mathbf{I} \mathbf{x} \right)^{1/2}, \qquad \mathbf{x} \in R^n,$$

ahol $\mathbf{I}$ az $(n \times n)$-es egységmátrix, az egész téren konvex függvény. (Az állítás bizonyítása a norma tulajdonságaiból azonnal adódik.)

A konvexitási tulajdonságot számos függvényművelet is megőrzi.

4.9. Tétel. Ha $A \subseteq R^n$ *konvex halmaz,* $f : A \to R$ *konvex függvény és* $\phi : R \to R$ *monoton növekvő konvex függvény, akkor a* $\phi f : A \to R$ *összetett függvény konvex.*

Ha $A \subseteq R^n$ *konvex halmaz,* $f_i : A \to R$, $i = 1, \ldots, m$, *konvex függvények és* $\lambda_i \geq 0$, *bármely* i*-re, akkor a* $\sum_{i=1}^m \lambda_i f_i$ *függvény konvex.*

Konvex függvények tetszőleges összességének pontonként vett szuprémuma konvex függvény.

Ha $\mathbf{f} : A \to R^m$ *konvex vektor függvény, azaz* $\mathbf{f}^T = (f_1, \ldots, f_n)$ *minden komponense konvex függvény, és* $\mathbf{B} : R^n \to R^m$ *lineáris leképezés, akkor az* $\mathbf{f}(\mathbf{B}\mathbf{x})$, $\mathbf{x} \in R^n$, *vektor függvény konvex.*

Bizonyítás. Az állítás a definíciókból közvetlenül adódik. ■

4.2. Definíció. *A konvex optimalizálási feladat (KO) a következő:*

$$\begin{gathered} \min \ f(\mathbf{x}) \\ g_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, m, \quad \mathbf{x} \in R^n, \end{gathered} \tag{4.9}$$

ahol az f *és* g_i, $i = 1, \ldots, m$, *függvények konvexek az egész téren vagy annak egy nyílt konvex részhalmazán.*

4.10. Tétel. *A (4.9) KO minden lokális minimuma globális minimum.*

Bizonyítás. A tétel bizonyítása indirekt úton történik. Tegyük fel, hogy a (KO) $\mathbf{x}^*$ pontbeli lokális minimuma nem globális. Ebből következik, hogy van olyan $\mathbf{x}^{**}$ megengedett megoldás, hogy $f(\mathbf{x}^{**}) < f(\mathbf{x}^*)$.

Tekintsük az $\mathbf{x}^*$ és az $\mathbf{x}^{**}$ pontokat összekötő $\mathbf{x}^*+\lambda(\mathbf{x}^{**}-\mathbf{x}^*)$, $0 \le \lambda \le 1$, egyenes szakaszt és az $\mathbf{x}^*$ pont olyan $U(\mathbf{x}^*)$ konvex környezetét, amelyben

$$f(\mathbf{x}^*) \le f(\mathbf{x}), \quad \mathbf{x} \in U(\mathbf{x}^*). \tag{4.10}$$

Mivel a g_i, $i = 1, \ldots, m$, függvények konvexitása miatt a (KO) megengedett tartománya konvex halmaz, ezért az $\mathbf{x}^*$ és az $\mathbf{x}^{**}$ pontokat összekötő egyenes szakasz minden pontja megengedett.

Legyen az $\bar{\mathbf{x}} = \mathbf{x}^* + \overline{\lambda}(\mathbf{x}^{**} - \mathbf{x}^*) \in U(\mathbf{x}^*)$, $\bar{\mathbf{x}} \ne \mathbf{x}^*$, pont tetszőlegesen választva az adott konvex környezetben. Az f függvény konvexitása, $0 < \overline{\lambda} < 1$ és $f(\mathbf{x}^{**}) < f(\mathbf{x}^*)$ miatt

$$f\big(\mathbf{x}^* + \overline{\lambda}(\mathbf{x}^{**} - \mathbf{x}^*)\big) \le (1-\overline{\lambda})f(\mathbf{x}^*) + \overline{\lambda}f(\mathbf{x}^{**}) < f(\mathbf{x}^*),$$

ami (4.10) miatt ellentmondás. Ezzel beláttuk az állítást. ■

A (4.9) KO esetén a (3.9) elsőrendű szükséges optimalitási feltétel egyben elegendő feltétel is, azaz az elsőrendű optimalitási feltétel teljesülése biztosítja a globális minimalitást.

4.11. Tétel. *Tegyük fel, hogy a (4.9) KO-ban az f és g_i, $i = 1, \ldots, m$, függvények differenciálhatóak és az $\mathbf{x}^*$ megengedett ponthoz találhatók olyan $\boldsymbol{\lambda} \in R^m$ Lagrange szorzók, amikre a*

$$\nabla_{\mathbf{x}} L(\mathbf{x}^*, \boldsymbol{\lambda}) = \nabla f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i \nabla g_i(\mathbf{x}^*) = 0, \tag{4.11}$$

$$\boldsymbol{\lambda} \ge \mathbf{0}, \tag{4.12}$$

$$\lambda_i g_i(\mathbf{x}^*) = 0, \qquad i = 1, \ldots, m, \tag{4.13}$$

feltételek teljesülnek. Akkor az $\mathbf{x}^*$ pont a KO globális minimuma.

Bizonyítás. A *4.9. Tétel* miatt

$$L(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}), \qquad \mathbf{x} \in R^n, \qquad \boldsymbol{\lambda} \geq \mathbf{0},$$

függvény konvex az $\mathbf{x}$ változóban. A *4.6. Tétel* miatt bármely két, $\mathbf{x}$ és $\mathbf{x}^*$ megengedett pontra teljesül az, hogy

$$f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) - f(\mathbf{x}^*) - \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}^*) \geq \nabla_{\mathbf{x}} L(\mathbf{x}^*, \boldsymbol{\lambda})\,(\mathbf{x} - \mathbf{x}^*). \tag{4.14}$$

A (4.14) egyenlőtlenségből a (4.11) és (4.13) egyenlőségek felhasználásával azt kapjuk, hogy

$$f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) \geq f(\mathbf{x}^*),$$

ami (4.12) miatt éppen az állítás. ∎

4.1. Példa. *Egy cég* 10 000 *dollárt akar reklámra költeni. Percenként 3000 dollárba kerül a hirdetés a televízióban, míg percenként* 1000 *dollárba a rádióban. Ha a cég x perc televíziós és y perc rádiós reklámidőt vesz, akkor $f(x, y) = -2x^2 - y^2 + xy + 8x + 3y$ bevétele lesz (ezer dollárban). Hogyan tudja a cég maximalizálni a bevételét?*

Megoldás. *A következő NLO feladatot akarjuk megoldani (a felírásban eltekintünk a nemnegativitási feltételektől):*

$$\max f(x, y) = -2x^2 - y^2 + xy + 8x + 3y$$

$$3x + y = 10, \qquad (x, y) \in R^2.$$

Ekkor $L(x, y, \mu) = -2x^2 - y^2 + xy + 8x + 3y + \mu(10 - 3x - y)$, $(x, y, \mu) \in R^3$. Felírjuk a

$$\frac{\partial L}{\partial x} = \frac{\partial L}{\partial y} = \frac{\partial L}{\partial \mu} = 0, \qquad (x, y, \mu) \in R^3,$$

egyenleteket. Azt kapjuk, hogy

$$\frac{\partial L}{\partial x} = -4x + y + 8 - 3\mu = 0, \tag{4.15}$$

$$\frac{\partial L}{\partial y} = -2y + x + 3 - \mu = 0, \tag{4.16}$$

$$\frac{\partial L}{\partial \mu} = 10 - 3x - y = 0, \qquad (x, y, \mu) \in R^3. \tag{4.17}$$

Vegyük észre, hogy $10 - 3x - y = 0$ *átírható* $3x + y = 10$ *alakra. A* (4.15) *egyenletből* $y = 3\mu - 8 + 4x$, (4.16)*-ból pedig* $x = \mu - 3 + 2y$ *adódik. Tehát* $y = 3\mu - 8 + 4(\mu - 3 + 2y) = 7\mu - 20 + 8y$, *azaz*

$$y = \frac{20}{7} - \mu, \tag{4.18}$$

$$x = \mu - 3 + 2\left(\frac{20}{7} - \mu\right) = \frac{19}{7} - \mu. \tag{4.19}$$

A (4.18) *és a* (4.19) *összefüggéseket* (4.17)*-be behelyettesítve kapjuk, hogy* $10 - 3\left(\frac{19}{7} - \mu\right) - \left(\frac{20}{7} - \mu\right) = 0$, *azaz* $4\mu - 1 = 0$, *azaz* $\mu = \frac{1}{4}$. *Innen* (4.18) *és* (4.19) *alapján*

$$y = \frac{20}{7} - \frac{1}{4} = \frac{73}{28},$$
$$x = \frac{19}{7} - \frac{1}{4} = \frac{69}{28}.$$

Az $f(x, y)$, $(x, y) \in R^2$, *függvény Hesse-mátrixa*

$$H(x, y) = \begin{bmatrix} -4 & 1 \\ 1 & -2 \end{bmatrix}, \qquad (x, y) \in R^2.$$

Mivel mindegyik elsőrendű főminor negatív és a Hesse mátrix determinánsa 7 *nagyobb mint* 0, *az* f *függvény konkáv. A feltétel lineáris, ezért a Lagrange-szorzók módszere az NLO feladat optimális megoldásához vezet.*

Tehát a cégnek $\frac{69}{28}$ *perc televíziós és* $\frac{73}{28}$ *perc rádiós reklámidőt kell vásárolnia. Mivel* $\mu = \frac{1}{4}$, *további* Δ *ezer dollár reklámra költése (kis* Δ *esetén) a cég bevételét közelítőleg 0,25*Δ *ezer dollárral növelné.*

Általában, ha a cég δ *ezer dollárt költene reklámra, megmutatható, hogy*

$\mu = \dfrac{11-\delta}{4}$. *Látható, hogy minél többet költenek reklámra, az újabb δ ezer dollárnyi reklámköltség hatása a bevétel növekedésére egyre kisebb.*

4.2. Példa. *Mutassuk meg, hogy adott $x_1, x_2, \dots, x_n$ számok esetén*

$$n\sum_{i=1}^{n} x_i^2 \geq \left(\sum_{i=1}^{n} x_i\right)^2$$

és az egyenlőség pontosan akkor teljesül, ha $x_1 = x_2 = \cdots = x_n$!

Megoldás. *Tegyük fel, hogy $x_1 + x_2 + \cdots + x_n = c$. Tekintsük a*

$$\begin{aligned} \min\ f(\mathbf{x}) = \sum_{i=1}^{n} x_i^2 & \\ \sum_{i=1}^{n} x_i = c, \qquad \mathbf{x} \in R^n, & \end{aligned} \tag{4.20}$$

NLO feladatot. A (4.20) *feladat megoldásához képezzük az*

$$L(x_1, x_2, \dots, x_n, \mu) = x_1^2 + x_2^2 + \cdots + x_n^2 + \mu(c - x_1 - x_2 - \cdots - x_n), \quad (\mathbf{x}, \mu) \in R^{n+1},$$

Lagrange-függvényt. Ekkor (4.20) *megoldásához olyan $(x_1, x_2, \dots, x_n, \mu) \in R^{n+1}$ vektort kell találni, amelyre*

$$\begin{aligned} \frac{\partial L}{\partial x_i} &= 2x_i - \mu = 0, \qquad i = 1, 2, \dots, n, \\ \frac{\partial L}{\partial \mu} &= c - x_1 - x_2 - \cdots - x_n = 0, \qquad (\mathbf{x}, \mu) \in R^{n+1}. \end{aligned}$$

A $\dfrac{\partial L}{\partial x_i} = 0$, $i = 1, \dots, n$, *összefüggésekből kapjuk, hogy* $2\overline{x}_1 = 2\overline{x}_2 = \cdots = 2\overline{x}_n = \overline{\mu}$, *azaz* $x_i = \dfrac{\overline{\mu}}{2}$. *A* $\dfrac{\partial L}{\partial \mu} = 0$ *összefüggésből* $c - \dfrac{n\overline{\mu}}{2} = 0$, *azaz* $\overline{\mu} = \dfrac{2c}{n}$ *adódik. A célfüggvény konvex (n konvex függvény összege) a feltétel pedig lineáris, tehát a Lagrange-szorzók módszerével előáll* (4.20) *optimális megoldása:*

$$\overline{x}_i = \frac{\left(\dfrac{2c}{n}\right)}{2} = \frac{c}{n} \qquad \text{és} \qquad f(\overline{\mathbf{x}}) = n\left(\frac{c^2}{n^2}\right) = \frac{c^2}{n}.$$

Ezért, ha

$$\sum_{i=1}^{n} x_i = c,$$

akkor

$$n\sum_{i=1}^{n} x_i^2 \geq n\left(\frac{c^2}{n}\right) = \left(\sum_{i=1}^{n} x_i\right)^2,$$

és egyenlőség pontosan akkor teljesül, ha $x_1 = x_2 = \cdots = x_n$.

Ha olyan $f(\mathbf{x})$, $\mathbf{x} \in R^n$, függvényt próbálunk meg maximalizálni, amely több függvény szorzataként áll elő, akkor gyakran könnyebb a $ln\, f$ függvényt maximalizálni helyette. Mivel a logaritmus függvény növekvő, ezért bármely olyan $\mathbf{x}^* \in R^n$, amely maximalizálja az $ln\, f(x_1, x_2, \ldots, x_n)$ függvényt az $(x_1, x_2, \ldots, x_n)$ értékek valamilyen lehetséges tartományán, maximalizálni fogja az $f(x_1, x_2, \ldots, x_n)$ függvényt is az $(x_1, x_2, \ldots, x_n)$ értékek ugyanezen lehetséges tartományán. A fentiekben természetesen feltesszük, hogy azok a függvények, amelyekre a *ln* függvényt alkalmazni akarjuk, csak pozitív értéket vesznek fel a megengedett tartományban.

5. fejezet

ÁLTALÁNOSÍTOTT KONVEX FÜGGVÉNYEK

Ebben a részben a konvex függvények általánosításai közül a kvázikonvex és a pszeudokonvex függvényeket tárgyaljuk. A kvázikonvex függvényeknek számos ekvivalens definicíója van. Itt az alsó szinthalmazok konvexitását emeljük ki. Megjegyezzük, hogy a kvázikonvex és a pszeudokonvex függvények értelmezési tartománya mindig konvex halmaz.

5.1. Definíció. *Legyen az f függvény az $A \subseteq R^n$ konvex halmazon értelmezve. Azt mondjuk, hogy az f kvázikonvex, ha az*

$$L(f, c) = \{\mathbf{x} \in A \mid f(\mathbf{x}) \leq c\} \tag{5.1}$$

alsó nívóhalmazok konvexek minden valós c értékre.

Minden konvex függvény kvázikonvex, de fordítva nem mindig igaz. A differenciálható kvázikonvex függvényeket Arrow és Enthoven (1961), valamint Crouzeix (1980) jellemezte.

5.1. Tétel. *(Arrow és Enthoven, 1961) Legyen az f függvény differenciálható a nyílt és konvex $A \subseteq R^n$ halmazon. Az f függvény akkor és csak akkor kvázikonvex, ha bármely $\mathbf{x}_1, \mathbf{x}_2 \in A$ esetén*

$$f(\mathbf{x}_1) \leq f(\mathbf{x}_2) \Longrightarrow \nabla f(\mathbf{x}_2)(\mathbf{x}_1 - \mathbf{x}_2) \leq 0, \tag{5.2}$$

ahol a ∇f sorvektor az f függvény gradiense.

Bizonyítás.

I. Ha az f függvény kvázikonvex, akkor

$$f(\mathbf{x}_1) \le f(\mathbf{x}_2) \Longrightarrow f\big(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2\big) \le f(\mathbf{x}_2)$$

bármely $0 < \lambda < 1$ értékre. Ebből következik, hogy

$$\frac{f\big(\mathbf{x}_2 + \lambda(\mathbf{x}_1 - \mathbf{x}_2)\big) - f(\mathbf{x}_2)}{\lambda} \le 0,$$

és pozitív λ értékekkel a 0-hoz tartva, azt kapjuk, hogy

$$\nabla f(\mathbf{x}_2)(\mathbf{x}_1 - \mathbf{x}_2) \le 0.$$

II. Ha $\mathbf{x}_1, \mathbf{x}_2 \in A$ két tetszőleges pont, akkor az általánosság megszorítása nélkül feltehetjük, hogy $f(\mathbf{x}_1) \le f(\mathbf{x}_2)$.

Az állítás bizonyításához elegendő azt belátni, hogy az alábbi

$$F(\lambda) = f\big(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2\big), \quad 0 \le \lambda \le 1,$$

egyváltozós függvény kvázikonvex, azaz $F(\lambda) \le F(0)$ bármely $0 \le \lambda \le 1$ értékre.

Tegyük fel, hogy valamely $0 \le \lambda \le 1$ értékre $F(\lambda) > F(0)$. Mivel $F(1) \le F(0)$, ezért van olyan $\lambda_0 > 0$ érték, amelyre $F(0) < F(\lambda_0)$ és $F'(\lambda_0) < 0$. Így,

$$F'(\lambda_0) = \nabla f(\mathbf{x}_0)(\mathbf{x}_1 - \mathbf{x}_2) < 0, \tag{5.3}$$

ahol $\mathbf{x}_0 = \lambda_0\mathbf{x}_1 + (1-\lambda_0)\mathbf{x}_2$. Mivel $f(\mathbf{x}_2) < f(\mathbf{x}_0)$, ezért az (5.2) tulajdonságból következik, hogy

$$\nabla f(\mathbf{x}_0)(\mathbf{x}_2 - \mathbf{x}_0) \le 0.$$

Átrendezés után és λ_0-al való osztás után azt kapjuk, hogy

$$\nabla f(\mathbf{x}_0)(\mathbf{x}_1 - \mathbf{x}_2) \ge 0,$$

ami ellentmond az (5.3) egyenlőtlenségnek, és így bizonyítottuk az állítást. ■

5.2. Tétel. (*Arrow és Enthoven, 1961) Ha az $f \in C^2$ függvény kvázikonvex az $A \subseteq R^n$ nyílt és konvex halmazon, akkor*

$$\mathbf{x} \in A,\ \mathbf{v} \in R^n,\ \nabla f(\mathbf{x})\mathbf{v} = 0 \quad \Longrightarrow \quad \mathbf{v}^T Hf(\mathbf{x})\mathbf{v} \geq 0, \tag{5.4}$$

ahol Hf az f függvény Hesse mátrixát jelenti.

Bizonyítás. A bizonyítás indirekt. Tegyük fel, hogy az f függvény kvázikonvex, $\nabla f(\mathbf{x})\mathbf{v} = 0$ és az állítással ellentétben $\mathbf{v}^T Hf(\mathbf{x})\mathbf{v} < 0$ valamilyen $\mathbf{x} \in A$ és $\mathbf{v} \in R^n$ vektorok esetén. A Hesse mátrix folytonossága miatt létezik olyan $\varepsilon > 0$ érték, amelyre

$$\|\tilde{\mathbf{x}} - \mathbf{x}\| < \varepsilon \Longrightarrow \mathbf{v}^T Hf\big(\lambda\tilde{\mathbf{x}} + (1-\lambda)\mathbf{x}\big)\mathbf{v} < 0, \quad \tilde{\mathbf{x}} \in A, \tag{5.5}$$

bármely $0 \leq \lambda \leq 1$ esetén. Ha $\tilde{\mathbf{x}} = \mathbf{x} + \mu\mathbf{v},\ 0 < \mu \parallel \mathbf{v} \parallel < \varepsilon$, választással, akkor (5.5) a következő formába írható:

$$(\tilde{\mathbf{x}} - \mathbf{x})^T Hf\big(\lambda\tilde{\mathbf{x}} + (1-\lambda)\mathbf{x}\big)(\tilde{\mathbf{x}} - \mathbf{x}) < 0. \tag{5.6}$$

A *Taylor tétel* alkalmazásával azt kapjuk, hogy

$$f(\tilde{\mathbf{x}}) = f(\mathbf{x}) + \nabla f(\mathbf{x})(\tilde{\mathbf{x}} - \mathbf{x}) + \frac{1}{2}(\tilde{\mathbf{x}} - \mathbf{x})^T Hf\big(\tilde{\lambda}\tilde{\mathbf{x}} + (1-\tilde{\lambda})\mathbf{x}\big)(\tilde{\mathbf{x}} - \mathbf{x})$$

teljesül valamely $0 < \tilde{\lambda} < 1$ értékre, és (5.6) miatt pedig az

$$f(\tilde{\mathbf{x}}) < f(\mathbf{x}) \tag{5.7}$$

egyenlőtlenség teljesül.

Ha $\hat{\mathbf{x}} = \mathbf{x}+\mu(-\mathbf{v})$, akkor hasonló meggondolással kapjuk, hogy

$$f(\hat{\mathbf{x}}) < f(\mathbf{x}). \tag{5.8}$$

Mivel $\mathbf{x} = \dfrac{1}{2}(\tilde{\mathbf{x}} + \hat{\mathbf{x}})$, az (5.7) és (5.8) egyenlőtlenségekből az következik, hogy az f

függvény nem kvázikonvex, ami ellentmondás. Ezzel bebizonyítottuk az állítást. ■

5.3. Tétel. *(Crouzeix, 1980) Ha a nyílt és konvex $A \subseteq R^n$ halmazon értelmezett $f \in C^2$ függvény teljesíti az*

$$\mathbf{x} \in A,\ \nabla f(\mathbf{x}) = 0 \quad \Longrightarrow \quad \mathbf{v}^T Hf(\mathbf{x})\mathbf{v} > 0,\ \forall \mathbf{v} \neq 0, \tag{5.9}$$

$$\mathbf{x} \in A,\ \mathbf{v} \in R^n,\ \nabla f(\mathbf{x})\mathbf{v} = 0 \quad \Longrightarrow \quad \mathbf{v}^T Hf(\mathbf{x})\mathbf{v} \geq 0, \tag{5.10}$$

feltételeket, akkor az f függvény kvázikonvex.

5.1. Következmény. *Legyen az $f \in C^2$ függvény értelmezve az $A \subseteq R^n$ nyílt és konvex halmazon, és tegyük fel, hogy $\nabla f(\mathbf{x}) \neq 0$, $\forall\ \mathbf{x} \in A$. Az f függvény akkor és csak akkor kvázikonvex, ha az (5.10) tulajdonság teljesül.*

Crouzeix (1980) egy megjegyzése szerint az (5.9) feltételt helyettesítve az

$$\mathbf{x} \in A,\ \nabla f(\mathbf{x}) = 0 \quad \Longrightarrow \quad f\text{-nek } \mathbf{x}\text{-ben globális minimuma van;} \tag{5.11}$$

feltétellel, az *5.3. Tétel* érvényben marad.

Mangasarian (1965) vezette be a differenciálható pszeudokonvex függvényekre vonatkozó, máig legszéleskörűbben alkalmazott definíciót.

5.2. Definíció. *Az $A \subseteq R^n$ nyílt és konvex halmazon értelmezett differenciálható f függvényt pszeudokonvexnek nevezzük, ha bármely $\mathbf{x}_1, \mathbf{x}_2 \in A$ esetén*

$$f(\mathbf{x}_1) < f(\mathbf{x}_2) \quad \Longrightarrow \quad \nabla f(\mathbf{x}_2)(\mathbf{x}_1 - \mathbf{x}_2) < 0, \tag{5.12}$$

vagy más formában

$$\nabla f(\mathbf{x}_2)(\mathbf{x}_1 - \mathbf{x}_2) \geq 0 \quad \Longrightarrow \quad f(\mathbf{x}_1) \geq f(\mathbf{x}_2).$$

A definíciókból következik, hogy minden differenciálható konvex függvény pszeudokonvex, és minden pszeudokonvex függvény kvázikonvex. A következő állítás, ami az előzők következménye, a pszeudokonvexitás jellemzését adja.

5.4. Tétel. *Legyen az $f \in C^2$ függvény az $A \subseteq R^n$ nyílt és konvex halmazon értelmezve. Az f függvény akkor és csak akkor pszeudokonvex, ha*

$$\mathbf{x} \in A,\ \nabla f(\mathbf{x}) = 0 \quad \Longrightarrow \quad f\textit{-nek az } \mathbf{x}\textit{-ben globális minimuma van}, \tag{5.13}$$

$$\mathbf{x} \in A,\ \mathbf{v} \in R^n,\ \nabla f(\mathbf{x})\mathbf{v} = 0 \quad \Longrightarrow \quad \mathbf{v}^T Hf(\mathbf{x})\mathbf{v} \geq 0. \tag{5.14}$$

Bizonyítás.

I. Ha az f függvény pszeudokonvex, akkor kvázikonvex is, és az *5.2. Tétel* szerint (5.14) teljesül. A pszeudokonvexitás definíciójából következik, hogy $\nabla f(\mathbf{x}_2) = \mathbf{0}$, $\mathbf{x}_2 \in A$, esetén $f(\mathbf{x}_2) \leq f(\mathbf{x})$, $\mathbf{x} \in A$, ami azt jelenti, hogy az (5.13) tulajdonság is teljesül.

II. Ha az (5.13) és (5.14) tulajdonságok teljesülnek, akkor az *5.3. Tétel* és (5.11) szerint az f függvény kvázikonvex. Az *5.1. Tétel* szerint, bármely $\mathbf{x}_1, \mathbf{x}_2 \in A$ esetén

$$f(\mathbf{x}_1) \leq f(\mathbf{x}_2) \quad \Longrightarrow \quad \nabla f(\mathbf{x}_2)(\mathbf{x}_1 - \mathbf{x}_2) \leq 0. \tag{5.15}$$

Tegyük fel, hogy az f függvény nem pszeudokonvex. Akkor léteznek olyan $\mathbf{x}_1^*, \mathbf{x}_2^* \in A$ vektorok, amelyekre az teljesül, hogy

$$f(\mathbf{x}_1^*) < f(\mathbf{x}_2^*) \quad \text{és} \quad \nabla f(\mathbf{x}_2^*)(\mathbf{x}_1^* - \mathbf{x}_2^*) \geq 0.$$

Az (5.15) tulajdonságból azt kapjuk, hogy

$$\nabla f(\mathbf{x}_2^*)(\mathbf{x}_1^* - \mathbf{x}_2^*) = 0. \tag{5.16}$$

Ha $\nabla f(\mathbf{x}_2^*) = 0$, akkor az (5.13)-ból következik, hogy az f függvénynek globális minimuma van az $\mathbf{x}_2^*$ pontban, ami ellentmond annak, hogy $f(\mathbf{x}_1^*) < f(\mathbf{x}_2^*)$.

Ha $\nabla f(\mathbf{x}_2^*) \neq 0$, akkor az $\{\mathbf{x} \in A | f(\mathbf{x}) = f(\mathbf{x}_2^*)\}$ hiperfelület $\mathbf{x}_2^*$ pontbeli érintőtere tartalmazza az $\mathbf{x}_1^*$ pontot az (5.16) képlet szerint. Másrészt, $f(\mathbf{x}_1^*) < f(\mathbf{x}_2^*)$, ezért az $\mathbf{x}_1^*$ pontnak létezik olyan $U(\mathbf{x}_1^*)$ környezete, hogy

$$f(\mathbf{x}) < f(\mathbf{x}_2^*), \qquad \mathbf{x} \in U(\mathbf{x}_1^*),$$

azaz az érintőtér mindkét oldalán vannak az $U(\mathbf{x}_1^*)$ halmaznak pontjai, ami ellentmond az f függvény kvázikonvexitásának. ∎

6. fejezet

LAGRANGE DUALITÁS ÉS NYEREGPONT

Ebben a részben bevezetjük a Lagrange duál problémát, majd bebizonyítjuk a gyenge és az erős dualitási tételt.

Tekintsük a következő nemlineáris optimalizálási feladatot:

$$\begin{gathered}\min \ f(\mathbf{x}) \\ h_j(\mathbf{x}) = 0, \qquad j = 1, \dots, p, \\ g_i(\mathbf{x}) \geq 0, \qquad i = 1, \dots, m, \\ \mathbf{x} \in A \subseteq R^n.\end{gathered} \tag{6.1}$$

A (6.1) primál feladat Lagrange duál feladata a következő:

$$\begin{gathered}\max_{(\boldsymbol{\mu},\boldsymbol{\lambda})} \left\{ \inf_{\mathbf{x}\in A} L(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\lambda}) \right\} \\ \boldsymbol{\mu} \in R^p, \quad \boldsymbol{\lambda} \geq \mathbf{0}, \quad \boldsymbol{\lambda} \in R^m,\end{gathered} \tag{6.2}$$

ahol $L(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\lambda})$, a (6.1) feladat Lagrange függvénye, a (3.2) képletben van megadva.

Adott nemlineáris optimalizálási feladat esetén, a feltételek különböző formában történő megadásától függően különböző duál feladatokat nyerünk, amik a feladat megoldásának hatékonyságára lehetnek hatással. Továbbá, a duál feladatban a feltételek figyelembevételére lehetőség van a Lagrange függvényben vagy az A halmaz meghatározásakor.

6.1. Példa.

Tekintsük a következő primál feladatot:

$$\min\ x_1^2 + x_2^2$$
$$g(\mathbf{x}) = x_1 + x_2 - 4 \geq 0,$$
$$x_1 \geq 0,\ x_2 \geq 0.$$

Az optimális megoldás $(x_1^*, x_2^*) = (2, 2)$, *ahol a célfüggvény értéke 8.*

A duál feladat a következő:

$$\max_{\lambda}\{\inf_{(x_1,x_2)} L(\mathbf{x},\lambda)\} = \max_{\lambda} \inf_{(x_1,x_2)} \{x_1^2 + x_2^2 - \lambda(x_1 + x_2 - 4) |\, x_1 \geq 0,\ x_2 \geq 0\}$$
$$\lambda \geq 0.$$

Mivel

$$\inf_{(x_1,x_2)} L(\mathbf{x},\lambda) = \inf_{x_1} \{x_1^2 - \lambda x_1 |\, x_1 \geq 0\} + \inf_{x_2}\{x_2^2 - \lambda x_2 \,|\, x_2 \geq 0\} + 4\lambda,$$

e függvény infimumának helye

$$x_1 = x_2 = \frac{\lambda}{2}, \qquad \textit{ha} \qquad \lambda \geq 0,$$
$$x_1 = x_2 = 0, \qquad \textit{ha} \qquad \lambda < 0.$$

Ebből következik, hogy a duál feladat explicit alakja

$$\max\ -\frac{1}{2}\lambda^2 + 4\lambda$$
$$\lambda \geq 0.$$

Mivel a célfüggvény konkáv, a $\lambda^* = 4$ pontban a duál feladatnak globális maximuma van, továbbá, a duál feladat optimumértéke 8, ami megegyezik a primál feladat optimumértékével.

6.1. Gyenge dualitási tétel.

Ha $\widehat{\mathbf{x}}$ megengedett megoldása a (6.1) primál feladatnak és $(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\lambda}})$ megengedett megoldása a (6.2) duál feladatnak, akkor

$$f(\widehat{\mathbf{x}}) \geq \inf_{\mathbf{x} \in A} L(\mathbf{x}, \widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\lambda}}). \tag{6.3}$$

Bizonyítás. A Lagrange függvény definíciójából és az $\widehat{\mathbf{x}}$, illetve $(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\lambda}})$ megengedettségéből következik, hogy

$$\inf_{\mathbf{x} \in A} L(\mathbf{x}, \widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\lambda}}) = \inf_{\mathbf{x} \in A} \left\{ f(\mathbf{x}) + \sum_{j=1}^{p} \widehat{\mu}_j h_j(\mathbf{x}) - \sum_{i=1}^{m} \widehat{\lambda}_i g_i(\mathbf{x}) \right\} \leq$$
$$f(\widehat{\mathbf{x}}) + \sum_{j=1}^{p} \widehat{\mu}_j h_j(\widehat{\mathbf{x}}) - \sum_{i=1}^{m} \widehat{\lambda}_i g_i(\widehat{\mathbf{x}}) \leq f(\widehat{\mathbf{x}}),$$

ami éppen az állítás. ■

6.1. Következmény. *A primál probléma infimuma nagyobb vagy egyenlő, mint a duál probléma szuprémuma, azaz*

$$\inf_{\mathbf{x}} \{ f(\mathbf{x}), \mathbf{x} \in M[\mathbf{h}, \mathbf{g}] \cap A \} \geq \sup_{(\boldsymbol{\mu}, \boldsymbol{\lambda})} \left\{ \inf_{\mathbf{x} \in A} \{ L(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\lambda}) | \; \boldsymbol{\mu} \in R^p, \quad \boldsymbol{\lambda} \geq 0 \right\}. \tag{6.4}$$

6.2. Következmény. *Ha $\widehat{\mathbf{x}}$ megengedett megoldása a (6.1) primál feladatnak, $(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\lambda}})$ megengedett megoldása a (6.2) duál feladatnak és*

$$f(\widehat{\mathbf{x}}) \leq \inf_{\mathbf{x} \in A} L(\mathbf{x}, \widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\lambda}}),$$

akkor $\widehat{\mathbf{x}}$ optimális megoldása a (6.1) primál feladatnak, a $(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\lambda}})$ pedig a (6.2) duál feladatnak.

6.3. Következmény. Ha

$$\inf_{\mathbf{x}} \{ f(\mathbf{x}), \mathbf{x} \in M[\mathbf{h}, \mathbf{g}] \cap A \} = -\infty, \text{ akkor } \inf_{\mathbf{x} \in A} L(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\lambda}) = -\infty, \quad \forall \boldsymbol{\mu} \in R^p, \; \boldsymbol{\lambda} \geq \mathbf{0}. \tag{6.5}$$

6.4. Következmény. *Ha*

$$\sup_{(\boldsymbol{\mu},\boldsymbol{\lambda})} \left\{ \inf_{\mathbf{x}\in A} L(\mathbf{x},\boldsymbol{\mu},\boldsymbol{\lambda}) \middle| \boldsymbol{\mu} \in R^p,\ \boldsymbol{\lambda} \geq \mathbf{0} \right\} = \infty, \tag{6.6}$$

akkor a primál problémának nincs megengedett megoldása.

A *6.1. Következmény* szerint a primál probléma infimuma nagyobb vagy egyenlő, mint a duál probléma szuprémuma. Ha a (6.4) egyenlőtlenség szigorú, akkor *dualitási résről* beszélünk.

Elválasztási tétel. *Ha $A \subseteq R^n$ nem üres, zárt konvex halmaz és $\mathbf{y} \notin A$, akkor létezik olyan $\mathbf{p} \in R^n$ nem nulla vektor és $\alpha \in R$ érték, amelyekre*

$$\mathbf{p}^T\mathbf{y} > \alpha \quad \text{és} \quad \mathbf{p}^T\mathbf{x} \leq \alpha, \qquad \mathbf{x} \in A.$$

Bizonyítás. Megmutatjuk, hogy mivel az $A \subseteq R^n$ nem üres, zárt konvex halmaz és $\mathbf{y} \notin A$, ezért létezik olyan egyedüli $\mathbf{x}^* \in A$ pont, amelyre

$$(\mathbf{x}-\mathbf{x}^*)^T(\mathbf{y}-\mathbf{x}^*) \leq 0, \qquad \mathbf{x} \in A,$$

és A pontjai közül ez az $\mathbf{x}^*$ pont van a legközelebb az $\mathbf{y}$ ponthoz az euklideszi metrikában mérve. A $\mathbf{p} = \mathbf{y} - \mathbf{x}^*$ és $\alpha = \mathbf{x}^{*T}(\mathbf{y}-\mathbf{x}^*)$ helyettesítésekkel pedig közvetlenül megkapjuk a kívánt egyenlőtlenségeket.

A tétel feltételeiből következik, hogy $\inf\{\|\mathbf{y}-\mathbf{x}\|\,|\mathbf{x} \in A\} = \gamma > 0$ és létezik olyan $\{\mathbf{x}_k\} \in A$ sorozat, amelyre $\|\mathbf{y}-\mathbf{x}_k\| \longrightarrow \gamma$. Megmutatjuk, hogy az $\{\mathbf{x}_k\}$ Cauchy sorozat, ezért van $\mathbf{x}^*$ határértéke. A *Paralelogramma tételt* (pl. Dancs és Puskás, 2001, 206. old.) alkalmazva azt kapjuk, hogy

$$\|\mathbf{x}_k - \mathbf{x}_m\|^2 = 2\|\mathbf{x}_k - \mathbf{y}\|^2 + 2\|\mathbf{x}_m - \mathbf{y}\|^2 - \|\mathbf{x}_k + \mathbf{x}_m - 2\mathbf{y}\|^2 =$$
$$2\|\mathbf{x}_k - \mathbf{y}\|^2 + 2\|\mathbf{x}_m - \mathbf{y}\|^2 - 4\|\frac{\mathbf{x}_k + \mathbf{x}_m}{2} - \mathbf{y}\|^2.$$

Mivel $(\mathbf{x}_k + \mathbf{x}_m)/2 \in A$, ezért a γ definíciójából adódik, hogy $\|\dfrac{\mathbf{x}_k + \mathbf{x}_m}{2} - \mathbf{y}\|^2 \geq \gamma^2$, így

$$\|\mathbf{x}_k - \mathbf{x}_m\|^2 \; \leq 2\|\mathbf{x}_k - \mathbf{y}\|^2 + 2\|\mathbf{x}_m - \mathbf{y}\|^2 - 4\,\gamma^2.$$

A k és m indexeket elég nagyra választva, az $\|\mathbf{x}_k - \mathbf{y}\|^2$ és az $\|\mathbf{x}_m - \mathbf{y}\|^2$ értékek tetszőlegesen közel lesznek a γ^2 értékhez, ezért $\|\mathbf{x}_k - \mathbf{x}_m\|^2$ is tetszőlegesen közel lesz a nullához. Ebből következik, hogy $\{\mathbf{x}_k\}$ Cauchy sorozat, aminek van $\mathbf{x}^*$ határértéke. Mivel az A halmaz zárt, ezért $\mathbf{x}^* \in A$. Azt kell még megmutatni, hogy csak egy ilyen tulajdonságú pont létezik. Tegyük fel, hogy létezik egy másik $\bar{\mathbf{x}}^*$ pont, amelyre $\|\mathbf{y} - \mathbf{x}^*\| = \|\mathbf{y} - \bar{\mathbf{x}}^*\| = \gamma$. Az A halmaz konvexitása miatt, $(\mathbf{x}^* + \bar{\mathbf{x}}^*)/2 \in A$. A háromszög egyenlőtlenséget alkalmazva

$$\|\mathbf{y} - \frac{\mathbf{x}^* + \bar{\mathbf{x}}^*}{2}\| \leq \frac{1}{2}\|\mathbf{y} - \mathbf{x}^*\| + \frac{1}{2}\|\mathbf{y} - \bar{\mathbf{x}}^*\| = \gamma.$$

Ha az egyenlőtlenség éles, akkor a γ definíciójával kerülünk ellentmondásba. Ezért egyenlőség teljesül és $\mathbf{y} - \mathbf{x}^* = \lambda(\mathbf{y} - \bar{\mathbf{x}}^*)$ valamely λ értékre. Mivel $\|\mathbf{y} - \mathbf{x}^*\| = \|\mathbf{y} - \bar{\mathbf{x}}^*\| = \gamma$, ezért $|\lambda| = 1$. Azonban $\lambda \neq -1$, mivel ellenkező esetben $\mathbf{y} = (\mathbf{x}^* + \bar{\mathbf{x}}^*)/2 \in A$, ami ellentmond az $\mathbf{y} \notin A$ feltételnek. Így $\lambda = 1$ és $\mathbf{x}^* = \bar{\mathbf{x}}^*$, ami az egyértelműséget bizonyítja.

A bizonyítás befejezéséhez meg kell mutatni, hogy az

$$(\mathbf{x} - \mathbf{x}^*)^T (\mathbf{x}^* - \mathbf{y}) \geq 0, \qquad \mathbf{x} \in A,$$

egyenlőtlenségek teljesülése szükséges és elegendő feltétele annak, hogy az $\mathbf{x}^* \in A$ pont legyen a legközelebb az $\mathbf{y}$ ponthoz.

Az elegendőség bizonyításához tekintsünk egy tetszőleges $\mathbf{x} \in A$ pontot. Akkor,

$$\|\mathbf{y} - \mathbf{x}\|^2 = \|\mathbf{y} - \mathbf{x}^* + \mathbf{x}^* - \mathbf{x}\|^2 = \|\mathbf{y} - \mathbf{x}^*\|^2 + \|\mathbf{x}^* - \mathbf{x}\|^2 + 2\,(\mathbf{x}^* - \mathbf{x})^T (\mathbf{y} - \mathbf{x}^*)\,.$$

Mivel a feltételezésünk szerint $(\mathbf{x}^* - \mathbf{x})^T (\mathbf{y} - \mathbf{x}^*) \geq 0$, és $\|\mathbf{x}^* - \mathbf{x}\|^2 \geq 0$, ezért azt kapjuk, hogy

$$\|\mathbf{y} - \mathbf{x}\|^2 \geq \|\mathbf{y} - \mathbf{x}^*\|^2, \qquad \mathbf{x} \in A,$$

ami azt jelenti, hogy az $\mathbf{x}^*$ pont van a legközelebb az A halmazban az $\mathbf{y}$ ponthoz.

A szükségesség bizonyításához tegyük fel, hogy

$$\|\mathbf{y}-\mathbf{x}\|^2 \geq \|\mathbf{y}-\mathbf{x}^*\|^2, \qquad \mathbf{x}\in A.$$

Legyen $\mathbf{x}\in A$ tetszőleges, akkor $\mathbf{x}^* + \lambda\left(\mathbf{x}-\mathbf{x}^*\right)\in A$ elegendően kicsiny $\lambda > 0$ értékre. Ezért,

$$\|\mathbf{y}-\mathbf{x}^*-\lambda\left(\mathbf{x}-\mathbf{x}^*\right)\|^2 \geq \|\mathbf{y}-\mathbf{x}^*\|^2,$$

és

$$\|\mathbf{y}-\mathbf{x}^*-\lambda\left(\mathbf{x}-\mathbf{x}^*\right)\|^2 = \|\mathbf{y}-\mathbf{x}^*\|^2 + \lambda^2\|\mathbf{x}-\mathbf{x}^*\|^2 + 2\lambda\left(\mathbf{x}-\mathbf{x}^*\right)^T\left(\mathbf{x}^*-\mathbf{y}\right).$$

A fenti relációkból következik, hogy

$$\lambda^2\|\mathbf{x}-\mathbf{x}^*\|^2 + 2\lambda\left(\mathbf{x}-\mathbf{x}^*\right)^T\left(\mathbf{x}^*-\mathbf{y}\right) \geq 0, \qquad \mathbf{x}\in A,$$

minden elegendően kicsiny $\lambda > 0$ értékre. Az egyenlőtlenség mindkét oldalát osztva a $\lambda > 0$ értékkel, majd a λ értékkel nullához tartva kapjuk az állítást. ■

Az elválasztási tételek a konvex analízis legfontosabb tételei közé tartoznak. Geometriailag a most bizonyított tétel azt jelenti, hogy ha adott egy pont tetszőleges konvex halmazon kívül, akkor létezik olyan a pontot tartalmazó hipersík, amelyik nem metszi a konvex halmazt. A következményként megfogalmazott *Legközelebbi pont tétel* és *Tompaszög-tétel* megtalálható, pl. a Dancs és Puskás (2001) jegyzet 273., illetve 275. oldalain.

6.5. Következmény (Legközelebbi pont tétel). *Ha $A\subseteq R^n$ zárt konvex halmaz, akkor tetszőleges $\mathbf{y}\in R^n$ ponthoz van az A halmaznak egyértelműen meghatározott legközelebbi pontja.*

6.6. Következmény (Tompaszög-tétel). *Ha $A \subseteq R^n$ zárt konvex halmaz és $\mathbf{x}^* \in A$ az A halmaznak egy $\mathbf{y} \notin A$ ponthoz legközelebb eső pontja, akkor*

$$(\mathbf{y}-\mathbf{x}^*)^T(\mathbf{x}-\mathbf{x}^*) \le 0, \quad \mathbf{x} \in A.$$

6.1. Lemma. *Legyen $A \subseteq R^n$ konvex halmaz, $\alpha : R^n \to R$, $-\mathbf{g} : R^n \to R^m$ konvex függvények és a $\mathbf{h} : R^n \to R^p$ leképezés affin, azaz $\mathbf{h}(\mathbf{x}) = \mathbf{A}\mathbf{x} - \mathbf{b}$, $\mathbf{x} \in R^n$, ahol $\mathbf{A}$ $(p \times n)$-es mátrix és $\mathbf{b} \in R^p$. Tekintsük a következő két rendszert:*

$$\alpha(\mathbf{x}) < 0, \quad \mathbf{h}(\mathbf{x}) = \mathbf{0}, \quad \mathbf{g}(\mathbf{x}) \ge \mathbf{0}, \quad \mathbf{x} \in A, \tag{6.7}$$

$$\delta\alpha(\mathbf{x}) + \boldsymbol{\mu}^T\mathbf{h}(\mathbf{x}) - \boldsymbol{\lambda}^T\mathbf{g}(\mathbf{x}) \ge 0,\ \mathbf{x} \in A,\ (\delta, \boldsymbol{\lambda}^T)^T \ge \mathbf{0},\ (\delta, \boldsymbol{\mu}^T, \boldsymbol{\lambda}^T)^T \ne \mathbf{0}. \tag{6.8}$$

Ha a (6.7) rendszernek nincs megoldása, akkor a (6.8) rendszernek van $(\delta, \boldsymbol{\mu}^T, \boldsymbol{\lambda}^T)^T$ megoldása. Ha $\delta > 0$, akkor az állítás megfordítása is igaz.

Bizonyítás. Tegyük fel, hogy a (6.7) rendszernek nincs megoldása. Mivel az $\alpha, -\mathbf{g}$ függvények konvexek, a $\mathbf{h}$ affin, ezért a

$$B = \{(\beta, \mathbf{r}^T, \mathbf{q}^T)^T | \, \beta > \alpha(\mathbf{x}), \ \ \mathbf{r} = \mathbf{h}(\mathbf{x}), \ \ \mathbf{q} \ge -\mathbf{g}(\mathbf{x}), \quad \text{valamely} \quad \mathbf{x} \in A \quad \text{pontban}\} \tag{6.9}$$

halmaz konvex. A feltevés szerint a (6.7) rendszernek nincs megoldása, ezért $(0, \mathbf{0}^T, \mathbf{0}^T)^T \notin B$. Ha $(0, \mathbf{0}^T, \mathbf{0}^T)^T \notin cl\, B$ vagy $(0, \mathbf{0}^T, \mathbf{0}^T)^T \in cl\, B$, ahol clB jelöli a B halmaz lezártját, akkor az *Elválasztási tétel* szerint létezik olyan $(\delta, \boldsymbol{\mu}^T, \boldsymbol{\lambda}^T)^T \ne \mathbf{0}$, amelyre

$$\delta\beta + \boldsymbol{\mu}^T\mathbf{r} + \boldsymbol{\lambda}^T\mathbf{q} \ge 0, \qquad \forall(\beta, \mathbf{r}^T, \mathbf{q}^T)^T \in clB. \tag{6.10}$$

$\Big($A $(0, \mathbf{0}^T, \mathbf{0}^T)^T \in cl\, B$ esetben létezik olyan $(\beta_k, \mathbf{r}_k^T, \mathbf{q}_k^T)^T \notin cl\, B$, $k = 1, 2, \ldots$ sorozat, amelyre $\lim\limits_{k\to\infty} (\beta_k, \mathbf{r}_k^T, \mathbf{q}_k^T)^T = (0, \mathbf{0}^T, \mathbf{0}^T)^T$. Az *Elválasztási tétel* szerint ehhez a sorozathoz megadható olyan $(\delta_k, \boldsymbol{\mu}_k^T, \boldsymbol{\lambda}_k^T)^T$, $k = 1, 2, \ldots$ sorozat, amelyre az teljesül, hogy $\|(\delta_k, \boldsymbol{\mu}_k^T, \boldsymbol{\lambda}_k^T)^T\| = 1, \forall k$ és

$$\delta_k\beta_k + \boldsymbol{\mu}_k^T\mathbf{r}_k + \boldsymbol{\lambda}_k^T\mathbf{q}_k < \delta_k\beta + \boldsymbol{\mu}_k^T\mathbf{r} + \boldsymbol{\lambda}_k^T\mathbf{q}, \qquad \forall(\beta, \mathbf{r}^T, \mathbf{q}^T)^T \in cl\, B.$$

Mivel a $(\delta_k, \boldsymbol{\mu}_k^T, \boldsymbol{\lambda}_k^T)^T$, $k = 1, 2, \dots$ sorozat korlátos, ezért van konvergens részsorozata. Az *Elválasztási tétel* felhasználásával nyert szigorú egyenlőtlenségekből tekintsük a konvergens részsorozathoz tartozókat, majd a *cl* B halmaz minden elemét lerögzítve és elvégezve a határátmenetet kapjuk az állítást.$\Big)$

Legyen $\widehat{\mathbf{x}} \in A$. Mivel a β értéke és a $\mathbf{q}$ vektor komponensei tetszőlegesen nagyok lehetnek, a (6.10) egyenlőtlenség csak akkor teljesülhet, ha $\delta \geq 0$ és $\boldsymbol{\lambda} \geq \mathbf{0}$. Továbbá,

$$(\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{r}}^T, \widehat{\mathbf{q}}^T)^T = \Big(\alpha(\widehat{\mathbf{x}}), \;\; \mathbf{h}(\widehat{\mathbf{x}})^T, \;\; -\mathbf{g}(\widehat{\mathbf{x}})^T\Big)^T \in clB.$$

A (6.10) egyenlőtlenségből következik, hogy

$$\delta\alpha(\widehat{\mathbf{x}}) + \boldsymbol{\mu}^T\mathbf{h}(\widehat{\mathbf{x}}) - \boldsymbol{\lambda}^T\mathbf{g}(\widehat{\mathbf{x}}) \geq 0. \tag{6.11}$$

Mivel a fenti egyenlőtlenség teljesül minden $\widehat{\mathbf{x}} \in A$ pontban, ezért a (6.8) rendszernek van megoldása.

Az állítás megfordításának az igazolásához tegyük fel, hogy a (6.8) rendszernek van $(\delta, \boldsymbol{\mu}^T, \boldsymbol{\lambda}^T)^T$ megoldása, amire teljesülnek a

$$\delta\alpha(\mathbf{x}) + \boldsymbol{\mu}^T\mathbf{h}(\mathbf{x}) - \boldsymbol{\lambda}^T\mathbf{g}(\mathbf{x}) \geq 0, \quad \mathbf{x} \in A, \quad \delta > 0, \qquad \boldsymbol{\lambda} \geq \mathbf{0},$$

egyenlőtlenségek. Legyen $\widehat{\mathbf{x}} \in A$ olyan vektor, amelyre $-\mathbf{g}(\widehat{\mathbf{x}}) \leq \mathbf{0}$ és $\mathbf{h}(\widehat{\mathbf{x}}) = \mathbf{0}$. Mivel $\boldsymbol{\lambda} \geq \mathbf{0}$, a fenti egyenlőtlenségből következik, hogy $\delta\alpha(\widehat{\mathbf{x}}) \geq 0$. A $\delta > 0$ és az $\alpha(\widehat{\mathbf{x}}) \geq 0$ egyenlőtlenségek miatt a (6.7) rendszernek nincs megoldása, ami az állítást bizonyítja. ■

A következő tételben megmutatjuk, hogy a primál és a duál feladat optimális célfüggvény értéke megegyezik a feltételi függvényekre vonatkozó konkávitási és megfelelően választott regularitási feltételek mellett.

6.2. Erős dualitási tétel.

Tegyük fel, hogy $A \subseteq R^n$ *nem üres konvex halmaz,* $-g_i : R^n \to R$, $i = 1, \dots, m$, *konvex függvények,* h_j, $j = 1, \dots, p$, *affin függvények, azaz* $\mathbf{h}(\mathbf{x}) = \mathbf{A}\mathbf{x} - \mathbf{b}$, *ahol* $\mathbf{A}$ *adott* $(p \times n)$*-es mátrix,* $\mathbf{b} \in R^p$ *rögzített vektor, és létezik olyan* $\hat{\mathbf{x}} \in A$ *vektor,* *amelyre*

$$\mathbf{g}(\hat{\mathbf{x}}) > \mathbf{0}, \qquad \mathbf{h}(\hat{\mathbf{x}}) = \mathbf{0}, \qquad \mathbf{0} \in int\, \mathbf{h}(A), \tag{6.12}$$

ahol $\mathbf{h}(A) = \{\mathbf{h}(\mathbf{x}) \mid \mathbf{x} \in A\}$. *Akkor*

$$\inf_{\mathbf{x}}\{f(\mathbf{x}), \mathbf{x} \in M[\mathbf{h}, \mathbf{g}] \cap A\} = \sup_{(\boldsymbol{\mu}, \boldsymbol{\lambda})}\left\{\inf_{\mathbf{x}\in A}\{L(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\lambda})|\ \boldsymbol{\mu} \in R^p,\ \boldsymbol{\lambda} \geq \mathbf{0}\right\}. \tag{6.13}$$

Ha a primál feladatban a célfüggvény értéke véges, akkor a duál feladatban is véges és felvétetik. Ha a primál feladat optimális megoldása $\mathbf{x}^*$, *a duál feladaté* $(\boldsymbol{\mu}^{*T}, \boldsymbol{\lambda}^{*T})^T$, *akkor* $\boldsymbol{\lambda}^{*T}\mathbf{g}(\mathbf{x}^*) = 0$.

Bizonyítás. Jelölje a primál feladat optimális célfüggvény értékét γ. Ha $\gamma = -\infty$, akkor a *6.1. Következmény* miatt az optimális duál célfüggvény érték is $-\infty$, ezért (6.13) igaz. Tegyük fel, hogy a γ érték véges, és tekintsük a következő rendszert:

$$f(\mathbf{x}) - \gamma < 0, \qquad \mathbf{h}(\mathbf{x}) = \mathbf{0}, \qquad \mathbf{g}(\mathbf{x}) \geq \mathbf{0}, \qquad \mathbf{x} \in A \subseteq R^n. \tag{6.14}$$

A γ definíciója miatt a (6.14) rendszernek nincs megoldása. A *6.1. Lemma* állításából következik, hogy a megfelelő (6.8) rendszernek létezik nem nulla $(\delta, \boldsymbol{\mu}^T, \boldsymbol{\lambda}^T)^T$, $(\delta, \boldsymbol{\lambda}^T)^T \geq \mathbf{0}$, megoldása.

Először megmutatjuk, hogy $\delta > 0$. Tegyük fel az állítással ellentétben, hogy $\delta = 0$. A feltevések miatt létezik olyan $\hat{\mathbf{x}} \in A$ vektor, amelyre $\mathbf{h}(\hat{\mathbf{x}}) = \mathbf{0}$ és $\mathbf{g}(\hat{\mathbf{x}}) > \mathbf{0}$. Ezt behelyettesítve a (6.14)-nek megfelelő (6.8) rendszerbe azt kapjuk, hogy $-\boldsymbol{\lambda}^T\mathbf{g}(\hat{\mathbf{x}}) \geq 0$. Mivel $\mathbf{g}(\hat{\mathbf{x}}) > \mathbf{0}$ és $\boldsymbol{\lambda} \geq \mathbf{0}$, ezért $\boldsymbol{\lambda}^T\mathbf{g}(\hat{\mathbf{x}}) \leq 0$ csak akkor lehetséges, ha $\boldsymbol{\lambda} = \mathbf{0}$. Mivel $(\delta, \boldsymbol{\lambda}^T)^T = \mathbf{0}$, ezért $\boldsymbol{\mu}^T\mathbf{h}(\mathbf{x}) \geq 0$, $\mathbf{x} \in A$. Azonban a $\mathbf{0} \in$ int $\mathbf{h}(A)$ feltétel miatt tudunk találni olyan $\mathbf{x} \in A$ vektort, amelyre $\mathbf{h}(\mathbf{x}) = -\varepsilon\boldsymbol{\mu}$, ahol $\varepsilon > 0$. Ezért $0 \leq \boldsymbol{\mu}^T\mathbf{h}(\mathbf{x}) = -\varepsilon\|\boldsymbol{\mu}\|^2$, amiből következik, hogy $\boldsymbol{\mu} = \mathbf{0}$. Tehát, a $\delta = 0$ feltevésből adódik, hogy $(\delta, \boldsymbol{\mu}^T, \boldsymbol{\lambda}^T)^T = \mathbf{0}$, ami ellentmondás. Így beláttuk, hogy $\delta > 0$.

A (6.14)-nek megfelelő (6.8) egyenlőtlenség

$$\delta\big(f(\mathbf{x})-\gamma\big)+\boldsymbol{\mu}^T\mathbf{h}(\mathbf{x})-\boldsymbol{\lambda}^T\mathbf{g}(\mathbf{x}) \geq 0, \quad \mathbf{x}\in A\subseteq R^n, \quad (\delta,\boldsymbol{\lambda}^T)^T \geq \mathbf{0}, \quad (\delta,\boldsymbol{\mu}^T,\boldsymbol{\lambda}^T)^T \neq 0. \tag{6.15}$$

Ezt osztva a δ értékkel:

$$f(\mathbf{x})+\frac{1}{\delta}\boldsymbol{\mu}^T\mathbf{h}(\mathbf{x})-\frac{1}{\delta}\boldsymbol{\lambda}^T\mathbf{g}(\mathbf{x}) \geq \gamma, \quad \mathbf{x}\in A\subseteq R^n, \tag{6.16}$$

ami azt mutatja, hogy

$$\inf_{\mathbf{x}\in A} L(\mathbf{x},\frac{1}{\delta}\boldsymbol{\mu},\frac{1}{\delta}\boldsymbol{\lambda}) \geq \gamma.$$

A *Gyenge dualitási tétel*t alkalmazva kapjuk a (6.13) egyenlőséget.

Tegyük fel, hogy $\mathbf{x}^*$ a primál feladat optimális megoldása és $f(\mathbf{x}^*)=\gamma$. A *6.1 Lemma* és (6.13) szerint a duál feladatnak is van optimális megoldása. A (6.16) egyenlőtlenséget tekintve az $\mathbf{x}^*$ pontban adódik, hogy $\boldsymbol{\lambda}^T\mathbf{g}(\mathbf{x}^*)\leq 0$. Mivel $\boldsymbol{\lambda}\geq\mathbf{0}$ és $\mathbf{g}(\mathbf{x}^*)\geq\mathbf{0}$, ezért $\boldsymbol{\lambda}^T\mathbf{g}(\mathbf{x}^*)=0$, ami az állítás. ∎

Az *Erős dualitási tétel*ben a (6.12) feltételek a *Slater feltétel* ($\exists\widehat{\mathbf{x}}\in A$ úgy, hogy $\mathbf{g}(\widehat{\mathbf{x}})>\mathbf{0}$) általánosításának tekinthetők. Az *Erős dualitási tétel* speciális eseteként adódik a lineáris optimalizálás dualitási tétele.

6.3. Nyeregpont tétel.

Ha $A\subseteq R^n$ nem üres halmaz, $f,g_i,h_j: R^n\to R$, $i=1,\ldots,m;\ j=1,\ldots,p$, függvények és léteznek $\mathbf{x}^*\in A\subseteq R^n, (\boldsymbol{\mu}^{*T},\boldsymbol{\lambda}^{*T})^T\in R^p\times R^m, \boldsymbol{\lambda}^*\geq\mathbf{0}$, úgy, hogy

$$\begin{gathered} L(\mathbf{x}^*,\boldsymbol{\mu},\boldsymbol{\lambda}) \leq L(\mathbf{x}^*,\boldsymbol{\mu}^*,\boldsymbol{\lambda}^*) \leq L(\mathbf{x},\boldsymbol{\mu}^*,\boldsymbol{\lambda}^*), \\ \mathbf{x}\in A\subseteq R^n,\ (\boldsymbol{\mu}^T,\boldsymbol{\lambda}^T)^T\in R^p\times R^m, \quad \boldsymbol{\lambda}\geq\mathbf{0}, \end{gathered} \tag{6.17}$$

akkor $\mathbf{x}^*$ a primál feladat és $(\boldsymbol{\mu}^{*T},\boldsymbol{\lambda}^{*T})^T$ a duál feladat megoldása.

Megfordítva, tegyük fel, hogy A konvex halmaz, $-g_i, i=1\ldots,m$, konvex, $h_j, i=1\ldots,p$, affin függvények és a (6.12) feltétel teljesül. Ha $\mathbf{x}^*$ a primál feladat egy optimális megoldása, akkor léteznek olyan $(\boldsymbol{\mu}^{*T},\boldsymbol{\lambda}^{*T})^T\in R^p\times R^m$, $\boldsymbol{\lambda}^*\geq\mathbf{0}$, vektorok, amelyekre teljesülnek a (6.17) egyenlőtlenségek.

Bizonyítás. Tegyük fel, hogy léteznek olyan

$$\mathbf{x}^* \in A \subseteq R^n, \quad (\boldsymbol{\mu}^{*T}, \boldsymbol{\lambda}^{*T})^T \in R^p \times R^m, \quad \boldsymbol{\lambda}^* \geq \mathbf{0},$$

vektorok, amelyekre a (6.17) egyenlőtlenségek teljesülnek. Mivel

$$\begin{gathered} f(\mathbf{x}^*) + \sum_{j=1}^{p} \mu_j h_j(\mathbf{x}^*) - \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}^*) = L(\mathbf{x}^*, \boldsymbol{\mu}, \boldsymbol{\lambda}) \leq L(\mathbf{x}^*, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*)\,, \\ \left(\boldsymbol{\mu}^T, \boldsymbol{\lambda}^T\right)^T \in R^p \times R^m, \quad \boldsymbol{\lambda} \geq \mathbf{0}, \end{gathered} \tag{6.18}$$

ebből következik, hogy

$$\mathbf{h}(\mathbf{x}^*) = \mathbf{0} \qquad \text{és} \qquad \mathbf{g}(\mathbf{x}^*) \geq \mathbf{0},$$

tehát, $\mathbf{x}^*$ a primál feladat egy megengedett megoldása. A (6.18) egyenlőtlenségből a $\boldsymbol{\lambda} = \mathbf{0}$, $\boldsymbol{\mu} = \mathbf{0}$ választás mellett következik, hogy $\boldsymbol{\lambda}^{*T}\mathbf{g}(\mathbf{x}^*) \leq \mathbf{0}$. Mivel $\boldsymbol{\lambda}^* \geq \mathbf{0}$ és $\mathbf{g}(\mathbf{x}^*) \geq \mathbf{0}$, ezért $\boldsymbol{\lambda}^{*T}\mathbf{g}(\mathbf{x}^*) = 0$.

A (6.17) egyenlőtlenségek alapján

$$\begin{gathered} f(\mathbf{x}^*) = f(\mathbf{x}^*) + \sum_{j=1}^{p} \mu_j^* h_j(\mathbf{x}^*) - \sum_{i=1}^{m} \lambda_i^* g_i(\mathbf{x}^*) = L(\mathbf{x}^*, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*) \leq \\ L(\mathbf{x}, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*) = f(\mathbf{x}) + \sum_{j=1}^{p} \mu_j^* h_j(\mathbf{x}) - \sum_{i=1}^{m} \lambda_i^* g_i(\mathbf{x})\,, \quad \mathbf{x} \in A, \end{gathered} \tag{6.19}$$

amiből következik, hogy $f(\mathbf{x}^*) \leq \inf_{\mathbf{x} \in A} L(\mathbf{x}, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*)\,.$

Mivel $\mathbf{x}^*$ a primál feladat egy megengedett megoldása és $\boldsymbol{\lambda}^* \geq \mathbf{0}$, a *6.2. Következményből* azt kapjuk, hogy $\mathbf{x}^*$ a primál feladat és $\left(\boldsymbol{\mu}^{*T}, \boldsymbol{\lambda}^{*T}\right)^T$ pedig a duál feladat optimális megoldása.

A fordított irány bizonyításához tegyük fel, hogy $\mathbf{x}^*$ a primál feladat egy optimális megoldása. Az *Erős dualitási tétel* miatt léteznek olyan $\left(\boldsymbol{\mu}^{*T}, \boldsymbol{\lambda}^{*T}\right)^T$, $\boldsymbol{\lambda}^* \geq \mathbf{0}$, Lagrange szorzók, amelyekre

$$f(\mathbf{x}^*) = \inf_{\mathbf{x} \in A} L(\mathbf{x}, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*) \qquad \text{és} \qquad \boldsymbol{\lambda}^{*T}\mathbf{g}(\mathbf{x}^*) = 0, \tag{6.20}$$

amiből következik, hogy

$$f(\mathbf{x}^*) = \inf_{\mathbf{x}\in A} L(\mathbf{x}, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*) \le f(\mathbf{x}) + \sum_{j=1}^{p} \mu_j^* h_j(\mathbf{x}) - \sum_{i=1}^{m} \lambda_i^* g_i(\mathbf{x}) = L(\mathbf{x}, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*)\,, \quad \mathbf{x} \in A \subseteq R^n.$$

Mivel $\boldsymbol{\lambda}^{*T}\mathbf{g}(\mathbf{x}^*) = \boldsymbol{\mu}^{*T}\mathbf{h}(\mathbf{x}^*) = 0$,

$$L(\mathbf{x}^*, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*) = f(\mathbf{x}^*) + \sum_{j=1}^{p} \mu_j^* h_j(\mathbf{x}^*) - \sum_{i=1}^{m} \lambda_i^* g_i(\mathbf{x}^*) \le L(\mathbf{x}, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*)\,, \quad \mathbf{x} \in A \subseteq R^n.$$

ami éppen (6.17) második egyenlőtlensége. Az első egyenlőtlenség triviálisan teljesül (6.17)-ben a $\boldsymbol{\lambda}^{*T}\mathbf{g}(\mathbf{x}^*) = 0$, $\mathbf{h}(\mathbf{x}^*) = \mathbf{0}$, $\mathbf{g}(\mathbf{x}^*) \ge \mathbf{0}$, $\boldsymbol{\lambda} \ge \mathbf{0}$, relációk miatt, ami a tétel állítását bizonyítja. ■

A *Karush-Kuhn-Tucker* és a nyeregpont feltételek között szoros kapcsolat van, mivel a bennük szereplő Lagrange szorzók konvexitási feltételek mellett megegyeznek. A dualitás elmélet mélyebb tárgyalása megtalálható pl. Ujvári (2001) jegyzetében.

7. fejezet

VÁLTOZÓ METRIKÁJÚ MÓDSZEREK

Ez a fejezet az NLO megoldására szolgáló iterációs módszerek családjába tartozó, *változó metrikájú módszer*ek elméleti megalapozásával foglalkozik. A változó metrikájú módszerek onnan kapták a nevüket, hogy a *csökkenési irány*[3] meghatározásához minden iterációs pontban bevezetünk egy szimmetrikus, pozitív definit mátrixot, ami egy skaláris szorzást is meghatároz. Az iterációs módszer azt jelenti, hogy az eljárás során generált pontsorozat minden eleme az előző lépésben nyert pontból származtatható olyan módon, hogy a pontsorozathoz tartozó függvény értékek csökkenő sorozatot alkotnak.

Az iteratív algoritmusok elmélete három részre osztható. Az első foglalkozik az algoritmusok építésével, figyelembe véve a feladat struktúráját és az informatikai lehetőségeket. A második rész fő kérdése a módszerek konvergenciája, még pontosabban a globális konvergencia vizsgálata, aminek megléte azt jelenti, hogy a módszer által generált pontsorozat a megengedett tartomány tetszőleges pontjából indulva tart (konvergál) egy (a) stacionárius ponthoz, egy (a) lokális minimumhoz vagy egy (a) globális minimumhoz. Sok fontos módszer osztály nem teljesíti ezt a tulajdonságot, és az induló pontoktól függően a generált pontsorozat divergens (nem konvergens) is lehet vagy nem a megoldáshoz konvergál. Az elmélet harmadik

[3] A célfüggvény értéke csökken vagy nem nő a csökkenési irányokban.

része a lokális konvergencia vizsgálatával foglalkozik, ami a konvergencia sebesség vizsgálatát jelenti.

Az e fejezetben szereplő tárgyalás speciális esete Rapcsák (1977) könyvében, illetve a [56, 58], cikkekben találhatónak.

Tekintsük az NLO-t a következő formában:

$$\begin{aligned} &\min f(\mathbf{x}) \\ &\mathbf{x} \in A \subseteq M \subseteq R^n, \end{aligned} \tag{7.1}$$

ahol

$$\begin{aligned} M = M[\mathbf{h}] &= \{\mathbf{x} \in R^n |\ h_j(\mathbf{x}) = 0, \quad j = 1, \dots, p\}, \\ TM[\mathbf{h}] &= \{\mathbf{v} \in R^n |\ \nabla h_j(\mathbf{x})\mathbf{v} = 0, \quad j = 1, \dots, p\}, \\ A = M[\mathbf{h}, \mathbf{g}] &= \{\mathbf{x} \in M |\ g_i(\mathbf{x}) \geq 0, \quad i = 1, \dots, m\}, \end{aligned} \tag{7.2}$$

R^n az n-dimenziós Euklideszi tér és $f, h_j,\ g_i,\ j = 1, \dots, p,\ i = 1, \dots, m$, kétszer folytonosan differenciálható függvények. Először stacionárius pont vagy lokális minimum meghatározására szolgáló változó metrikájú módszerekre adunk meg általános algoritmus leírást egyenlőség feltételek esetén.

Tegyük fel, hogy a LICQ regularitási feltétel teljesül minden $\mathbf{x} \in M[\mathbf{h}]$ pont esetén. Vezessünk be az n-dimenziós Euklideszi téren egy $\mathbf{G}_1(\mathbf{x}),\ \mathbf{x} \in R^n$, mátrix (értékű) függvényt, ami a következő tulajdonságokkal rendelkezik:

1) $\mathbf{G}_1(\mathbf{x}),\ \mathbf{x} \in R^n$, szimmetrikus mátrixok;

2) $\mathbf{G}_1(\mathbf{x}),\ \mathbf{x} \in R^n$, pozitív definit mátrixok;

3) a $\mathbf{G}_1(\mathbf{x}),\ \mathbf{x} \in R^n$, mátrixok által meghatározott kvadratikus alakok értéke nem változik reguláris (egyetlen pontban sem szinguláris Jacobi mátrixú) koordináta transzformációk esetén.

Az első két tulajdonság alapján, a $\mathbf{G}_1$ és a $\mathbf{G}_1^{-1}$ leképzések az n-dimenziós Euklideszi tér minden pontjában meghatároznak egy-egy skaláris szorzást az Euklideszi térre vonatkozóan. A harmadik tulajdonságot explicit formában nem használjuk fel az anyagban.

Legyen $\mathbf{x}_0 \in M$ az induló pont, $\mathbf{x}_k \in M$ a k-adik iterációs lépésben nyert megengedett megoldás, $\mathbf{D}_k$ $(n \times n)$-es szimmetrikus mátrix, aminek $TM_{\mathbf{x}_k}$ invariáns altere minden k esetén, továbbá $\mathbf{D}_k$ pozitív definit a $TM_{\mathbf{x}_k}$ altéren minden iterációs pontban úgy, hogy az altereken vett legkisebb saját értékek alulról egyenletesen korlátosak. Ez utóbbi tulajdonság azt jelenti, hogy létezik olyan $\varepsilon > 0$ érték, aminél minden a $TM_{\mathbf{x}_k}$ altéren vett legkisebb sajátérték nagyobb vagy egyenlő. Az M halmazban haladó görbét *geodetikus*nak nevezzük, ha a második derivált vektora minden pontban merőleges az M halmaz érintősíkjára. Be lehet látni, hogy minden pontból minden irányban pontosan egy geodetikus indul, ami az adott pont egy kicsiny környezetében egyértelmű. Ezt a fogalmat ki lehet terjeszteni arra az esetre is, amikor az R^n téren a skaláris szorzatot szimmetrikus és pozitív definit mátrix függvény határozza meg (Spivak, 1979).

Vezessük be a következő jelölést:

$$\nabla^G f^T = (\mathbf{I} - \mathbf{G}_1^{-1} J\mathbf{h}^T (J\mathbf{h}\mathbf{G}_1^{-1} J\mathbf{h}^T)^{-1} J\mathbf{h})\mathbf{G}_1^{-1} \nabla f^T,$$

ahol $J\mathbf{h}$ a $\mathbf{h} : R^n \to R^p$ leképezés $p \times n$-es Jacobi mátrix leképezése, $\mathbf{I}$ az $n \times n$-es egységmátrix és $\mathbf{G}^{-1}$ az $n \times n$-es inverz mátrix leképezés.

Belátható, hogy az M halmazon értelmezett, $\nabla^{\mathbf{G}} f^T$ leképezés tetszőleges pontban a $\mathbf{G}_1^{-1} \nabla f^T$ vektor ortogonális leképzése a $\mathbf{G}_1$ skaláris szorzat szerint a ponthoz tartozó TM altérre [56]. A $\mathbf{G}_1$ leképezés minden M halmazbeli pontban értelmezett TM altéren meghatároz (indukál) egy skaláris szorzást, amit $\mathbf{G}$ jelöl.

Az általános algoritmus váz a (7.1) NLO stacionárius pontjának vagy lokális minimumának a meghatározására:

1. Határozzuk meg a csökkenési irányt a következő módon:

$$\mathbf{p}_k = -\mathbf{D}_k^2 \nabla^G f(\mathbf{x}_k)^T. \tag{7.3}$$

2.

$$\mathbf{x}_{k+1} = \gamma_{\mathbf{x_k}}(t_k, \mathbf{p}_k), \tag{7.4}$$

ahol a t_k lépéshossz meghatározása történhet az alábbi

$$t_k = \arg\min\{f(\gamma_{\mathbf{x_k}}(t, \mathbf{p}_k))|\ t \geq 0\} \tag{7.5}$$

$\mathbf{x}_k$ pontból a $\mathbf{p}_k$ irányban induló geodetikus görbe menti minimalizálás elvégzésével vagy valamilyen heurisztikus szabály segítségével (pl. Luenberger, 1973). Az Armijo szabály[3] esetében $\alpha \in \left(0, \frac{1}{2}\right)$ és $t_k = 2^{-l_k}$, ahol l_k az a legkisebb egész szám, amelyre teljesül a következő egyenlőtlenség:

$$f(\gamma_{\mathbf{x}_k}(2^{-l_k},\ \mathbf{p}_k)) \leq f(\mathbf{x}_k) - \alpha 2^{-l_k} \|\nabla^G f(\mathbf{x}_k)\mathbf{G}_1(\mathbf{x}_k)\mathbf{D}_k^2 \nabla^G f(\mathbf{x}_k)^T\|. \tag{7.6}$$

Ez az általános algoritmus váz számos ismert nemlineáris optimalizálási algoritmust tartalmaz. A $\mathbf{D}_k = \mathbf{I},\ k = 1, 2, \ldots,$ esetben:

ha

$$J\mathbf{h}(\mathbf{x}) = 0, \qquad \mathbf{G}_1(\mathbf{x}) = \mathbf{I}, \qquad \mathbf{x} \in R^n,$$

akkor $\mathbf{p}_k = -\nabla f^T$ és a csökkenő irányok módszerét kapjuk;

ha

$$J\mathbf{h}(\mathbf{x}) = 0, \qquad \mathbf{G}_1(\mathbf{x}) = Hf(\mathbf{x}), \qquad \mathbf{x} \in R^n,$$

ahol Hf a célfüggvény Hesse mátrixa, akkor $\mathbf{p}_k = -H^{-1}\nabla f^T$ és a Newton módszert kapjuk;

ha

$$J\mathbf{h}(\mathbf{x}) = \mathbf{A}, \qquad \mathbf{x} \in R^n_+,$$

ahol $\mathbf{A}$ teljes rangú $(p \times n)$-es mátrix és

$$\mathbf{G}_1(\mathbf{x}) = \begin{pmatrix} \frac{1}{x_1^2} & & \\ & \ddots & \\ & & \frac{1}{x_n^2} \end{pmatrix}, \qquad \mathbf{x} \in R^n_+,$$

[3] lásd pl. Ortega and Rheinboldt, Academic Press, 1970.

ahol R^n_+ a pozitív ortánst jelöli, a lineáris optimalizálás affin vektor mezőjét kapjuk a Karmarkar (1990) híres cikkében bevezetett lineáris optimalizálási kanonikus alakra; ha

$$\mathbf{G}_1(\mathbf{x}) = \mathbf{I}, \qquad \mathbf{x} \in R^n_+,$$

akkor $\mathbf{p}_k = (\mathbf{I} - J\mathbf{h}^T(J\mathbf{h}J\mathbf{h}^T)^{-1}J\mathbf{h})\nabla f^T$ és a gradiens vetítési módszert kapjuk; ha

$$\mathbf{G}_1(\mathbf{x}) = Hf(\mathbf{x}), \qquad \mathbf{x} \in R^n,$$

akkor vetített Newton-típusú módszert kapunk;
ha

$$J\mathbf{h}(\mathbf{x}) = \mathbf{A}, \qquad \mathbf{x} \in R^n_+,$$

ahol $\mathbf{A}$ teljes rangú $(p \times n)$-es mátrix és

$$\mathbf{G}_1(\mathbf{x}) = Hf(\mathbf{x}), \qquad \mathbf{x} \in R^n_+,$$

ahol

$$f(\mathbf{x}) = \mathbf{c}^T\mathbf{x} - \mu \sum_{i=1}^{n} \log x_i, \qquad \mathbf{x} \in R^n_+,$$

akkor a standard lineáris optimalizálási feladat logaritmikus büntető függvényét kapjuk, csökkenő iránynak a projektív vektor mezőt, és a μ paraméter megfelelően választott értéke mellett a vetített Newton módszert.

Kvázi-Newton, SQP és konjugált gradiens módszereket a $\mathbf{D}_k$ mátrixok iterációs pontokbeli megfelelő választásával lehet nyerni.

7.1. Newton módszer[5]

Alapvető módszer a numerikus analízisben, operációkutatásban, optimalizáláselméletben és az irányításelméletben.

Alapötlet

Legyen $f : R \to R$ differenciálható függvény. Oldjuk meg az

$$f(x) = 0, \qquad x \in R, \tag{7.1.1}$$

egyenletet. Indulva az $x_0 \in R$ pontból, tekintsük az f függvény lineáris közelítését az x_o pont környezetében:

$$f(x_0 + h) \approx f(x_0) + f'(x_0)h, \qquad h \in R, \tag{7.1.2}$$

és oldjuk meg az adódó

$$f(x_0) + f'(x_0)h = 0, \qquad h \in R, \tag{7.1.3}$$

egyenletet. Így az

$$x_{k+1} = x_k - \big(f'(x_k)\big)^{-1} f(x_k), \qquad k = 0, 1, \ldots \tag{7.1.4}$$

interatív módszert kapjuk, amit először *Newton* javasolt 1669-ben polinomok gyökének meghatározására. A módszert az

$$f(x) = x^3 - 2x - 5 = 0$$

példán mutatta be, ahol az induló pont $x_0 = 2$ volt. A következő lépésben az

$$f(2 + h) = h^3 + 6h^2 + 10h - 1, \qquad h \in R,$$

[5]Polyak, B.T., 2007-es cikke alapján tárgyaljuk.

függvénynek tekintve a lineáris közelítését a

$$10h - 1 = 0, \qquad h \in R,$$

egyenletet kapjuk, amiből $x_1 = 2 + 0,1 = 2.1$, és az iterációt ismételve a sorozat gyorsan konvergál a megoldáshoz.

A (7.1.4) módszert tetszőleges differenciálható függvényekre *Raphson, J.*, javasolta 1690-ben, ezért gyakran *Newton-Raphson* módszerként hivatkoznak rá. *Fourier* 1818-ban bizonyította, hogy a módszer kvadratikusan konvergens a gyök(ök) egy környezetében, *Cauchy* pedig 1829-ben és 1847-ben kiterjesztette a módszert több dimenzióra és a módszer segítségével bizonyította, hogy egy egyenletnek létezik megoldása. *Fine* 1916-ban bebizonyította az n-dimenziós esetben a módszer konvergenciáját a megoldás létezésének feltételezése nélkül, míg *Bennet*– szintén 1916-ban – kiterjesztette a módszert végtelen dimenzióra. *Kantorovich* 1948-ban fejlesztette tovább a módszert függvényterekre.

A *Newton* módszerről részletesebb ismertetés található *Kantorovich-Akilov* (1959, 1977), *Ostrowski* (1960) és *Ortega-Rheinboldt* (1970) könyveiben.

Konvergencia eredmények

Kantorovich (1948) az

$$f(x) = 0, \qquad x \in X, \tag{7.1.5}$$

egyenletet tekintette, ahol $f : X \to Y$ és X, Y *Banach* terek. A javasolt módszer

$$x_{k+1} = x_k - \big(f'(x_k)\big)^{-1} f(x_k), \qquad k = 0, 1, \ldots, \tag{7.1.6}$$

ahol $f'(x_k)$ az x_k pontbeli $f(x_k)$ nemlineáris operátor deriváltja és $\big(f'(x_k)\big)^{-1}$ az inverze.

Tétel (Kantorovich, 1948). *Legyen f definiálva és kétszer folytonosan differenciálható a $B = \{x \in X \mid \| x - x_0 \| \leq r\}$ gömbben úgy, hogy*

- $f'(x_0)$ *invertálható,*
- $\| \big(f'(x_0)\big)^{-1} f(x_0) \| \leq \eta,$
- $\| \big(f'(x_0)\big)^{-1} f''(x) \| \leq K, \ x \in B,$ *és*
- $h = K\eta < \dfrac{1}{2}, \quad r \geq \dfrac{1 - \sqrt{1 - 2h}}{h}\eta.$

Akkor a (7.1.5) *egyenletnek van* $x^* \in B$ *megoldása, a* (7.1.6) *módszer jól definiált és az* x_k *sorozat kvadratikusan konvergál az* x^* *megoldáshoz, azaz*

$$\| x_k - x^* \| \leq \frac{n}{h2^k}(2h)^{2k}. \tag{7.1.7}$$

Ez volt az első numerikus eredmény funkcionál analízisben. Mivel a tételben nincs feltéve, hogy van megoldás, ezért az állítás egzisztencia tétel is.

Egy másik változat:

Tétel (Mysovskikh, 1950). *Legyen* f *definiálva és kétszer folytonosan differenciálható a*

$$B = \{x \in X \mid \| x - x_0 \| \leq r\}$$

gömbben úgy, hogy

- $f'(x)$, $x \in B$, *invertálható,*
- $\| \left(f'(x)\right)^{-1} \| \leq \beta, \ \| f''(x) \| \leq K, \quad \mathbf{x} \in B,$
- $\| f(x_0) \| \leq \eta$ *és*
- $h = K\beta^2\eta < 2, \qquad r \geq \beta\eta \sum_{n=0}^{\infty} \left(\frac{h}{2}\right)^{2^n - 1}.$

Akkor a (7.1.5) *egyenletnek van* $x^* \in B$ *megoldása, a* (7.1.6) *módszer jól definiált és az* x_k *sorozat kvadratikusan konvergál az* x^* *megoldáshoz, azaz*

$$\| x_k - x^* \| \leq \frac{\beta\eta(h/2)^{2^k - 1}}{1 - (h/2)^{2^k}}. \tag{7.1.8}$$

Ez utóbbi állításban f B fölötti invertálhatósága van feltételezve, de $h < 2$, míg az előző állításban $h < 1/2$.

A (7.1.6) *Newton* módszerben minden iterációs pontban ki kell számítani a deriváltat, majd utána meg kell oldani egy lineáris rendszert, ami komoly számítástechnikai nehézségeket eredményezhet. A módosított *Newton* módszerben csak az induló pontban van meghatározva a leképzés inverze. Ez nem rontja el a konvergenciát, azonban a konvergencia sebesség nem kvadratikus, hanem lineáris.

A lokális konvergencia tételek kritikus feltétele a gömb r sugarának meghatározása, ami azt mutatja meg, hogy $\| f(x_0) \|$ mennyire kicsi, azaz x_0 milyen közel van a megoldáshoz.

Cayley 1879-ben észrevette, hogy a módszer globálisan nem minden kezdőpont esetén konvergens. *Cayley* példája a következő: oldjuk meg az $x^3 = 1$ egyenletet a *Newton* módszerrel. Így $f(x) = x^3 - 1, \quad$ és

$$x_{k+1} = x_k - \frac{x_k^3 - 1}{3x_k^2} = \frac{2x_k}{3} + \frac{1}{3x_k^2}, \tag{7.1.9}$$

amiből következik, hogy minden iterációs sorozat, amiben $x_k = 0$ valamilyen k értékre, nem folytatható tovább.

Egy lehetőség ennek elkerülésére a lépéshossz bevezetése. Ezt nemlineáris optimalizálási feladatok esetén vizsgáljuk.

<u>*Newton* módszer a feltétel nélküli optimalizálásra</u>

$$\min f(\mathbf{x}), \qquad \mathbf{x} \in R^n,$$

egy alapvető iterációs eljárás:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k H^{-1} f(\mathbf{x}_k) \nabla f(\mathbf{x}_k)^T,$$

ahol $\alpha_k - t$ az

$$f\Big(\mathbf{x}_k - \alpha H^{-1} f(\mathbf{x}_k) \nabla f(\mathbf{x}_k)^T\Big), \qquad \alpha \in R,$$

egyváltozós függvény α szerinti minimalizálásával kapjuk.

8. fejezet

A VÁLTOZÓ METRIKÁJÚ MÓDSZEREK KONVERGENCIÁJA

Két globális konvergencia tételt bizonyítunk, amihez szükséges az alábbi állítás.

8.1. Tétel (Kirszbraun). *Ha H Hilbert tér, A a H tér részhalmaza, $\Phi : A \to H$ és*

$$\|\Phi(x_1) - \Phi(x_2)\| < L\|x_1 - x_2\|, \qquad \forall x_1, x_2 \in A, \tag{8.1}$$

valamely L konstans esetén, akkor a Φ leképezés kiterjeszthető a teljes H térre oly módon, hogy a (8.1) egyenlőtlenségek ugyanazzal az L Lipschitz konstanssal teljesülnek.

Jelölje W_k az

$$\{\mathbf{x} \in M \subseteq R^n |\ f(\mathbf{x}) \leq f(\mathbf{x}_k)\}$$

nívóhalmaz $\mathbf{x}_k$ pontot tartalmazó összefüggő részhalmazát (komponensét).

8.2. Tétel. *Ha f folytonosan differenciálható, $W_0 \subseteq M$ kompakt halmaz, az $\mathbf{x}_k$ pontsorozatot (7.3), (7.4) és (7.5) generálja, és*

$$\mathbf{D}_k \nabla^G f(\mathbf{x}_k)^T \in TM_{\mathbf{x}_k}, \qquad k = 1, 2, \ldots, \tag{8.2}$$

$$\text{a } \mathbf{D}_k \nabla^G f(\mathbf{x}_k)^T, \forall k, \textit{ leképezések teljesítik a Lipschitz feltételt,} \tag{8.3}$$

$$\begin{gathered}\text{a } \mathbf{D}_k, \forall k, \textit{ mátrixok minden iterációs pontban pozitív definitek a} \\ TM_{\mathbf{x}_k} \textit{ altereken és a kvadratikus formák alulról egyenletesen korlátosak,}\end{gathered} \tag{8.4}$$

$$\mathbf{D}_k \, \mathbf{G}_1(\mathbf{x}_k) = \mathbf{G}_1(\mathbf{x}_k)\mathbf{D}_k, \qquad k = 1, 2, \ldots, \tag{8.5}$$

akkor az $\mathbf{x}_k$ *sorozat vagy véges lépés után stacionárius pontot ad, vagy a végtelen sorozat bármely torlódási pontja stacionárius. Ha a stacionárius pontbeli függvényérték egyedüli, akkor a teljes sorozat konvergál a stacionárius ponthoz.*

Bizonyítás. Ha $\mathbf{x}_k$ stacionárius pont, akkor $\nabla^G f(\mathbf{x}_k) = \mathbf{0}$, és az algoritmus nem generál új iterációs pontot, így egy megállási kritérium alapján az eljárás a k-adik lépésben befejeződik.

Tegyük fel, hogy $\mathbf{x}_k$ nem stacionárius pont. Ezért $\mathbf{p}_k$ csökkenési irány, mivel

$$\begin{gathered}\nabla^G f(\mathbf{x_k})\mathbf{G}_1(\mathbf{x_k})\mathbf{p_k} = -\nabla^G f(\mathbf{x_k})\mathbf{G}_1(\mathbf{x_k})\mathbf{D}_k^2 \nabla^G f(\mathbf{x_k})^T = \\ -\nabla^G f(\mathbf{x_k})\mathbf{D}_k \mathbf{G}_1(\mathbf{x_k})\mathbf{D}_k \nabla^G f(\mathbf{x_k})^T < 0.\end{gathered} \tag{8.6}$$

Jelölje az $\mathbf{x}_k$ pontból a $\mathbf{p}_k$ irányban induló geodetikus ívet $\gamma_{\mathbf{x}_k}(t, \ \mathbf{p}_k)$. A W_0 halmaz kompaktságából következik, hogy bármely W_k halmaz is kompakt, mivel az $\{f(\mathbf{x}_k)\}$ sorozat monoton csökkenő. A *Hopf-Rinow tétel* [65] miatt, adott $\mathbf{x}_k$ és $\mathbf{p}_k$ esetén létezik olyan $\gamma_{\mathbf{x}_k}(t, \mathbf{p}_k)$ geodetikus ív, ami minden $t \in R_+$ esetén értelmezve van és $\gamma_{\mathbf{x}_k}(t, \ \mathbf{p}_k) \in W_k, t \in R_+$, vagy a geodetikus ív értelmezési tartománya zárt intervallum, aminek egyik végpontjában az ív a W_k határán végződik, az ív többi része pedig a W_k halmazban fut.

Legyen

$$\bar{t}_k = \limsup J_k,$$

ahol

$$J_k = \{t \in R_+ | \gamma_{\mathbf{x}_k}(t, \ \mathbf{p}_k) \text{ értelmezve van és } f\big(\gamma_{\mathbf{x}_k}(t, \mathbf{p}_k)\big) < f\big(\gamma_{\mathbf{x}_k}(0, \mathbf{p}_k)\big) = f(\mathbf{x}_k)\}.$$

A J_k halmaz (8.6) szerint nem üres és

$$\bar{t}_k = +\infty, \qquad \gamma_{\mathbf{x}_k}(t, \mathbf{p}_k) \in W_k, \qquad t \in [0, +\infty),$$

vagy $\bar{t}_k$ véges érték és $\gamma_{\mathbf{x}_k}(t, \mathbf{p}_k) \in W_k$, $t \in [0, \bar{t}_k]$.

Ha az első esetben az egyváltozós függvénynek nincs lokális minimuma, akkor – mivel a függvény monoton csökkenő és alulról korlátos –, a $t_k \to +\infty$ határátmenettel egy W_k-beli pontot határozunk meg. Ebből következik, hogy mindkét esetben a (7.5) szabály szerinti lépéshossz jól definiált és $t_k \in (0, \bar{t}_k)$, azaz az $f\big(\gamma_{\mathbf{x}_k}(t_k, \mathbf{p}_k)\big)$ függvénynek minimuma van a t_k pontban és $f(\gamma_{\mathbf{x}_k}(t_k, \mathbf{p}_k)) = f(\mathbf{x}_{k+1})$.

Ha $\alpha \in (0, \frac{1}{2})$, akkor az

$$\nabla^G f(\gamma_{\mathbf{x}_k}(t, \mathbf{p}_k))\mathbf{G}_1(\gamma_{\mathbf{x}_k}(t, \mathbf{p}_k))\frac{d\gamma_{\mathbf{x}_k}(t, \mathbf{p}_k)}{dt} = \alpha\nabla^G f(\mathbf{x}_k)\mathbf{G}_1(\mathbf{x}_k)\mathbf{p}_k \tag{8.7}$$

egyenletnek a legkisebb megoldása $\hat{t}_k \in (0, t_k)$, mivel a

$$\nabla^G f(\gamma_{\mathbf{x}_k}(t, \mathbf{p}_k))\mathbf{G}_1(\gamma_{\mathbf{x}_k}(t, \mathbf{p}_k))\frac{d\gamma_{\mathbf{x}_k}(t, \mathbf{p}_k)}{dt}$$

folytonos függvény minden értéket felvesz $\nabla^G f(\mathbf{x}_k)\mathbf{G}_1(\mathbf{x}_k)\mathbf{p}_k$ és 0 között, valamint

$$\nabla^G f(\gamma_{\mathbf{x}_k}(t, \mathbf{p}_k))\mathbf{G}_1(\gamma_{\mathbf{x}_k}(t,\ \mathbf{p}_k))\frac{d\gamma_{\mathbf{x}_k}(t, \mathbf{p}_k)}{dt} < \alpha\nabla^G f(\mathbf{x}_k)\mathbf{G}_1(\mathbf{x}_k)\mathbf{p}_k \tag{8.8}$$

teljesül bármely $t \in [0, \hat{t}_k]$ értékre.

Az $\{f(\mathbf{x}_k)\}$ sorozat monoton csökkenő és alulról korlátos, mivel a folytonos f függvény felveszi a minimumát a kompakt W_0 halmazon. Ezért a függvényértékek sorozata konvergens.

Terjesszük ki az iterációs pontokban értelmezett $\mathbf{D}_k\nabla^G f(\mathbf{x}_k)$ leképezést egy az R^n-ben definiált vektorértékű leképezéssé, ami *Kirszbraun tétele* miatt lehetséges.

Két esetet különböztetünk meg.

a) Létezik a $\hat{t}_k$ sorozatnak olyan $\{\hat{t}_{k_i}\}$ részsorozata, amelyik nullához tart. Az általánosság megszorítása nélkül feltehetjük, hogy a $\{\hat{t}_{k_i}\}$ sorozathoz tartozó

$\{\hat{\mathbf{x}}_{k_i}\}$ sorozat és a $\{t_{k_i}\}$ sorozathoz tartozó $\{\mathbf{x}_{k_i}\}$ sorozat is az $\hat{\mathbf{x}} \in W_0$ ponthoz tart.

A (8.7) egyenlőségből a $\mathbf{G}_1$ függvény, a $\nabla^G f$ vektor mező és a $\mathbf{D}$ leképezés folytonosságából nyerjük, hogy

$$\mathbf{D}(\hat{\mathbf{x}})^T\mathbf{G}_1(\hat{\mathbf{x}})\mathbf{D}(\hat{\mathbf{x}}) = \alpha\mathbf{D}(\hat{\mathbf{x}})^T\mathbf{G}_1(\hat{\mathbf{x}})\mathbf{D}(\hat{\mathbf{x}}), \tag{8.9}$$

amiből következik, hogy $\mathbf{D}(\hat{\mathbf{x}}) = \mathbf{0}$. Ha $\nabla^G f(\hat{\mathbf{x}}) \neq \mathbf{0}$, akkor $\nabla^G f(\mathbf{x}_{k_i}) \neq \mathbf{0}$ bármely elég nagy i indexre, azaz a gradiens legalább egy komponense nagyobb, mint egy állandó érték, így a $\mathbf{D}_k$ mátrixok $TM_{\mathbf{x}_k}$ feletti pozitív definitsége miatt $\nabla^G f(\hat{\mathbf{x}})\mathbf{D}(\hat{\mathbf{x}}) > \mathbf{0}$, ami ellentmondás. Ezért, $\nabla^G f(\hat{\mathbf{x}}) = \mathbf{0}$ és $\hat{\mathbf{x}}$ stacionárius pont.

b) Létezik olyan $\beta > 0$ érték, amelyre $\hat{t}_{k_i} \geq \beta$ bármely k_i esetén. Tekintsük a $\{\hat{t}_{k_i}\}$ és $\{t_{k_i}\}$ index sorozatokhoz tartozó $\{\hat{\mathbf{x}}_{k_i}\}$ és $\{\mathbf{x}_{k_i}\}$ részsorozatokat, amelyek konvergálnak az $\hat{\mathbf{x}} \in W_0$ ponthoz. Tegyük fel, hogy $\hat{\mathbf{x}}$ nem stacionárius pont. Akkor,

$$\|\mathbf{D}(\hat{\mathbf{x}})^T\mathbf{G}_1(\hat{\mathbf{x}})\mathbf{D}(\hat{\mathbf{x}})\| = \delta > 0. \tag{8.10}$$

A $\mathbf{G}_1$ mátrix függvény, a $\nabla^G f$ vektormező és a vektor értékű $\mathbf{D}$ folytonosságból következik, hogy

$$\|\frac{df(\gamma_{\mathbf{x}_{k_i}}(0,\mathbf{p}_{k_i}))}{dt}\| = \|\nabla^G f(\mathbf{x}_{k_i})\mathbf{G}_1(\mathbf{x}_{k_i})\mathbf{p}_{k_i}\| > \delta/2 \tag{8.11}$$

teljesül egy konstansnál nagyobb k_i indexek esetén. A középérték tétel és a (8.8) egyenlőtlenség felhasználásával azt kapjuk, hogy

$$\begin{gathered} f(\mathbf{x}_{k_{i+1}}) \leq f(\mathbf{x}_{k_i+1}) \leq f(\hat{\mathbf{x}}_{k_i}) = f(\mathbf{x}_{k_i}) + \\ \hat{t}_{k_i}\nabla^G f(\gamma_{\mathbf{x}_{k_i}}(\tilde{t},\mathbf{p}_{k_i}))\mathbf{G}_1(\gamma_{\mathbf{x}_{k_i}}(\tilde{t},\mathbf{p}_{k_i}))\frac{d\gamma_{\mathbf{x}_{k_i}}(\tilde{t},\mathbf{p}_{k_i})}{dt} < \\ f(\mathbf{x}_{k_i}) + \hat{t}_{k_i}\alpha\nabla^G f(\mathbf{x}_{k_i})\mathbf{G}_1(\mathbf{x}_{k_i})\mathbf{p}_{k_i} < f(\mathbf{x}_{k_i}) - \tfrac{1}{2}\alpha\beta\delta, \quad \tilde{t} \in (0,\hat{t}_{k_i}), \end{gathered} \tag{8.12}$$

ami ellentmond annak, hogy az $\{f(\mathbf{x}_{k_i})\}$ sorozat konvergál az $f(\hat{\mathbf{x}})$ értékhez.

Ezért, $\hat{\mathbf{x}}$ stacionárius pont.

Végül, tegyük fel, hogy $\mathbf{x}^*$ és $\mathbf{x}^{**}$ két különböző W_0 halmazbeli torlódási pontja az $\{\mathbf{x}_k\}$ sorozatnak. A *8.2. Tétel* eddig bizonyított része alapján $\mathbf{x}^*$ és $\mathbf{x}^{**}$ stacionárius pontok. Mivel az $\{f(\mathbf{x}_k)\}$ sorozat konvergens, ezért $f(\mathbf{x}^*) = f(\mathbf{x}^{**})$, ami nem lehetséges, ha a stacionárius értékek egyedüliek. ■

8.3. Tétel. *Ha f folytonosan differenciálható, $W_0 \subseteq M$ kompakt halmaz, az $\mathbf{x}_k$ pontsorozatot (7.3), (7.4) és (7.6) generálja, és*

$$\mathbf{D}_k \nabla^G f(\mathbf{x}_k)^T \in TM_{\mathbf{x}_k}, \qquad k = 1, 2, \dots, \tag{8.13}$$

$$\text{a } \mathbf{D}_k \nabla^G f(\mathbf{x}_k)^T, \forall k, \text{ leképzések teljesítik a Lipschitz feltételt,} \tag{8.14}$$

$$\begin{gathered}\text{a } \mathbf{D}_k, \forall k, \text{ mátrixok minden iterációs pontban pozitív definitek a } TM_{\mathbf{x}_k} \\ \text{altereken és a kvadratikus formák alulról egyenletesen korlátosak,}\end{gathered} \tag{8.15}$$

$$\mathbf{D}_k \mathbf{G}_1(\mathbf{x}_k) = \mathbf{G}_1(\mathbf{x}_k)\mathbf{D}_k, \qquad k = 1, 2, \dots, \tag{8.16}$$

akkor az $\mathbf{x}_k$ sorozat vagy véges lépés után stacionárius pontot ad, vagy a végtelen sorozat bármely torlódási pontja stacionárius. Ha a stacionárius pontbeli függvényérték egyedüli, akkor a teljes sorozat konvergál a stacionárius ponthoz.

Bizonyítás. Ha $\mathbf{x}_k$ stacionárius pont, akkor $\nabla^G f(\mathbf{x}_k) = \mathbf{0}$, és az algoritmus nem generál új iterációs pontot, így egy megállási kritérium alapján az eljárás a k-adik lépésben befejeződik.

Tegyük fel, hogy $\mathbf{x}_k$ nem stacionárius pont. A (8.6) képlet alapján $\mathbf{p}_k$ csökkenési irány. Jelölje az $\mathbf{x}_k$ pontból a $\mathbf{p}_k$ irányban induló geodetikus ívet $\gamma_{\mathbf{x}_k}(t, \mathbf{p}_k)$. A W_0 halmaz kompaktságából következik, hogy bármely W_k halmaz is kompakt, mivel az $\{f(\mathbf{x}_k)\}$ sorozat monoton csökkenő. A *Hopf-Rinow tétel* miatt, adott $\mathbf{x}_k$ és $\mathbf{p}_k$ esetén létezik olyan $\gamma_{\mathbf{x}_k}(t, \mathbf{p}_k)$ geodetikus ív, ami minden $t \in R_+$ esetén értelmezve van és $\gamma_{\mathbf{x}_k}(t, \mathbf{p}_k) \in W_k, t \in R_+$, vagy a geodetikus ív értelmezési tartománya zárt intervallum, aminek egyik végpontjában az ív a W_k határán végződik, az ív többi része pedig a W_k halmazban fut. Ezért a lépéshosszat megadó (7.6) szabály jól definiált.

Az $\{f(\mathbf{x}_k)\}$ sorozat monoton csökkenő és alulról korlátos, mivel a folytonos f függvény felveszi a minimumát a kompakt W_0 halmazon. Ezért, a függvényértékek sorozata konvergens.

Terjesszük ki az iterációs pontokban értelmezett $\mathbf{D}_k\nabla^G f(\mathbf{x}_k)$ leképezést egy az R^n-ben definiált vektor értékű leképezéssé, ami *Kirszbraun tétele* miatt lehetséges.

Két esetet különböztetünk meg.

a) Létezik a t_k sorozatnak olyan $\{t_{k_i} = 2^{-l_{k_i}}\}$ részsorozata, amelyik nullához tart. Az általánosság megszorítása nélkül feltehetjük, hogy a $\{t_{k_i}\}$ sorozathoz tartozó $\{\mathbf{x}_{k_i}\}$ sorozat az $\hat{\mathbf{x}} \in W_0$ ponthoz tart.
A (7.6) lépéshossz szabályban a *Taylor tételt* alkalmazva azt kapjuk, hogy

$$\begin{aligned}
&\alpha 2^{-l_{k_i}} \|\nabla^G f(\mathbf{x}_{k_i})\mathbf{D}_k\mathbf{G}_1(\mathbf{x}_{k_i})\mathbf{D}_k\nabla^G f(\mathbf{x}_{k_i})^T\| \le f(\mathbf{x}_{k_i}) - f(\gamma_{\mathbf{x}_{k_i}}(2^{-l_{k_i}},\ \mathbf{p}_{k_i})) = \\
&f(\mathbf{x}_{k_i}) - f(\mathbf{x}_{k_i}) - 2^{-l_{k_i}}\nabla^G f(\gamma_{\mathbf{x}_{k_i}}(\tilde{t},\mathbf{p}_{k_i}))\mathbf{G}_1(\gamma_{\mathbf{x}_{k_i}}(\tilde{t},\mathbf{p}_{k_i}))\frac{d\gamma_{\mathbf{x}_{k_i}}(\tilde{t},\mathbf{p}_{k_i})}{dt} = \\
&-2^{-l_{k_i}}\nabla^G f(\gamma_{\mathbf{x}_{k_i}}(\tilde{t},\mathbf{p}_{k_i}))\mathbf{G}_1(\gamma_{\mathbf{x}_{k_i}}(\tilde{t},\mathbf{p}_{k_i}))\frac{d\gamma_{\mathbf{x}_{k_i}}(\tilde{t},\mathbf{p}_{k_i})}{dt}, \qquad \tilde{t} \in (0, 2^{-l_{k_i}}),
\end{aligned} \tag{8.17}$$

amiből az következik, hogy

$$\begin{aligned}
&\alpha\|\nabla^G f(\mathbf{x}_{k_i})\mathbf{D}_k\mathbf{G}_1(\mathbf{x}_{k_i})\mathbf{D}_k\nabla^G f(\mathbf{x}_{k_i})^T\| \le \\
&-\nabla^G f(\gamma_{\mathbf{x}_{k_i}}(\tilde{t},\mathbf{p}_{k_i}))\mathbf{G}_1(\gamma_{\mathbf{x}_{k_i}}(\tilde{t},\mathbf{p}_{k_i}))\frac{d\gamma_{\mathbf{x}_{k_i}}(\tilde{t},\mathbf{p}_{k_i})}{dt}, \qquad \tilde{t} \in (0, 2^{-l_{k_i}}).
\end{aligned} \tag{8.18}$$

A (8.18) egyenlőtlenségből és a $\mathbf{G}_1$ mátrix függvény, a $\nabla^G f$ vektormező és a vektor értékű $\mathbf{D}$ függvény folytonosságából azt kapjuk, hogy

$$(1+\alpha)\mathbf{D}(\hat{\mathbf{x}})^T\mathbf{G}_1(\hat{\mathbf{x}})\mathbf{D}(\hat{\mathbf{x}}) \le 0, \tag{8.19}$$

és $\mathbf{D}(\hat{\mathbf{x}}) = \mathbf{0}$. Ha $\nabla^G f(\hat{\mathbf{x}}) \ne \mathbf{0}$, akkor $\nabla^G f(\mathbf{x}_{k_i}) \ne \mathbf{0}$ bármely elég nagy i indexre, azaz a gradiens legalább egy komponense nagyobb, mint egy állandó érték, így a $\mathbf{D}_k$ mátrixok $TM_{\mathbf{x}_k}$ feletti pozitív definitsége miatt $\nabla^G f(\hat{\mathbf{x}})\mathbf{D}(\hat{\mathbf{x}}) > \mathbf{0}$, ami ellentmondás. Ezért, $\nabla^G f(\hat{\mathbf{x}}) = \mathbf{0}$ és $\hat{\mathbf{x}}$ stacionárius pont.

b) Létezik olyan $\beta > 0$ érték, amelyre $t_{k_i} \geq \beta$ bármely k_i esetén. Tekintsük a $\{t_{k_i}\}$ index sorozathoz tartozó $\{\mathbf{x}_{k_i}\}$ pont sorozatot, ami tart az $\hat{\mathbf{x}} \in W_0$ ponthoz. A (7.6) lépéshossz szabályból következik, hogy

$$\alpha 2^{-l_{k_i}} \|\nabla^G f(\mathbf{x}_{k_i}) \mathbf{D}_k \mathbf{G}_1(\mathbf{x}_{k_i}) \mathbf{D}_k \nabla^G f(\mathbf{x}_{k_i})^T\| \leq f(\mathbf{x}_{k_i}) - f(\gamma_{\mathbf{x}_{k_i}}(2^{-l_{k_i}}, \mathbf{p}_{k_i})). \tag{8.20}$$

A (8.20) egyenlőtlenség jobb oldala nullához tart, $2^{-l_{k_i}} \geq \beta$ és a $\mathbf{G}_1$ mátrix függvény, $\nabla^G f$ vektor mező és a $\mathbf{D}$ vektor függvény folytonossága azt adja, hogy

$$\mathbf{D}(\hat{\mathbf{x}})^T \mathbf{G}_1(\hat{\mathbf{x}}) \mathbf{D}(\hat{\mathbf{x}}) = 0, \tag{8.21}$$

és $\mathbf{D}(\hat{\mathbf{x}}) = \mathbf{0}$, amiből $\nabla^G f(\hat{\mathbf{x}}) = \mathbf{0}$.

Végül tegyük fel, hogy $\mathbf{x}^*$ és $\mathbf{x}^{**}$ két különböző W_0 halmazhoz tartozó torlódási pontja az $\{\mathbf{x}_k\}$ sorozatnak. A *8.3 Tétel* eddig bizonyított része alapján $\mathbf{x}^*$ and $\mathbf{x}^{**}$ stacionárius pontok. Mivel az $\{f(\mathbf{x}_k)\}$ sorozat konvergens, ezért $f(\mathbf{x}^*) = f(\mathbf{x}^{**})$, ami nem lehetséges, ha a stacionárius értékek egyedüliek. ■

9. fejezet

SPECIÁLIS STRUKTÚRÁJÚ NEMLINEÁRIS OPTIMALIZÁLÁSI FELADATOK

9.1. Mechanikai erőegyensúly és nemlineáris optimalizálás

Farkas Gyula 1901-ben publikálta a homogén, lineáris egyenlőtlenség-rendszerekre vonatkozó alaptételét, amelyre napjainkban nagyon sok optimalizáláselméleti cikk szerzője hivatkozik. Azonban úgy tűnik, hogy ennek az eredménynek a mechanikai erőegyensúllyal való kapcsolata sem a matematikában, sem a fizikában nem ismert eléggé. Valószínűleg ez az egyik oka annak, hogy a mechanikai erőegyensúllyal kapcsolatos matematikai problémák rendszerezése és egységes tárgyalása eddig nem került előtérbe és a különböző munkákban speciális esetek szerepelnek.

Az analitikus mechanika a klasszikus mechanika egyik ága, amelynek az alapproblémája tömegpontrendszerek mozgásának és egyensúlyi állapotának jellemzése. Fizikusok és matematikusok már hosszú idő óta foglalkoznak ezzel a problémával és jóllehet majdnem minden, mechanikával foglalkozó könyvben található ezzel a témával foglalkozó rész, mégis, a probléma matematikai tárgyalása nem tekinthető lezártnak.

A probléma - amit a newtoni mechanika keretei között tárgyalunk - egyik legáltalánosabb megfogalmazása a következő:

Legyen adott R^3-ban (a *3*-dimenziós Euklideszi térben) n darab tömegpont, amelyekre a $\mathbf{P}_1, \dots, \mathbf{P}_n$ aktív erők hatnak. Ezen kívül legyenek adottak a kényszerek, amik a tömegpontok mozgására az egyedüli korlátozást jelentik.

A feladat a rendszer mechanikai állapotának a meghatározása, vagy speciális esetben a rendszer egyensúlyi állapotának a jellemzése.

A rendszer mechanikai állapotának meghatározása a mozgásegyenletek felírását jelenti. Akkor mondjuk, hogy *a mechanikai rendszer erőegyensúlyban van*, ha minden tömegpontra az aktív erők és a reakcióerők összege nulla, ahol a *reakcióerők*nek azokat az erőket nevezzük, amelyek a tömegpontokat a kényszernek megfelelő pályán tartják.

Az anyagi pontrendszer helyzetét és mozgását a helykoordináták időfüggvényei segítségével adhatjuk meg a $3n$-dimenziós térben. Itt azt az esetet mutatjuk be, amikor a pontrendszer egy lehetséges mozgását egy $\mathbf{x}(t) : [0, 1] \subseteq R \longrightarrow R^{3n}$ kétszer folytonosan differenciálható vektorfüggvényt adja meg, amely eleget tesz a kényszerfeltételeknek. Az $\mathbf{x}'(t)$, $\mathbf{x}''(t)$, $t \in [0, 1]$, vektorok a sebességet, illetve a gyorsulást jelentik.

Az anyagi pontrendszer erőegyensúlyának jellemzéséhez a virtuális munka elvét használjuk fel. A virtuális munka elve a klasszikus mechanika egyik legrégebbi összefüggése, amit Bernoulli mondott ki először 1717-ben az egyenlőség típusú kényszerfeltételek esetére. Az elvet egyenlőtlenség típusú kényszerfeltételek esetére Fourier mondta ki 1798-ban, majd Gauss 1829-ben. Azonban már Cournot 1827-ben és Osztrogradszkij 1838-ban rájött arra, hogy a virtuális munka elvének felhasználásakor nehézséget okoz az, hogy szerepelnek benne a virtuális elmozdulások. Ezért egyenlőség és egyenlőtlenség típusú kényszerfeltételek esetén kiküszöbölték ezeket, azaz felírták a *Farkas tétel állítását*, de nem bizonyították. Ezt Farkas bizonyította először 1898-ben. A *virtuális munka elv*éről, illetve annak történeti vonatkozásairól részletesen olvashatunk a [47], [48] cikkekben. Ez elv szerint, ha a pontrendszer erőegyensúlyban van és a súrlódás elhanyagolható, akkor az aktív erők munkája a virtuális elmozdulások irányában kisebb vagy egyenlő, mint nulla, azaz ha a pontrendszerre ható erőket a $\mathbf{P} \in R^{3n}$ sorvektor jelöli, a virtuális elmozdulásokat a $\mathbf{v} \in R^{3n}$ oszlopvektor, akkor a virtuális munka elve szerint bármely $\mathbf{v}$ virtuális elmozdulásra teljesül a

$$\mathbf{P}^T \mathbf{v} \leq 0 \tag{9.1.1}$$

egyenlőtlenség.

Itt olyan mechanikai rendszerekkel foglalkozunk, ahol a súrlódás elhanyagolható. A külön hivatkozás nélkül felhasznált fogalmak a [3], [22] könyvekben megtalálhatók.

A tömegpontrendszerek mozgásánál az egyik legegyszerűbb eset az, mikor a pontrendszer konzervatív erőtérben mozog, a kényszerek egyenlőség és egyenlőtlenség feltételekkel vannak megadva és csak a helykoordinátáktól függenek. Ebben az esetben az erőegyensúly jellemzéséhez a virtuális munka elve helyett kiindulhatunk a speciálisabb *Courtivron-elv*ből, amely szerint erőegyensúly esetén a $V : R^{3n} \to R$ potenciálfüggvénynek stacionárius pontja van az adott kényszerfeltételek mellett, azaz az alábbi nemlineáris optimalizálási feladatnak az $\mathbf{x}_0$ egyensúlyi pont *Karush-Kuhn-Tucker pontja*:

$$\begin{gathered} \min V(\mathbf{x}) \\ g_i(\mathbf{x}) \ \geq 0, \quad i = 1, \dots, m, \\ h_j(\mathbf{x}) = 0, \quad j = 1, \dots, p, \qquad \mathbf{x} \in R^{3n}, \end{gathered} \tag{9.1.2}$$

ahol $g_i, h_j \in C^2, \;\; i = 1, \dots, m; \;\; j = 1, \dots, p.$

Ha a pontrendszer nem konzervatív erőtérben mozog, vagy a kényszerfeltételek között nem véges alakok is szerepelnek, akkor a *Courtivron-elv*et nem tudjuk alkalmazni. Ebben az esetben az általánosan érvényes virtuális munka elvét kell közvetlenül felhasználni.

Az optimalizáláselméletből, illetve az analitikus mechanikával foglakozó munkákból ismert, hogy ha a (9.1.2) feladat valamely szélsőértékhelyén vagy nyeregpontjában egy regularitási feltétel teljesül, akkor ebben a pontban az optimalitás elsőrendű szükséges feltétele is teljesül, ami nem más mint a (3.9) összefüggések, azaz a virtuális munka elvének az alkalmazásával a mechanikai egyensúlyra kapott szükséges feltétel.

A virtuális munka elvének az alkalmazásához is szükség van regularitási feltételre, így ha a (9.1.2) feladatnál az optimalitási feltételek származtatásához is és a virtuális munka elvének az alkalmazásához is ugyanazt a regularitási feltételt használjuk - a Karush-Kuhn-Tucker regularitási feltételt -, akkor ebben a klasszikus esetben a két elv megegyezik.

9.2. Hiperbolikus vagy lineáris törtprogramozás

Martos B. 1960-ban figyelt fel arra, hogy az a feladatosztály, ahol két lineáris függvény hányadosának a szélsőértékeit kell meghatározni, sok lényeges tulajdonságban megegyezik a lineáris feladatokkal, és a világon elsőként oldotta meg a lineáris törtprogramozási, vagy az általa bevezetett terminológiát használva, a hiperbolikus programozási feladatot[6], megelőzve az ugyanezzel a problémával foglalkozó amerikai és német matematikusokat. Minderről bővebben lehet olvasni Martos (1975) könyvében és Rapcsák (2005) cikkében.

Az ismertetendő modellben a cél a termelés hatékonyságának növelése. A hatékonyság mérésére szokásos az egységnyi költségre eső árbevételt alkalmazni. Jelölje $c_1, \dots, c_n$ az egyes termékek egy egységének előállítási költségét, $p_1, ..., p_n$ az eladásukból származó árbevételt. Az összes költség és az összes árbevétel kiszámításánál ismét élünk a linearitási feltétellel. Az adott c_j, $j = 1, \dots, n$, értékekből képezzük a $\mathbf{c}$ vektort és a p_j, $j = 1, \dots, n$, értékekből a $\mathbf{p}$ vektort. Ekkor a termelés hatékonyságát – amelyet maximalizálni szeretnénk lineáris korlátozó feltételek mellett –

[6]Ezt a részt Komáromi (2002) jegyzete és Charnes-Cooper (1962) módszere (lásd [4]) alapján tárgyaljuk.

a

$$\frac{\mathbf{p}^T\mathbf{x}}{\mathbf{c}^T\mathbf{x}} = \frac{\sum_{j=1}^{n} p_j x_j}{\sum_{j=1}^{n} c_j x_j}$$

hányados fejezi ki.

A modell a következő:

$$\begin{gathered} \max \frac{\mathbf{p}^T\mathbf{x} + p_0}{\mathbf{c}^T\mathbf{x} + c_0} \\ \mathbf{A}\mathbf{x} \leq \mathbf{b}, \qquad \mathbf{x} \geq \mathbf{0}, \end{gathered} \tag{9.2.1}$$

ahol $\mathbf{A}$ $m \times n$-es mátrix, $\mathbf{c}, \mathbf{p}, \mathbf{x}, \mathbf{y} \in R^n$, $\mathbf{b} \in R^m$ és $p_0, c_0 \in R$. Ez a modell a hiperbolikus, vagy lineáris törtprogramozás körébe tartozik, amit a következőképpen alakítunk át lineáris optimalizálási feladattá. Vezessük be a $t = \dfrac{1}{\mathbf{c}^T\mathbf{x} + c_0}$ változót. Tegyük fel, hogy $\mathbf{c}^T\mathbf{x} + c_0 > 0$, $\mathbf{x} \in \{\mathbf{x} \in R^n \mid \mathbf{A}\mathbf{x} \leq \mathbf{b}, \mathbf{x} \geq 0\}$, és a megengedett tartomány kompakt. Legyen $\mathbf{y} = t\mathbf{x}$, így a következő lineáris optimalizálási feladathoz jutunk:

$$\begin{gathered} \max \mathbf{p}^T\mathbf{y} + p_0 t \\ \mathbf{A}\mathbf{y} - t\mathbf{b} \leq \mathbf{0}, \qquad \mathbf{c}^T\mathbf{y} + c_0 t = 1, \qquad \mathbf{y} \geq \mathbf{0}, \qquad t \geq 0. \end{gathered} \tag{9.2.2}$$

Megjegyezzük, hogy ha $(\mathbf{y}^T, t)^T$ megengedett megoldása a (9.2.2) feladatnak, akkor $t > 0$. Ha ugyanis $t = 0$ lenne, akkor létezne olyan $\mathbf{y} \neq \mathbf{0}$ vektor, amelyre $A\mathbf{y} \leq \mathbf{0}$ és $\mathbf{y} \geq \mathbf{0}$ teljesülne, ami ellentmondana a megengedett halmaz kompaktságának.

A két feladat ekvivalens a következő értelemben: ha $\mathbf{x}$ megoldja az (9.2.1) feladatot, akkor $t = \dfrac{1}{\mathbf{c}^T\mathbf{x} + c_0} > 0$ és $\mathbf{y} = t\mathbf{x}$ megoldja a (9.2.2) feladatot. Fordítva, ha $\mathbf{y}$ és t együttesen megoldja a (9.2.2) feladatot, akkor $t > 0$ és $\mathbf{x} = \dfrac{1}{t}\mathbf{y}$ megoldja az (9.2.1) feladatot. Megállapítható tehát, hogy a kompakt megengedett tartománnyal rendelkező hiperbolikus optimalizálási feladatok visszavezethetők olyan lineáris optimalizálási feladatokra, amelyekben eggyel több változó és eggyel több feltétel van.

9.2.1. Példa. *Tekintsük a következő optimalizálási feladatot*:

$$\min \frac{-2x_1 + x_2 + 2}{x_1 + 3x_2 + 4}$$

$$-x_1 + x_2 \leq 4,$$
$$2x_1 + x_2 \leq 14,$$

$$x_2 \leq 6, \qquad x_1 \geq 0, \qquad x_2 \geq 0.$$

Oldjuk meg ezt a feladatot Charnes és Cooper módszerével. A $(0,0)$ *pont megengedett és ebben a pontban, továbbá a megengedett megoldások teljes tartományában a célfüggvény nevezője pozitív. Az ekvivalens lineáris optimalizálási feladat a következő*:

$$\begin{aligned} \min\ -2y_1 + y_2 + 2t & \\ -y_1 + y_2 - 4t & \leq 0, \\ 2y_1 + y_2 - 14t & \leq 0, \\ y_2 - 6t & \leq 0, \\ y_1 + 3y_2 + 4t & = 1, \end{aligned}$$

$$y_1 \geq 0, \qquad y_2 \geq 0, \qquad t \geq 0.$$

A lineáris optimalizálási feladat optimális megoldása: $y_1 = 7/11, y_2 = 0$, $t = 1/11$. *Így az eredeti hiperbolikus optimalizálási feladat megoldása*:

$$x_1 = y_1/t = 7 \quad \textit{és} \quad x_2 = y_2/t = 0.$$

9.3. Kvadratikus programozás

Tekintsünk egy olyan NLO feladatot, amelynek a célfüggvénye $x_1^{k_1}x_2^{k_2}\dots x_n^{k_n}$ alakú kifejezések összegeként áll elő. Az $x_1^{k_1}x_2^{k_2}\dots x_n^{k_n}$ kifejezés foka $k_1+k_2+\dots k_n$. Tehát, az $x_1^2x_2$ kifejezés foka 3, az x_1x_2 kifejezésé pedig 2. Egy olyan NLO feladatot, amelynek lineárisak a feltételei, a célfüggvénye pedig 0, 1 vagy 2 fokú, $x_1^{k_1}x_2^{k_2}\dots x_n^{k_n}$ alakú kifejezések összege, *kvadratikus optimalizálási feladatnak* nevezünk. (Itt feltételezzük, hogy $k_1, k_2, \dots, k_n$ egész.)

Számos módszer alkalmazható kvadratikus programozási feladatok megoldására (lásd Bazaraa és Shetty (1993, 11. fejezet)). Mi itt a kvadratikus optimalizálási portfólió kiválasztására történő alkalmazását mutatjuk be Winston (2003) alapján.

9.3.1. A kvadratikus optimalizálás és a portfólió kiválasztás

Tekintsünk egy befektetőt, aki adott mennyiségű pénzét különböző befektetésekben helyezheti el. Általában azt tételezzük fel, hogy a befektető maximalizálni akarja a befektetései (portfóliója) utáni várható hozamot, miközben azt is biztosítani akarja, hogy a portfólió kockázata kicsi legyen, ahol a kockázatot a portfólió hozamának szórásnégyzetével becsüljük. Sajnos, a magas várható hozamú részvények esetében általában a hozam szórásnégyzete (kockázata) is magas. Ezért a portfólió kiválasztási feladat egyik gyakori megközelítése az, hogy választunk egy elfogadható, minimális várható hozam értéket, majd ezt a várható hozamszintet teljesíteni tudó portfóliók közül megkeressük a minimális szórásnégyzetűt. Például, a befektető keresheti azt a minimális szórásnégyzetű portfóliót, amelynek a várható hozama 12%. Az elfogadható minimális várható hozam szintjének változtatásával a befektető különböző portfóliókat kaphat, és azokat összehasonlíthatja.

Ezekkel a meggondolásokkal a portfólió kiválasztási feladatot kvadratikus optimalizálási feladatra vezethetjük vissza. Ehhez szükségünk van azokra a szabályokra, amelyek a valószínűségi változók összege várható értékének és szórásnégyzetének meghatározására vonatkoznak.

Tudjuk, hogy adott $S_1, S_2, \ldots, S_n$ valószínűségi változók, valamint a, b és k konstansok esetén

$$E(S_1 + S_2 + \ldots + S_n) = E(S_1) + E(S_2) + \ldots + E(S_n), \tag{9.3.1}$$

$$\mathrm{var}\,(S_1 + S_2 + \ldots + S_n) = \mathrm{var}\,S_1 + \mathrm{var}\,S_2 + \ldots \mathrm{var}\,S_n + \sum_{i \neq j} \mathrm{cov}\;(S_i, S_j), \tag{9.3.2}$$

$$E(kS_i) = kE(S_i), \tag{9.3.3}$$

$$\mathrm{var}\,(kS_i) = k^2 \mathrm{var}\,S_i, \tag{9.3.4}$$

$$\mathrm{cov}\,(aS_i, bS_j) = ab\;\mathrm{cov}\,(S_i, S_j), \tag{9.3.5}$$

ahol E a várható értéket, *var* a szórásnégyzetet és *cov* két valószínűségi változó kovarianciáját jelenti.

A következő példán megmutatjuk, hogyan lehet a portfólió kiválasztási feladatot kvadratikus optimalizálási feladatként felírni.

9.3.1. Példa. *Tegyük fel, hogy 1000 dollárunk van, és azt három részvénybe fektethetjük. Legyen S_i az a valószínűségi változó, amely az i-edik részvénybe fektetett egy dollár utáni éves hozamot képviseli. Tehát ha S_i értéke* 0.12, *akkor az év elején az i-edik részvénybe fektetett 1 dollár az év végén* 1.12 *dollárt ér. A következő információkkal rendelkezünk:* $E(S_1) = 0.14$, $E(S_2) = 0.11, E(S_3) = 0.10, var\,S_1 = 0.20, var\,S_2 = 0.08, var\,S_3 = 0.18$, $cov\,(S_1, S_2) = 0.05, cov\,(S_1, S_3) = 0.02, cov\,(S_2, S_3) = 0.03$. *Állítsunk fel egy kvadratikus optimalizálási feladatot, amelynek segítségével meghatározhatjuk azt a portfóliót, ahol az éves várható hozam legalább* 12%, *és az ilyen portfóliók közül minimális az éves hozam szórásnégyzete!*

Megoldás. *Jelölje x_j, hogy hány dollárt fektetünk a j-edik részvénybe, $j = 1, 2, 3$. Ekkor a portfólió éves hozama $(x_1S_1 + x_2S_2 + x_3S_3)/1000$, a portfólió éves hozamának várható értéke pedig ((9.3.1) és (9.3.3) alapján)*

$$\frac{x_iE(S_1) + x_2E(S_2) + x_3E(S_3)}{1000}.$$

Annak biztosítására, hogy a portfólió várható hozama legalább 12% *legyen, a következő feltételt kell megfogalmazni:*

$$\frac{0.14x_1 + 0.11x_2 + 0.10x_3}{1000} \geq 0.12,$$

azaz

$$0.14x_1 + 0.11x_2 + 0.10x_3 \geq 0.12(1000) = 120.$$

Természetesen, az $x_1 + x_2 + x_3 = 1000$ *feltételt is fel kell írni. Tételezzük fel, hogy csak nem-negatív összeget lehet részvénybe fektetni (azaz a rövidre eladás nem megengedett), ezért az* $x_1, x_2, x_3 \geq 0$ *feltételeket is csatolni kell. Célunk egyszerűen a portfólió éves hozama szórásnégyzetének a minimalizálása.*

Tudjuk (9.3.2)*-ből, hogy a portfólió hozamának szórásnégyzete a következő:*

$$\begin{aligned} var(x_1S_1 + x_2S_2 + S_3) &= var(x_1S_1) + var(x_2S_2) + var(x_3S_3) \\ &+ 2cov(x_1S_1, x_2S_2) + 2cov(x_1S_1, x_3S_3) \\ &+ 2cov(x_2S_2, x_3S_3) \\ &= x_1^2 var\, S_1 + x_2^2 var\, S_2 + x_3^2 var\, S_3 + 2x_1x_2 cov(S_1, S_2) \\ &+ 2x_1x_3 cov(S_1, S_3) + 2x_2x_3 cov(S_2, S_3) \\ &(a\ (9.3.4)\ \textit{és}\ (9.3.5)\ \textit{egyenletek alapján}) \\ &= 0.20x_1^2 + 0.08x_2^2 + 0.18x_3^2 + 0.10x_1x_2 \\ &+ 0.04x_1x_3 + 0.06x_2x_3. \end{aligned}$$

Vegyük észre, hogy a portfólió szórásnégyzetére vonatkozó utolsó kifejezés másodfokú kifejezések összege. Tehát, olyan NLO feladatunk van, ahol a feltételek lineárisak, a célfüggvény pedig másodfokú kifejezesek összegéből áll. A legalább 12% *várható hozamú portfóliók közül a legkisebb szórásnégyzetű portfóliót a következő kvadratikus* *optimalizálási feladat megoldásával kaphatjuk meg:*

$$\min\ f(x_1,x_2,x_3) = 0.20x_1^2 + 0.08x_2^2 + 0.18x_3^2 + 0.10x_1x_2 + 0.04x_1x_3 + 0.06x_2x_3$$
$$0.14x_1 + 0.11x_2 + 0.10x_3 \geq 120,$$
$$x_1 + x_2 + x_3 = 1000,$$
$$x_1, x_2, x_3 \geq 0. \tag{9.3.6}$$

Mivel a célfüggvény konvex, a feltételek pedig lineárisak, konvex optimalizálási feladatot kapunk. Optimalizálási programcsomag segítségével megkaphatjuk a (9.3.6) *feladat optimális megoldását, ami* $f(\mathbf{x}^*) = 75238$ *(dollár)*^{2}, $x_1^* = 380.95\$$, $x_2^* = 476.19\$$ $x_3^* = 142.86\$$. *Mivel, az első feltétel egyenlőségként teljesül, az optimális portfólió várható hozama pontosan* 12%. *A portfólió éves hozamának szórása dollárban* $(75238)^{1/2} = 274.30\$$.

Egy befektetési portfólió hozamát gyakran közelítik normális eloszlással. Tudjuk, hogy a normális eloszlás nagyjából 95%*-os valószínűséggel vesz fel a várható értéktől legfeljebb két szórásnyi eltérésű értékeket. Tehát* 95%*-ig biztosak lehetünk abban, hogy az optimális portfólió éves dollárhozama* $-428.60\$ = 120\$ - 2\,(274.30\$)$ *és* $665.60\$ = 120\$ + 2(274.30\$)$ *között lesz.*

Megjegyzések

1. A kvadratikus optimalizálás optimális portfóliók meghatározására történő alkalmazásának ötlete Markowitz (1959) dolgozatából származik, és részét képezi azoknak az eredményeknek, amelyek alapján Markowitz Közgazdasági Nobel-díjat kapott.

2. Megvizsgálható (Winston, 2003), hogy miként lehet aktuális adatok felhasználásával megbecsülni egy befektetés várható hozamát és szórásnégyzetét, továbbá két különböző befektetés esetén a hozamok közötti kovarianciát.

3. A valóságban tranzakciós költségek is fellépnek, amikor befektetéseket veszünk vagy adunk el. Megvizsgálható (Winston, 2003), hogyan módosítják a tranzakciós költségek a portfólió optimalizálási modelleket.

9.4. Entrópia optimalizálás (EO)

Az entrópia optimalizálás primál feladata:

$$\begin{aligned} \min\ \mathbf{c}^T\mathbf{x} + \sum_{i=1}^{n} x_i \log x_i \\ \mathbf{A}\mathbf{x} = \mathbf{b}, \\ \mathbf{x} > \mathbf{0}, \end{aligned} \tag{PEO}$$

ahol $\mathbf{A}$ $m \times n$-es mátrix, $\mathbf{b} \in R^m$ és $\mathbf{c} \in R^n$. A PEO konvex optimalizálási feladat.

Legyen $(\log \mathbf{x})^T = (\log x_1, \ldots, \log x_n)$. Így a PEO célfüggvénye $\mathbf{c}^T\mathbf{x} + \mathbf{x}^T \log \mathbf{x}$, $\mathbf{x} > 0$, alakban írható.

A 6. Fejezetben ismertetett módon vezessük le az PEO Lagrange-duál feladatát. Ha az $\mathbf{x} > \mathbf{0}$ nemnegativitási feltételt

$$\mathbf{x} \in A = \{\mathbf{x} \in R^n \mid \mathbf{x} > \mathbf{0}\}$$

alakban írjuk, akkor a PEO feladat *Lagrange*-duálja:

$$\max_{\boldsymbol{\mu} \in R^m} \psi(\boldsymbol{\mu}),$$

ahol (DEO)

$$\psi(\boldsymbol{\mu}) = \inf_{\mathbf{x} \in A} L(\mathbf{x}, \boldsymbol{\mu}),$$

és a PEO Lagrange függvénye:

$$L(\mathbf{x}, \boldsymbol{\mu}) = \mathbf{c}^T\mathbf{x} + \mathbf{x}^T \log \mathbf{x} - \boldsymbol{\mu}^T(\mathbf{A}\mathbf{x} - \mathbf{b}), \qquad \boldsymbol{\mu} \in R^m, \qquad \mathbf{x} \in A.$$

A Lagrange függvény $\mathbf{x}$ változó szerinti gradiense

$$\nabla_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\mu}) = \mathbf{c} + \mathbf{e} + \log \mathbf{x} - \mathbf{A}^T\boldsymbol{\mu},$$

ahol $\mathbf{e}^T = (1, \ldots, 1)$.

A Lagrange függvény rögzített $\boldsymbol{\mu}$ mellett konvex az $\mathbf{x}$ változóban, ezért akkor veszi fel a globális minimumát egy $\mathbf{x}^*$ pontban, ha a gradiens vektor nulla, azaz létezik $\mathbf{x}^* \in A$, amelyre

$$\mathbf{c} + \mathbf{e} + \log \mathbf{x}^* - \mathbf{A}^T\boldsymbol{\mu} = \mathbf{0}.$$

A $\mathbf{c} = -\mathbf{e} - \log \mathbf{x}^* + \mathbf{A}^T\boldsymbol{\mu}$ kifejezést visszahelyettesítve a Lagrange függvénybe azt kapjuk, hogy

$$L(\mathbf{x}^*, \boldsymbol{\mu}) = -\mathbf{e}^T\mathbf{x}^* - \mathbf{x}^{*T}\log \mathbf{x}^* + \boldsymbol{\mu}^T\mathbf{A}\mathbf{x}^* + \mathbf{x}^{*T}\log \mathbf{x}^* - \boldsymbol{\mu}^T(\mathbf{A}\mathbf{x}^* - \mathbf{b}) = \mathbf{b}^T\boldsymbol{\mu} - \mathbf{e}^T\mathbf{x}^*,$$
$$\boldsymbol{\mu} \in R^m.$$

Mivel

$$\log \mathbf{x}^* = -\mathbf{c} - \mathbf{e} + \mathbf{A}^T\boldsymbol{\mu},$$

ezért

$$x_i^* = e^{\mathbf{a}_i^T\boldsymbol{\mu} - c_i - 1}, \qquad i = 1, \ldots, n.$$

Az entrópia optimalizálás duál feladata:

$$\max\ \mathbf{b}^T\boldsymbol{\mu} - \sum_{i=1}^{n} e^{\mathbf{a}_i^T\boldsymbol{\mu} - c_i - 1}, \qquad \boldsymbol{\mu} \in R^m, \tag{DEO}$$

ami konvex feltétel nélküli optimalizálási feladat.

Dualitási eredmények

Az entrópia optimalizálási feladat itt szereplő dualitási tételei Kas és Klafszky [31] cikkében találhatók.

9.4.1. Lemma (Gyenge dualitás). *Ha* $\mathbf{x}$ *a PEO feladat megengedett megoldása és* $\boldsymbol{\mu}$ *a* DEO *feladat megengedett megoldása, akkor*

$$\mathbf{c}^T\mathbf{x} + \sum_{i=1}^{n} x_i \log x_i \geq \mathbf{b}^T\boldsymbol{\mu} - \sum_{i=1}^{n} e^{\mathbf{a}_i^T\boldsymbol{\mu} - c_i - 1},$$

és egyenlőség akkor és csak akkor teljesül, ha

$$x_i = e^{\mathbf{a}_i^T\boldsymbol{\mu} - c_i - 1}, \qquad i = 1, \ldots, n. \tag{9.4.1}$$

A PEO feladat megengedett megoldásainak halmazát jelölje $\mathcal{P}$.

9.4.1. Következmény. *Ha* (9.4.1) *fennáll egy adott* $\mathbf{x} \in \mathcal{P}$ és $\boldsymbol{\mu} \in R^m$ *esetén, akkor mindkettő optimális megoldás és a két optimum érték megegyezik.*

Azt mondjuk, hogy a PEO kielégíti a *Slater*-regularitási feltételt, ha létezik olyan $\hat{\mathbf{x}} > \mathbf{0}$ vektor, amelyre $\mathbf{A}\hat{\mathbf{x}} = \mathbf{b}$.

9.4.1. Dualitási tétel. *Tegyük fel, hogy mind a PEO, mind a DEO feladatnak van megengedett megoldása.*

Akkor a PEO feladatnak van optimális megoldása, és a primál optimum érték és a duál optimum érték szuprémuma megegyezik.

Akkor és csak akkor létezik $\boldsymbol{\mu}^* \in R^m$ *duál optimális megoldás, ha a PEO kielégíti a Slater-regularitási feltételt. A primál és a duál optimum érték megegyezik.*

9.5. Geometriai optimalizálás[7] (GO)

Legyenek az $\mathbf{a}_i \in R^m$, $i = 1, \ldots, n$, $\mathbf{b} \in R^m$ és $\mathbf{c} \in R^m$ vektorok adottak és legyen az $I = \{1, \ldots, n\}$ index halmaz az $I_1, I_2, \ldots, I_r$ diszjunkt index halmazokra bontva.

A geometriai optimalizálás primál feladata:

$$\max \mathbf{b}^T\mathbf{y}$$
$$g_k(\mathbf{y}) = \sum_{i \in I_k} e^{\mathbf{a}_i^T\mathbf{y} - c_i} \leq 1, \qquad k = 1, \ldots, r, \qquad \mathbf{y} \in R^m. \tag{PGO}$$

A PGO konvex optimalizálási feladat, ami az $I_1 = \{1\}$, $I_2 = \{2\}, \ldots, I_n = \{n\}$, esetben a lineáris optimalizálási feladatot adja.

A geometriai optimalizálás alkalmazásaiban a PGO feltételi függvények gyakran

$$\sum_{i \in I_k} \alpha_i \prod_{j=1}^{m} \tau_j^{a_{ij}}, \qquad k = 1, \ldots, r,$$

pozinomiális alakban jelennek meg, ahol az $\alpha_i > 0$, $i \in I_k$, $k = 1, \ldots, r$, értékek adottak és a változók csak pozitív értékeket vesznek fel, azaz $\tau_j > 0$, $j = 1, \ldots, m$.

[7] A geometriai optimalizálást Klafszky E. (1973) munkája alapján tárgyaljuk.

A pozinomiális alak előnye, hogy a $\tau_j = e^{y_j}$, $j = 1, \dots, m$, egy-egy értelmű transzformáció konvex függvényeket eredményez az $\mathbf{y}$ változóban:

$$\sum_{i \in I_k} \alpha_i \prod_{j=1}^{m} (e^{y_j})^{a_{ij}} = \sum_{i \in I_k} \alpha_i \, e^{\sum\limits_{j=1}^{m} a_{ij}\, y_j}, \qquad k = 1, \dots, r, \qquad \mathbf{y} \in R^m.$$

A g_k, $k = 1, \dots, r$, függvényekről a logaritmikus konvexitást is be lehet látni a *Hölder*-egyenlőtlenség segítségével.[8]

A geometriai optimalizálás duál feladata:

$$\begin{aligned} \min\ & \mathbf{c}^T \mathbf{x} + \psi(\mathbf{x}) \\ & \mathbf{A}\mathbf{x} = \mathbf{b}, \\ & \mathbf{x} \geq \mathbf{0}, \end{aligned} \tag{DGO}$$

ahol

$$\psi(\mathbf{x}) = \sum_{k=1}^{r} \left(\sum_{i \in I_k} x_i \log x_i - \left(\sum_{i \in I_k} x_i \right) \log \left(\sum_{i \in I_k} x_i \right) \right).$$

A DGO konvex optimalizálási feladat, ami $r = 1$ és $\sum_{i \in I} x_i = 1$ esetén a primál entrópia optimalizálási feladatot adja. A duál feladat levezetése Duffin, Peterson és Zener [10] könyvében megtalálható.

Dualitási eredmények

A PGO és a DGO feladatok megengedett megoldásainak halmazát jelölje $\mathcal{P}$ és $\mathcal{D}$. Azt mondjuk, hogy a PGO kielégíti a *Slater*-regularitási feltételt, ha létezik olyan $\mathbf{y} \in R^m$ vektor, amelyre az egyenlőtlenség feltételek szigorú egyenlőtlenséggel teljesülnek.

9.5.1. Lemma (Gyenge dualitás). *Ha* $\mathbf{y} \in \mathcal{P}$ és $\mathbf{x} \in \mathcal{D}$, *akkor*

$$\mathbf{b}^T \mathbf{y} \leq \mathbf{c}^T \mathbf{x} + \psi(\mathbf{x}),$$

---

[8] A Hölder-egyenlőtlenség szerint

$$\sum_{i=1}^{n} \alpha_i^{\lambda} \beta_i^{1-\lambda} \leq \left(\sum_{i=1}^{n} \alpha_i \right)^{\lambda} \left(\sum_{i=1}^{n} \beta_i \right)^{1-\lambda},$$

ahol $(\alpha_1, \alpha_2, \dots, \alpha_n) \geq 0, (\beta_1, \beta_2, \dots, \beta_n) \geq 0$ és $0 \leq \lambda \leq 1$.

és egyenlőség akkor és csak akkor teljesül, ha

$$x_j = e^{\mathbf{a}_j^T \mathbf{y} - c_j} \sum_{i \in I_k} x_i, \qquad j \in I_k, \qquad k = 1, \ldots, r. \tag{9.5.1}$$

9.5.1. Következmény. *Ha* (9.5.1) *teljesül valamely* $\mathbf{y} \in \mathcal{P}$ *és* $\mathbf{x} \in \mathcal{D}$ *esetén, akkor mindkettő optimális megoldás és a két optimum érték megegyezik.*

9.5.1. Dualitási tétel.

1. *Ha a PGO kielégíti a Slater-regularitási feltételt és véges optimum értéke van, akkor a duál feladatnak létezik optimális megoldása, és a két optimum érték megegyezik.*
2. *Ha mind a PGO, mind a DGO feladatnak van megengedett megoldása, akkor a primál célfüggvény szuprémuma megegyezik a duál célfüggvény infimumával.*
3. *Ha a PGO feladatnak van megengedett megoldása és véges szuprémuma, akkor a duál feladatnak is van megengedett megoldása és a primál szuprémum megegyezik a duál infimummal.*

9.5.1. Feladat. *Vezessük le a geometriai optimalizálási feladat DGO duálját a primál PGO feladatból a Lagrange-dualitást felhasználva.*

A duál levezetéséhez azt kell felhasználni, hogy

$$\psi(\mathbf{x}) = \sum_{k=1}^{r} \log \frac{\prod_{i \in I_k} x_i^{x_i}}{\left(\sum_{i \in I_k} x_i\right)^{\left(\sum_{i \in I_k} x_i\right)}},$$

és ezt a függvényt az általánosított számtani-mértani közepek közötti egyenlőtlenség felhasználásával becsülhetjük.[9]

[9]Az általánosított számtani-mértani egyenlőtlenség szerint

$$\left(\frac{\sum_{i=1}^{n} \alpha_i}{\sum_{i=1}^{n} \beta_i}\right)^{\sum_{i=1}^{n} \beta_i} \geq \prod_{i=1}^{n} \left(\frac{\alpha_i}{\beta_i}\right)^{\beta_i},$$

ahol $\alpha = (\alpha_1 \ldots, \alpha_n) \geq 0$, $\beta = (\beta_1 \ldots, \beta_n) > 0$, és egyenlőség akkor és csak akkor teljesül, ha $\alpha = \lambda\beta$ valamely nemnegatív λ esetén.

9.6. Lineáris szemidefinit optimalizálás[10] (LSDO)

Legyenek $\mathbf{A}_0, \mathbf{A}_1 \ldots, \mathbf{A}_n \in R^{m\times m}$ szimmetrikus mátrixok, $\mathbf{c} \in R^n$ adott vektor és $\mathbf{x} \in R^n$ az ismeretlen vektor.

A lineáris szemidefinit optimalizálás primál feladata:

$$\min \mathbf{c}^T\mathbf{x}$$
$$\mathbf{A}_0 + \sum_{k=1}^{n} x_k\mathbf{A}_k \preceq 0, \qquad \mathbf{x} \in R^n, \tag{PLSDO}$$

ahol a $\preceq 0$ szimbólum azt jelenti, hogy a bal oldali mátrixnak negatív szemidefinitnek kell lennie. A PLSDO primál feladat konvex optimalizálási feladat, mivel negatív szemidefinit mátrixok bármely konvex kombinációja is negatív szemidefinit.

A lineáris szemidefinit optimalizálás duál feladata:

$$\max tr(\mathbf{A}_0\mathbf{W})$$
$$c_k + tr(\mathbf{A}_k\mathbf{W}) = 0, \qquad k = 1, \ldots, n, \tag{DLSDO}$$
$$\mathbf{W} \succeq 0,$$

ahol $\mathbf{W} \in R^{m\times m}$ a változók mátrixa, a $\succeq$ szimbólum azt jelenti, hogy a bal oldali mátrixnak pozitív szemidefinitnek kell lennie és a tr leképezés, a mátrix nyoma, adott mátrix esetén a főátlóban levő elemek összegét adja. A DLSDO feladat konvex optimalizálási feladat, ami abból következik, hogy a mátrix nyoma a mátrix lineáris függvénye, és pozitív szemidefinit mátrixok konvex kombinációja is pozitív szemidefinit.

9.6.1. Dualitási tétel. *Ha $\mathbf{x} \in R^n$ primál megengedett megoldás, $\mathbf{W} \in R^{m\times m}$ duál megengedett megoldás, akkor*

$$\mathbf{c}^T\mathbf{x} \geq tr(\mathbf{A}_0\mathbf{W}), \tag{9.6.1}$$

[10] A lineáris szemidefinit optimalizálást Shapiro (2001) cikke alapján tárgyaljuk.

és egyenlőség akkor és csak akkor áll fenn, ha

$$\left(\mathbf{A}_0 + \sum_{k=1}^{n} \mathbf{x}_k \mathbf{A}_k\right) \mathbf{W} = 0. \tag{9.6.2}$$

A lineáris szemidefinit optimalizálási feladat nemlinearitása miatt erős dualitás csak bizonyos regularitási feltétel, pl. a *Slater*-regularitási feltétel esetén teljesül. A PLSDO akkor *Slater*-reguláris, ha létezik olyan $\mathbf{x} \in R^n$, amelyre az $\left(\mathbf{A}_0 + \sum_{k=1}^{n} \mathbf{x}_k \mathbf{A}_k\right)$ mátrix pozitív definit. A DLSDO akkor *Slater*-reguláris, ha létezik olyan $\mathbf{W}$ $m \times m$-es szimmetrikus pozitív definit mátrix, amelyre $c_k + tr(\mathbf{A}_k \mathbf{W}) = 0$, $k = 1, \ldots, n$. A definíciókból következik, hogy lineáris szemidefinit optimalizálás esetén a *Slater*-regularitási feltételek megegyeznek a belsőpont-feltételekkel.

Irodalomjegyzék

[1] Arrow, K.J. and Enthoven, A.C., Quasi-concave programming, *Econometrica* 29 (1961) 779-800.

[2] Avriel, M., Diewert, W.E., Schaible, S. and Zang, I., Generalized concavity, *Plenum Press*, 1988.

[3] Banach, S., Mechanics, *Nauki*, Warszawa, Wroclaw, 1951.

[4] Bazaraa, M.S. and Shetty, C. M., Foundations of optimization, *Springer-Verlag*, Berlin, Heidelberg, New York, 1976.

[5] Bazaraa, M.S. and Shetty, C. M., Nonlinear programming, theory and algorithms, *John Wiley and Sons*, New York, 1979.

[6] Bennet, A. A., Newton's method in general analysis, *Proceedings of National Academy of Sciences*, USA 2 (10) (1916) 592-598.

[7] Crouzeix, J.P., On second-order conditions for quasiconvexity, *Mathematical Programming* 18 (1980) 349-352.

[8] Dancs, I. és Puskás Cs., Vektorterek, *Aula Kiadó*, 2001.

[9] Dancs, I., Magyarkúti Gy., Medvegyev P. és Tallos P., Bevezetés a matematikai analízisbe, *Aula Kiadó*, 2003.

[10] Deák, I., Bevezetés a stochasztikus programozásba, *Operációkutatás* 3, *Aula Kiadó*, Budapest, 2003.

[11] Duffin, R.J., Peterson, E.L and Zener, C., Geometric programming, *John Wiley* & *Sons*, New York, 1967.

[12] Egorychev, G.P., A solution of van der Waerden's permanent problem, *Dokl. Akad. Nauk SSSP* 258 (1981) 1041-44. (in Russian) *Translated in Soviet Math. Dokl.* 23 (1981) 619-622.

[13] Falikman, D.I., A proof of van der Waerden's conjecture on a permanent of a doubly stochastic matrix, *Mat. Zametki* 29 (1981) 931-939.

[14] Farkas, J., Theorie der einfachen Ungleichungen, *Journal für die Reine und Angewandte Mathematik* 124 (1901) 1-27.

[15] Farkas, J., Beitrage zu den Grundlagen der analytischen Mechanik, *Journal für die Reine und Angewandte Mathematik* 131 (1906) 165-201.

[16] Fenchel, W., Convex cones, sets and functions, Mimeographed lecture notes, *Princeton University*, Princeton, New Jersey, 1951.

[17] Fenchel, W., Über konvexe Funktionen mit voreschriebenen Niveaumannigfaltigkeiten, *Math. Z.* 63 (1956) 496-506.

[18] Fiacco, A.V. and McCormick, G.P., Nonlinear programming, sequential unconstrained minimization techniques, *John Wiley and Sons*, New York, 1968.

[19] Fine, H., On Newton's method of approximation, *Proceedings of National Academy of Sciences*, USA 2 (9) (1916) 546-552.

[20] Forgó, F., Nonconvex programming, *Akadémiai Kiadó*, Budapest, 1988.

[21] Fourier, J., Mémoire sur le statique, *Journal de l'École Polytechnique* 5 (1798).

[22] Gantmacher, F., Lectures in analytical mechanics, *Mir Publishers*, Moscow, 1970.

[23] Giannessi, F., Theorems of the alternative and optimality conditions, *Journal of Optimization Theory and Applications* 42 (1984) 331-365.

[24] Halmos E. és Rapcsák T., Statikailag határozatlan rácsos tartók minimális súlyra történő méretezése, *Alkalmazott Matematikai Lapok* 3 (1977) 171-183.

[25] Halmos, E. és Rapcsák, T., Minimum weight design of the statically indeterminate trusses, *Mathematical Programming Study* 9 (1978) 109-111.

[26] Hunyadi L. és Vita L., Statisztika közgazdászoknak, *Központi Statisztikai Hivatal*, Budapest, 2002.

[27] Kantorovich, L. V., On Newton's method for functional equations, *Doklady AN SSSR* 59 (7) (1948) 1237-1240.

[28] Kantorovich, L. V. and Akilov, G. P., Functional analysis in normed spaces, *Fizmatgiz*, Moscow, 1959.

[29] Kantorovich, L. V. and Akilov, G. P., Functional analysis, *Nauka*, Moscow, 1977.

[30] Karush, W., Minima of function of several variables with inequalities as side conditions, Master's Thesis, *Department of Mathematics*, *University of Chicago*, 1939.

[31] Kas, P. and Klafszky, E., On the duality of the mixed entropy programming, *Optimization* 27 (1993) 253-258.

[32] Klafszky, E., Geometriai programozás és néhány alkalmazása, Kandidátusi értekezés, MTA Számítástechnikai és Automatizálási Kutató Intézet, Tanulmányok 8/1973.

[33] Komáromi, É., Lineáris programozás, *Operációkutatás* 2, *Aula Kiadó*, Budapest, 2002.

[34] Kuhn, H.W. and Tucker, A. W., Nonlinear programming, in: Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, *University of California Press*, Berkeley, California, 1950.

[35] Lagrange, J.L., Mécanique analytique 1-11, *Paris*, 1788.

[36] Luenberger, D.G., Introduction to linear and nonlinear programming, *Addison-Wesley Publishing Company Inc.*, 1973.

[37] Mangasarian, O.L., Pseudo-convex functions, *Society for Industrial and Applied Mathematics Journal on Control* 3 (1965) 281-290.

[38] Mangasarian, O.L., Nonlinear programming, *McGraw-Hill Book Company*, 1969.

[39] Martos, B., Nonlinear programming: theory and methods, North-Holland, Amsterdam; *Akadémiai Kiadó*, Budapest, 1975.

[40] Mayer, J., Stochastic linear programming algorithms, *Gordon and Breach*, 1998.

[41] Mysovskikh, L. P., On convergence on L. V. Kantorovich's method for functional equations and its applications, *Doklady AN SSSR* 70 (4) 565-568.

[42] Ortega, J. M. and Rheinboldt, W. C., Iterative solution of nonlinear equations in several variables, *Academic Press*, New York-London, 1970.

[43] Ostrogradsky, M., Mémoire sur les déplacement instantanés des systèmes assujettis á des conditions variables, Mémoires de l'Académie Impériale des Sciences de Saint-Petersbourg, *Sixieme Série* 1 (1838) 565-600.

[44] Ostrowski, A. M., Solution of equations and systems of equations, *Academic Press*, Basel, 1960.

[45] Pintér, J., Global optimization in action; Continuous and Lipschitz optimization: algorithms, implementations and applications, *Kluwer Academic Publishers*, 1996.

[46] Polyak, B. T., Newton's method and its use in optimization, European *Journal of Operational Research* 181 (2007) 1086-1096.

[47] Prékopa A., Az optimalizáláselmélet kialakulásának történetéről, *Alkalmazott Matematikai Lapok* 4 (1978) 165-191.

[48] Prékopa, A., On the development of optimization theory, *The American Mathematical Monthly* 87 (1980) 527-542.

[49] Prékopa, A., Stochastic programming, *Kluwer Academic Publishers*, 1995.

[50] Prékopa A., Rapcsák T. és Zsuffa I., Egy új módszer sorbakapcsolt tározórendszerek tervezésére, *Alkalmazott Matematikai Lapok*, 2 (1976) 189-201.

[51] Prékopa, A., Rapcsák, T. and Zsuffa, I., Serially linked reservoir system design using stochastic programming, *Water Resources Research* 14 (1978) 672-678.

[52] Rapcsák T., Autóbuszok erőátviteli láncának optimális méretezése mechanikus sebességváltó esetén, *Alkalmazott Matematikai Lapok* 4 (1978) 229-243.

[53] Rapcsák T., Lineáris programozási modell egy tereprendezési feladat megoldására, *Alkalmazott Matematikai Lapok* 7 (1981) 99-105.

[54] Rapcsák, T., A linear programming model for the optimal levelling of an irrigation surface, *European Journal of Operational Research* 13 (1983) 369-373.

[55] Rapcsák, T., The optimal power transmission of buses in case of a mechanical speed gear, *Advances in Management Studies* 2 (1983) 1-22.

[56] Rapcsák, T. and Thang, T.T., On nonlinear coordinate representations of smooth optimization problems, *Journal of Optimization Theory and Applications* 86 (1995) 459-489.

[57] Rapcsák, T., Smooth nonlinear optimization in R^n, *Kluwer Academic Publishers*, 1997.

[58] Rapcsák, T., Variable metric methods along geodesics, in: New trends in mathematical programming, eds.: F. Giannessi, S. Komlósi and T. Rapcsák, *Kluwer Academic Publishers* (1998) 257-275.

[59] Rapcsák, T., Mechanikai egyensúly és egyensúlyi rendszerek, Új utak a magyar operációkutatásban; In memoriam Farkas Gyula, szerk.: Komlósi Sándor és Szántai T., *Dialóg Campus Kiadó* (1999) 32-42.

[60] Rapcsák, T., Martos Béla optimalizáláselméleti munkásságának méltatása az Egerváry-emlékplakett átadása alkalmából, *Alkalmazott Matematikai Lapok* 23 (2006) 1-4.

[61] Roberts, A.W. and Varberg, D.E., Convex functions, *Academic Press*, New York and London, 1973.

[62] Rockafellar, R.T., Convex analysis, *Princeton University Press*, Princeton, New Yersey, 1970.

[63] Roos, C., Terlaky, T. and Vial, J.-Ph., Theory and algorithms for linear optimization. An interior point approach, *Wiley*, Chichester, UK, 1997.

[64] Shapiro, A., Semidefinite programminng: optimality conditions and stability, Encyclopedia of Optimization, eds: C.A. Floudas and P.M. Pardalos, *Kluwer Academic Publishers* 5 (2001) 138-143.

[65] Spivak, M., A comprehensive introduction to differential geometry I-V, *Publish or Perish*, Berkley, 1979.

[66] Stahl J., Optimumszámítás, Budapesti Közgazdaságtudományi Egyetem, *Aula Kiadó*, 1992.

[67] Udriste, C., Convex functions and optimization methods on Riemannian manifolds, *Kluwer Academic Publishers*, 1994.

[68] Ujvári, M., A szemidefinit programozás alkalmazásai a kombinatorikus optimalizálásban, Egyetemi jegyzet, *ELTE Eötvös Kiadó*, 2001.

[69] van der Waerden, B.L., Aufgabe 45, *Iber. Deutsch. Math.-Verein* 35 (1926) 117.

[70] Winston, W.L., Operációkutatás, Módszerek és alkalmazások, *Aula*, 2003.

[71] Zalai E., Matematikai közgazdaságtan, *KJK KERSZÖV*, Budapest, 2000.