Alkalmazott matematikai lapok

1990-91/1-2

A MAGYAR TUDOMÁNYOS AKADÉMIA
MATEMATIKAI ÉS FIZIKAI TUDOMÁNYOK
OSZTÁLYÁNAK KÖZLEMÉNYEI

15. KÖTET

ALKALMAZOTT MATEMATIKAI LAPOK

A MAGYAR TUDOMÁNYOS AKADÉMIA MATEMATIKAI ÉS FIZIKAI TUDOMÁNYOK OSZTÁLYÁNAK KÖZLEMÉNYEI

XV. kötet 1-2. szám

Szerkesztőség és kiadóhivatal: 1117 Budapest, Bogdánfy út 10/B.

Az Alkalmazott Matematikai Lapok változó terjedelmű füzetekben jelenik meg, és olyan eredeti tudományos cikkeket publikál, amelyek a gyakorlatban, vagy más tudományokban közvetlenül felhasználható új matematikai eredményt tartalmaznak, illetve már ismert, de színvonalas matematikai apparátus újszerű és jelentős alkalmazását mutatják be. A folyóirat közöl cikk formájában megírt, új tudományos eredménynek számító programokat, és olyan, külföldi folyóiratban már publikált dolgozatokat, amelyek magyar nyelven történő megjelentetése elősegítheti az elért eredmények minél előbbi, széles körű hazai felhasználását. A szerkesztőbizottság bizonyos időnként lehetővé kívánja tenni, hogy a legjobb cikkek nemzetközi folyóiratok különszámaként angol nyelven is megjelenhessenek.

A folyóirat feladata a Magyar Tudományos Akadémia III. (Matematikai és Fizikai) Osztályának munkájára vonatkozó közlemények, könyvismertetések stb. publikálása is.

A kéziratok a főszerkesztőhöz, vagy a szerkesztőbizottság bármely tagjához beküldhetők. A főszerkesztő címe:

Benczúr András, főszerkesztő
1117 Budapest, Bogdánfy út 10/B.

Közlésre el nem fogadott kéziratokat a szerkesztőség lehetőleg visszajuttat a szerzőhöz, de a beküldött kéziratok megőrzéséért vagy továbbításáért felelősséget nem vállal.

Az Alkalmazott Matematikai Lapok előfizetési ára kötetenként 192 forint (a XVI. kötettől 850 forint). Megrendelések a szerkesztőség címén lehetségesek (átutalásokat az ELTE MNB 232–90142–6207 számlaszámára a 9015567 munkaszám és 694 utalványozási kód megjelöléssel kérjük).

A Magyar Tudományos Akadémia III. (Matematikai és Fizikai) Osztálya a következő idegen nyelvű folyóiratokat adja ki:

1. Acta Mathematica Hungaricae,
2. Acta Physica Hungaricae,
3. Studia Scientiarum Mathematicarum Hungarica.

Alkalmazott Matematikai Lapok 15 (1990–91) 1–90

KÖZÖNSÉGES DIFFERENCIÁLEGYENLETEK MEGOLDÁSAINAK STABILITÁSÁRÓL MECHANIKAI ALKALMAZÁSOKKAL

HATVANI LÁSZLÓ

Szeged

A dolgozat témája LJAPUNOV direkt módszerének nem–autonóm rendszerekre, illetve a változók egy részére vonatkozó (parciális) stabilitásvizsgálatokra való továbbfejlesztése, tökéletesítése és klasszikus mechanikai feladatokra történő alkalmazása. A 2. fejezetben olyan tételeket adunk, amelyek negatív szemidefinit (de nem negatív definit) deriválttal rendelkező *Ljapunov–függvények* segítségével lokalizálják a megoldások pozitív határhalmazait. A 3. fejezetben ezeket az eredményeket egyensúlyi helyzet aszimptotikus stabilitását biztosító feltételek levezetésére használjuk. A 4. fejezetben a parciális határhalmazok lokalizálásának problémáját oldjuk meg. Az 5. fejezetben olyan rendszerekkel foglalkozunk, amelyeknek jobboldala valamilyen értelemben konvergens, ha a nem–ellenőrzött koordináták, illetve az idő végtelenbe tart. Az így keletkező „határegyenlet" tulajdonságaiból következtetünk az eredeti egyenlet megoldásainak megfelelő tulajdonságaira. A 6. fejezetben olyan *Ljapunov–függvények* elméletét dolgozzuk ki, amelyek két függvény összegeként állnak elő úgy, hogy az összegnek a rendszer szerinti deriváltja az egyik tag felhasználásával becsülhető.

Tartalomjegyzék

Bevezetés

A stabilitás fogalma a mechanikából ered. Mint ismeretes, egy mechanikai rendszer állapotát helyzete és sebessége határozza meg, amelyeket közös néven állapothatározóknak nevezünk. Egy egyensúlyi helyzetet stabilisnak mondunk, ha a rendszer mozgásai során az állapothatározóknak a nyugalmi állapottól való eltérése tetszőlegesen kicsiny marad, feltéve, hogy a rendszer az egyensúlyi helyzethez elegendően közelről indult elegendően kicsiny kezdő sebességgel. Ha az eltérés még nullához is tart, amint az idő végtelenbe tart, az egyensúlyi helyzetet aszimptotikusan stabilisnak mondjuk. Ha az egyensúlyi helyzet nem stabilis, akkor instabilisnak nevezzük. Például, a matematikai inga pályájának legalsó pontja stabilis egyensúlyi helyzet, a pálya legfelső pontja viszont instabilis.

Tekintsünk most általánosabban egy olyan, időben változó rendszert vagy folyamatot, amelynek matematikai modellje egy

$$\dot{x} = X(x,t) \qquad (t \geq 0, \quad x \in R^k) \tag{0.1}$$

közönséges differenciálegyenlet–rendszer; t az időt jelöli, x pedig a rendszert jellemző állapothatározókból álló vektor. Jelölje $x(t;x_0,t_0)$ a (0.1) egyenlet $x(t_0;x_0,t_0) = x_0$ feltételnek eleget tevő megoldását. Az x_0 kezdeti állapot meghatározása méréssel történik, tehát nem abszolút pontos. Ha a valódi x_0 állapotra méréssel a ξ közelítés adódott, akkor modellünk alapján arra a következtetésre jutunk, hogy a rendszer a t időpillanatban az $x(t;\xi,t_0)$ állapotban található, pedig $x(t;x_0,t_0)$ lesz a valódi állapota. A modell csak akkor használható a gyakorlatban a rendszer leírására, ha ez a két állapot kicsit tér el egymástól, feltéve, hogy ξ elég jól közelíti x_0–t. Ugyanazzal a problémával kerültünk szembe az általános (0.1) differenciálegyenlet egy tetszőleges $x(t;x_0,t_0)$ megoldására vonatkozóan, mint az előbb egy mechanikai rendszer egyensúlyi állapotának tanulmányozásánál.

A stabilitás egzakt matematikai elméletének megalapozója a kiemelkedő orosz mechanikus és matematikus, A. M. LJAPUNOV. 1892–ben megjelent doktori értekezésében — elvonatkoztatva a mechanikai rendszer egyensúlyi állapotától — megadta a (0.1) egyenlet tetszőleges megoldása stabilitásának ma is használatos definícióját : az $x(t;x_0,t_0)$ megoldás stabilis, ha az $|x(t;\xi,t_0) - x(t;x_0,t_0)|$ eltérés tetszőlegesen kicsiny t-ben egyenletesen a $[t_0, \infty)$ intervallumon, feltéve, hogy $|\xi - x_0|$ elég kicsiny. Ha az eltérésről még azt is tudjuk, hogy 0-hoz tart, ha $t \to \infty$, akkor a megoldás aszimptotikusan stabilis.

Az $y = x - x(t; x_0, t_0)$ új függő változó bevezetésével az $x = x(t; x_0, t_0)$ kiszemelt megoldás az $y \equiv 0$ megoldásba megy át, és y-ra ugyancsak egy (0.1) típusú egyenlet teljesül, ezért feltehetjük, hogy (0.1)-nek az $x \equiv 0$ megoldása és mi ennek a megoldásnak a stabilitási tulajdonságait vizsgáljuk.

Ebben a dolgozatban olyan feltételeket adok meg, amelyeknek teljesülése a megoldások ismerete nélkül ellenőrizhető, és amelyek biztosítják, hogy a nem-autonóm (0.1) rendszer 0-megoldása aszimptotikusan stabilis, illetve instabilis legyen. Az eredmények legtöbbje alkalmas nemcsak az állapothatározók összességére, hanem azoknak egy részére vonatkozó, un. parciális stabilitási tulajdonságok megállapítására is. A tételek egy része kifejezetten parciális stabilitási tulajdonságokra vonatkozik. Ide tartoznak azok az eredmények, amelyekben arra a régi problémára adok megoldást, hogy miként lehet parciális aszimptotikus stabilitást, illetve instabilitást biztosítani akkor, ha a nem-ellenőrzött koordináták a megoldások mentén nem-korlátosak is lehetnek.

A tételek jelentős részéhez az elméleti mechanika stabilitáselméletének néhány klasszikus problémájával foglalkozva jutottam el. Ezek egyike annak meghatározása, hogy holonóm szkleronóm mechanikai rendszer stabilis egyensúlyi állapota milyen surlódási és giroszkópikus erők hatására válik aszimptotikusan stabilissá. Ezzel a kérdéssel kis rezgésekre már KELVIN és P. G. TAIT [53] is foglalkozott. Érdekes, hogy a stacionárius esetben, vagyis amikor a ható erők explicit módon nem függnek az időtől, teljes disszipáció esetén az aszimptotikus stabilitás a mechanikai tapasztalatok alapján (l. csillapított rezgőmozgás) egészen kézenfekvőnek látszik, bizonyítására mégis egészen 1966-ig kellett várni (L. SALVADORI [47]), amikor az invarianciaelv már ismertté vált. Már a tapasztalat is mutatja, hogy az instacionárius eset jóval bonyolultabb, hiszen ha a surlódási erő fékező hatása az idő függvényében túl kicsi vagy túl nagy, akkor nincs aszimptotikus stabilitás. Matematikailag a nehézségeket úgy foglalhatjuk össze, hogy az invarianciaelv az instacionárius esetre nem alkalmazható. Ebben az esetben a probléma a következő alakot ölti: Adjunk meg olyan, az erők időtől való függését szabályozó feltételeket, amelyek teljesülése esetén az egyensúlyi állapot aszimptotikusan stabilis. Dolgozatomban a disszipáció különböző teljességi fokainak bevezetésével megadok ilyen feltételeket. Bebizonyítom továbbá azt a tapasztalat által ugyancsak kézenfekvőnek látszó régi sejtést, hogy az egyensúlyi állapot a sebességekre vonatkozóan teljes disszipáció esetén tetszőleges nem-korlátos surlódási erők mellet is aszimptotikusan stabilis.

Tekintsük most át részletesebben az egyes fejezetek tartalmát, kitérve arra, hogy milyen előzmények birtokában születtek a tételek és hogyan kapcsolódnak a témakör alapvető eredményeihez.

Ismeretes, hogy a differenciálegyenletek, köztük a mechanikai rendszerek mozgásegyenletei általában nem integrálhatók. Ezért a megoldások (mozgások) stabilitási tulajdonságait a megoldások ismerete nélkül, közvetlenül az egyenletben szereplő függvények (potenciális energia, mozgási energia, ható erők) ismeretében kell megállapítanunk. A. M. LJAPUNOV említett értekezésében egy ilyen módszernek az alapjait is lefektette. Észrevette, hogy DIRICHLET–nek a konzervatív mechanikai rendszer egyensúlyi helyzetének stabilitására vonatkozó tételre adott bizonyítása, amely a teljes mechanikai energiának a mozgások mentén való invarianciáján, azaz az energia megmaradásának elvén alapult, átvihető a (0.1) általános egyenletre. Az így keletkezett *Ljapunov-féle direkt módszer* során az egyenletben szereplő függvények felhasználásával konstruálunk egy, az állapothatározók halmazán értelmezett V függvényt, egy u.n. *Ljapunov-függvényt*, és — csupán az egyenletrendszert használva — meghatározzuk annak a rendszer megoldásai (a mozgások) mentén vett $\dot{V}$ deriváltját. A V és $\dot{V}$ függvényekkel megfogalmazhatók olyan feltételek, amelyek teljesülése esetén a rendszer triviális megoldása (az egyensúlyi helyzet) stabilis, illetve aszimptotikusan stabilis, vagy instabilis. Az első ilyen tételeket LJAPUNOV adta meg. Az aszimptotikus stabilitásról szóló alaptételének a jelen értekezésben is különösen fontos szerepe lesz, ezért példaként ezt idézzük : Ha a (0.1) rendszerhez létezik olyan $V(x,t)$ függvény, amely (1) pozitív definit; (2) $V(x,t) \rightrightarrows 0$, ha $x \to 0$, $t \in R_+$; (3) $\dot{V}(x,t)$ negatív definit, akkor a (0.1) rendszer 0–megoldása aszimptotikusan stabilis.

A direkt módszer mind elméleti, mind gyakorlati szempontból nagyon hasznosnak bizonyult. A 40–50–es évektől kezdődően, amikor kiderült, hogy az irányítási rendszerek konstruálásában és stabilizálásában is a leghasznosabb módszer, és amikor az elektronikus számítógépek megjelenése révén a numerikus közelítés előtérbe került, különösen nagy lendülettel kezd fejlődni. *Princetonban*, 1949–ben újranyomják LJAPUNOV több, mint 50 éves doktori értekezését. Mind Keleten, mind Nyugaton számos monográfia készül a témából. Konferenciasorozatokat szerveznek, és a meginduló szovjet és amerikai differenciálegyenletes folyóiratoknak is egyik központi témája a stabilitáselmélet. Igen sok, a gyakorlatban fontos rendszerhez sikerül *Ljapunov-függvényt* konstruálni, az alaptételekről pedig bebizonyítják, hogy bizonyos módosításokkal megfordíthatókká válnak. Mindazonáltal, a módszer alkalmazásának máig legnehezebb problémája az adott rendszerhez alkalmas *Ljapunov-függvény* konstruálása, amelyre nincs általános algoritmus. Ezért fontos vizsgálni, hogyan lehet gyengíteni, vagy akár csak átfogalmazni az alaptételek egy–egy feltételét. Különösen sok nehézséget okoz annak a feltételnek a teljesítése, hogy a *Ljapunov-függvénynek* a rendszer szerinti deriváltja legyen negatív definit. Például, mint már említettük, mechanikai tapasztalatok alapján nyilvánvalónak látszik, hogy holonóm szkleronóm mechanikai rendszer stabilis egyensúlyi állapota instacionárius surlódási erők hatására teljes disszipáció esetén aszimptotikusan stabilissá válik. Ugyanakkor LJAPUNOV aszimptotikus stabilitásáról szóló tételével ezt nem tudjuk levezetni, mivel a kézenfekvő *Ljapunov-függvénynek*, a teljes mechanikai energiának a mozgások mentén vett deriváltja nem negatív definit, hanem csak negatív szemidefinit. Felmerül a kérdés : Milyen kiegészítő feltételek mellett tudunk szemidefinit deriválttal rendelkező *Ljapunov-függvény* birtokában aszimptotikus stabilitást állítani?

Ebben a problémakörben az első lényeges eredmény E. A. BARBASIN és N. N. KRASZOVSZKIJ híres tétele [64] : Legyen (0.1) autonóm, azaz tekintsünk egy

$$\dot{x} = X(x) \qquad (t \geq 0,\ x \in R^k;\ X(0) = 0) \tag{0.2}$$

rendszert, és tegyük fel, hogy létezik olyan $V : R^k \to R$ függvény, amelyre teljesülnek a következők : (1) V pozitív definit; (2) $\dot{V}(x) \leq 0$; (3) a $\dot{V}^{-1}(0)$ halmaz nem tartalmazza a (0.2) egyenlet egyetlen teljes trajektóriáját sem a $\{0\}$ kivételével. Ekkor a (0.2) egyenlet 0–megoldása aszimptotikusan stabilis.

Ez a tétel rendkívül alkalmazhatónak és gyümölcsözőnek bizonyult, így sok dolgozat foglalkozik a (0.1) nem–autonóm rendszerre való kiterjesztésével (Z. ARTSTEIN [2–4], T. A. BURTON [9], J. KATO [33], J. P. LASALLE [35, 36], V. M. MATROSZOV [69], N. ONUCHIC, L. R. ONUCHIC, P. TABOAS [41], A. SZ. OZIRANER, V. V. RUMJANCEV [73], N. ROUCHE, P. HABETS, M. LALOY [46], L. SALVADORI [48], A. A. SESZTAKOV [51], T. YOSHIZAWA [56–58]). LASALLE [36] bebizonyította, hogy ha a *Ljapunov–függvény* deriváltja egy $\dot{V}(x,t) \leq -U(x) \leq 0$ egyenlőtlenségnek tesz eleget, és $X(x,t)$ a t–nek bizonyos értelemben korlátos függvénye, akkor a (0.1) rendszer megoldásainak trajektóriái az $U^{-1}(0)$ halmazhoz konvergálnak, ha $t \to \infty$. Az alkalmazások során kiderült, hogy a tételben megkívánt egyenlőtlenség sokszor túl szigorú. Például, ha mechanikai rendszerre surlódási erők hatnak, és a surlódási együtthatók az időtől függnek, akkor a mechanikai energia deriváltjára egy

$$\dot{V}(x,t) \leq -\phi(t)\, U(x) \qquad (\phi(t) \geq 0,\ U(x) \geq 0) \tag{0.3}$$

típusú egyenlőtlenség teljesül, ahol $\phi(t)$ időnként még a 0 értéket is felveheti.

A 2. fejezetben olyan lokalizációs lemmát adok meg, amelynek alkalmazásával bebizonyítható, hogy ha $\dot{V}$–ra (0.3) teljesül, és ϕ integrálisan pozitív, akkor a megoldások trajektóriái az $U^{-1}(0)$ halmazhoz konvergálnak. Sőt, az így kapott tétel egy másik lényeges szempontból is általánosítja LASALLE eredményét, amennyiben benne az $X(x,t)$ korlátossága helyett egy $W : R^k \to R^r$ segédfüggvény $\dot{W}(x,t)$ deriváltjának korlátosságát követeljük meg, ami — ahogyan a felsorolt példák mutatják — a tételt sokkal rugalmasabbá, alkalmazhatóbbá teszi. Ennek az az ára, hogy R^k–ban be kell vezetni egy új, a W által generált környezetrendszert, amellyel R^k esetleg nem lesz lokálisan kompakt. A fejezet második felében az előbbi környezetrendszeren alapuló módszer továbbfejlesztésével olyan lokalizációs tételeket bizonyítok, amelyek — szemben a korábbi eredményekkel — az $X(x,t)$, illetve $\dot{W}(x,t)$ függvényektől nem kívánnak semmilyen korlátossági feltételt.

A 3. fejezet alkalmazott jellegű. A lokalizációs eredmények felhasználásával megadom a *Barbasin–Kraszovszkij–tétel* egy kiterjesztését nem–autonóm rendszerekre, amelynek segítségével elegendő feltételek adhatók nem–lineáris másodrendű differenciálegyenlet 0–megoldása, illetve instacionárius disszipatív és giroszkópikus erők hatása alatt álló mechanikai rendszer egyensúlyi állapota aszimptotikus stabilitására, instabilitására.

A 4. és 5. fejezet parciális stabilitási tulajdonságokkal foglalkozik. A parciális stabilitás fogalmát V. V. RUMJANCEV [74] vezette be 1957-ben.

A mechanikában gyakran tapasztalható, hogy egy egyensúlyi helyzet az állapothatározók nem mindkét típusára, hanem vagy csak a sebességekre, vagy csak a helykoordinátákra vonatkozóan rendelkezik valamilyen stabilitási tulajdonsággal. Például egy vízszintes sík lapon surlódás hatása alatt mozgó golyó sebessége nullához tart, de egy kiszemelt (a mozgásoktól független) egyensúlyi helyzettől való eltérése nem. Más szóval, az egyensúlyi helyzetek csak a sebességekre vonatkozóan aszimptotikusan stabilisak. Általában, a parciális stabilitási tulajdonságok az állapothatározóknak csak egy részére, az un. ellenőrzött koordinátákra tesznek megállapításokat.

A direkt módszernek a parciális aszimptotikus stabilitásra és instabilitásra való kiterjesztése során felmerülő problémák közül a legsúlyosabb az, hogy a nem-ellenőrzött koordinátákról a mozgások során még korlátosságot sem tudunk, így a klasszikus módszerben oly fontos szerepet játszó kompaktsági meggondolásokat nem alkalmazhatjuk (pl. az invarianciaelv nem alkalmazható). Ezen úgy szoktak segíteni, hogy feltételezik a nem-ellenőrzött koordinátáknak a mozgások során való korlátosságát (pl. [72, 73, 44, 61]). Ez a feltétel azonban közvetlenül nem ellenőrizhető, „a priori" ismeretet tételez fel a megoldásokról, így kívánatos lenne tőle megszabadulni. A. SZ. OZIRANER [72] példája mutatja, hogy elhagyni nem lehet, így az a kérdés, hogyan lehetne valamilyen ellenőrizhető feltétellel helyettesíteni.

A 4. fejezetben autonóm rendszerekkel foglalkozva bevezetem a parciális határhalmaz fogalmát, majd ezek lokalizálása céljából a következő alternatívát bizonyítom: ha egy megoldás mentén az ellenőrzött koordináták vektora korlátos, akkor vagy a nem-ellenőrzött koordinátákból álló vektor normája végtelenhez tart, ha az idő végtelenhez tart, vagy pedig a megoldás határhalmaza a *Ljapunov-függvény* egy szinthalmazán fekszik. A lokalizáláshoz tehát csak az első esettel kell foglalkozni. A 4. fejezetben ehhez a *Ljapunov-függvényre* adok meg alkalmas, közvetlenül ellenőrizhető feltételeket. Az 5. fejezetben feltételezem, hogy a (0.2) egyenlet jobboldalán $X(y,z)$ valamilyen értelemben konvergál egy $X^*(y)$ függvényhez, ha $|z| \to \infty$ (ahol $x = (y,z)^T$, és z a nem-ellenőrzött koordináták vektora), és a határegyenletek módszerének [50] alkalmas módosításával a határegyenletre adok megfelelő feltételeket.

A 6. fejezet eredményei egyaránt sorolhatók a *Ljapunov-elméletehez* és a differenciálegyenlőtlenségek elméletéhez. Disszipatív rendszerekben a sebességekre vonatkozó aszimptotikus stabilitás problémájának megoldásához itt olyan tételeket bizonyítok, amelyek egy $(V_1(x,t) + V_2(x,t))^{\cdot} \leq -\phi(t)c(V_1(x,t))$ típusú egyenlőtlenség teljesülését tételezik fel ($\phi(t) \geq 0$; $c(r) \geq 0$ és monoton növő), és azt állítják, hogy $V_1(x,t) \to 0$ és $V_2(x,t)$-nek létezik a határértéke a megoldások mentén, ha $t \to \infty$.

Már az eddigi példákból is kiderült, hogy a *Ljapunov-féle direkt módszer* igazából csak egy jó programot ad arra, hogy a konkrét rendszerekre hogyan állapítsunk meg stabilitási tulajdonságokat. A legtöbb tétel ilyen szerkezetű : „Ha a rendszerhez létezik olyan *Ljapunov-függvény*, hogy ..., akkor a rendszer 0–megoldása ... ." *Ljapunov-függvények* konstruálására viszont nincs általános algoritmus. E. A. BARBASIN [62] írja az eddig meglévő partikuláris módszereket összegyűjtő könyvében, hogy nem–lineáris rendszerekhez alkalmas *Ljapunov-függvény* megtalálása ma sokszor még szerencse dolga (hallottam már a *Ljapunov-függvény-keresés* művészetéről is). Mindenesetre, hogy egy általános *Ljapunov-típusú tétel* ér–e valamit, vagy sem, az dönti el, hogy reális modellekhez tudunk–e találni olyan *Ljapunov-függvényt*, amely a kívánt tulajdonságokat teljesíti. Ezért úgy láttam célszerűnek, hogy az egyes alkalmazásokat, példákat közvetlenül az illusztrálandó alaptételek után tárgyaljam, így ugyanaz a modell a dolgozatban több helyen is előfordul.

A kifejtés módszerének alapelve : az eredményekről egy olyan áttekintést adni, amely biztosítja, hogy az olvasó tájékozódjék az alapvető problémákban és a lehetőségekhez mérten könnyen kivehesse és nyomon követhesse az alapvető gondolatokat. Ezért az eredményeket legtöbbször nem teljes általánosságukban, néha csak egy–egy példán illusztrálva ismertetem, az általános esetre csak utalok, vagy vázlatos kifejtést adok. A fejezeteket az eredmények irodalmi előzményeinek, utóhatásának, visszhangjának rövid áttekintésével zárom.

Más szerzők eredményeit az idézett tételek sorszámaihoz tett $*$–gal különböztetem meg saját eredményeimtől (pl. 4.1.* TÉTEL).

1. fejezet

Alapfogalmak, előzmények

Ebben a dolgozatban, mint már említettük, közönséges differenciálegyenlet–rendszerek megoldásainak aszimptotikus viselkedésével foglalkozunk, tehát az állapottér (fázistér) a valós koordinátájú $x = \mathrm{col}(x^1, \dots, x^k)$ oszlopvektorok R^k tere, a t idő az R valós számegyenesen, illetve az $R_+ := [0, \infty)$ félegyenesen változik. Tegyük fel, hogy az R^k téren adott egy $|\cdot| : R^k \to R_+$ norma. Ha $x, y \in R^k$, akkor $d(x,y) := |x - y|$ a két elem távolságát jelöli.

Az R^k térhez az egyetlen ∞ ideális pontnak a hozzávételével az R^k_∞ teret kapjuk, miközben tetszőleges $x \in R^k_\infty$–ra $d(x, \infty) := 1/|x|$. Ha $K, H \subset R^k_\infty$, akkor távolságukat a $d(K, H) := \inf\{d(x,y) : x \in K,\ y \in H\}$ képlet definiálja. Ha $p \in R^k_\infty$, és $\rho > 0$ adott, akkor $B(p, \rho)$ a p körüli ρ sugarú nyitott gömböt jelöli (a dimenzió hangsúlyozására használjuk a $B_k(p, \rho)$ jelölést is). Ha $H \subset R^k_\infty$ és $\rho > 0$, akkor $B(H, \rho)$ az R^k_∞ tér H–hoz ρ–nál közelebb eső pontjainak halmazát jelöli.

Adott $K \subset R^k_\infty$ halmaz és $u : [0, \omega) \to R^k$ függvény esetén az $u(t) \to K$ $(t \to \omega - 0)$ konvergencia azt jelenti, hogy $d(K, u(t)) \to 0$ $(t \to \omega - 0)$.

Tekintsük az

$$\dot{x} = X(x, t) \tag{1.1}$$

differenciálegyenlet–rendszert, ahol az $X : \Gamma \to R^k$ függvény legalább folytonos; $\Gamma := G \times R_+; G \subset R^k$ adott összefüggő nyitott halmaz. Jelölje $x(t) = x(t; x_0, t_0)$ az

$$(1.2) \qquad \dot{x} = X(x,t), \quad x(t_0) = x_0 \quad ((x_0, t_0) \in \Gamma)$$

kezdetiérték–probléma egy tetszőleges megoldását. „Az X függvény legalább folytonos" feltétel azt jelenti, hogy X–ről csupán folytonosságot teszünk fel minden olyan állításban, amelynek bizonyítása közben csak az (1.2) probléma megoldásának létezését használjuk. Néhány állítás bizonyításában arra is szükségünk lesz, hogy $x(t; x_0, t_0)$ x_0–nak folytonos függvénye, ilyenkor azt is feltesszük, hogy teljesül ennek a tulajdonságnak egy elegendő feltétele (pl. X x–ben *Lipschitz–feltételnek* tesz eleget). Ha az (1.1) egyenlet jobboldala t–től független, vagyis az egyenlet

$$(1.3) \qquad \dot{x} = X(x)$$

alakú, más szóval, autonóm, akkor $x(t; x_0, t_0)$ nem függ lényegesen t_0–tól, hiszen $x(t; x_0, t_0) = x(t - t_0; x_0, 0)$, tehát (1.2) helyett csak az

$$(1.4) \qquad \dot{x} = X(x), \quad x(0) = x_0 \qquad (x_0 \in G)$$

kezdetiérték–problémát tekintjük, amelynek egy tetszőleges megoldását $x(t; x_0)$–lal jelöljük.

Gyakran fogjuk vizsgálni az $x^1, \dots, x^k$ állapothatározókoordináták egy részére vonatkozó, un. *parciális* stabilitási tulajdonságok feltételeit is. Ekkor az $x \in R^k$ vektornak egy $x = \mathrm{col}(y, z)$ partícióját tekintjük, ahol $y \in R^m, z \in R^n$ $(1 \le m \le k,\ n := k - m)$. Az y vektor az x vektor azon koordinátáiból áll, amelyekre vonatkozóan az adott tulajdonságot megkívánjuk — ezeket *ellenőrzött* koordinátáknak fogjuk hívni. Ebben az esetben az (1.1) egyenletet az

$$(1.5) \qquad \dot{y} = Y(y,z,t), \quad \dot{z} = Z(y,z,t)$$

alakban is használjuk, és mindig feltesszük, hogy a jobboldalak a $\Gamma_m = G_m \times R_+$ halmazon vannak értelmezve, ahol $G_m = B_m(0,h) \times R^n$, $0 < h \le \infty$. Következetesen Γ'_m–vel jelöljük a Γ_m halmaz egy y–ban kompakt $B_m(0,h') \times R^n$ részhalmazát, ahol $0 < h' < h$. Hallgatólagosan mindig feltesszük továbbá, hogy az (1.5) $x(t) = (y(t), z(t))^T$ megoldásai *z–folytathatók*, vagyis ha $y(t)$ korlátos a $[t_0, T)$ intervallumon $(t_0 < T < \infty)$, akkor az $x(t)$ megoldás folytatható a $[t_0, T]$ intervallumra (azaz egyetlen megoldásnak sem lehet olyan „szökési ideje", ahol csak a nem–ellenőrzött koordináták vektora „szökik").

Az (1.1) és (1.3) egyenlet megoldásait geometriailag az integrálgörbékkel és trajektóriákkal lehet ábrázolni.

Tekintsük az (1.1) (vagy az (1.3)) egyenlet egy $x = \phi(t)$ megoldását. Ismeretes (pl. [15]), hogy ϕ–nek létezik egy jobbra maximális $\hat{\phi}$ folytatása, amely egy $[t_0, \omega)$

intervallumon van értelmezve, ahol $t_0 < \omega \leq \infty$, és $\hat{\phi}(t)$ G határához tart, ha $t \to \omega - 0$ (ha $\omega < \infty$, akkor szökési időnek nevezzük).

A továbbiakban az egész dolgozat során, ha tekintjük egy egyenlet egy tetszőleges megoldását, akkor azon mindig jobbra maximálisan folytatott megoldást értünk ($\phi \equiv \hat{\phi}$).

A $t \to \phi(t)$ $(t_0 \leq t < \omega)$ függvény grafikonját az R^{k+1} kibővített fázistérben a ϕ megoldás *integrálgörbéjének* nevezzük, a $\gamma^+(\phi) := \{\phi(t) : t_0 \leq t < \omega\} \subset R^k$ halmazt pedig a megoldás *pozitív féltrajektóriájának* (félpályájának) hívjuk. Ha ϕ az (1.3) autonóm egyenlet megoldása, akkor feltehető, hogy balra is maximálisan folytatott, azaz $\phi : (\alpha, \omega) \to R^k$ $(-\infty \leq \alpha < \omega \leq \infty)$. Ekkor a $\gamma(\phi) := \{\phi(t) : \alpha < t < \omega\} \subset R^k$ halmazt a megoldáshoz tartozó (teljes) *trajektóriának* (pályának) nevezzük.

A $H \subset G$ halmazt az (1.3) egyenletre nézve *invariánsnak* nevezzük, ha bármely $x_0 \in H$ ponthoz létezik az (1.4) problémának olyan $\phi : (\alpha, \omega) \to R^k$ megoldása, amelyre $\gamma(\phi) \subset H$.

Legyen $\phi : [t_0, \omega) \to R^k$ megoldása (1.1)-nek (vagy (3.1)-nek). Azt mondjuk, hogy $p \in \bar{G}$ *pozitív határpontja* ϕ-nek, ha létezik olyan $\{t_n\}$ sorozat, amelyre $t_n \to \omega - 0$, $\phi(t_n) \to p$ $(n \to \infty)$ teljesül. A ϕ megoldás összes pozitív határpontjának halmazát $\Omega(\phi)$-vel jelöljük, és a ϕ megoldás *pozitív határhalmazának* nevezzük.

A határhalmazok tulajdonságait foglalja össze az alábbi két lemma, amelyek bizonyítása [46] III. függelékében található.

1.1.* LEMMA. *Legyen ϕ (1.1)-nek tetszőleges megoldása. Ekkor*

a) $\overline{\gamma^+(\phi)} = \gamma^+(\phi) \cup \Omega(\phi)$;

b) $\Omega(\phi)$ *zárt;*

c) ha $\gamma^+(\phi)$ *korlátos, akkor* $\Omega(\phi)$ *nem üres, kompakt, összefüggő, és* $\phi(t) \to \Omega(\phi)$ $(t \to \omega - 0)$; *továbbá, ha* M *zárt és* $\phi(t) \to M$ $(t \to \omega - 0)$, *akkor* $\Omega(\phi) \subset M$;

d) ha $G \cap \Omega(\phi)$ *nem üres, akkor* $\omega = \infty$.

1.2.* LEMMA. *Ha ϕ megoldása az autonóm (1.3) egyenletnek, akkor $\Omega(\phi) \cap G$ invariáns az (1.3) egyenletre nézve.*

Az 1.1.* lemma *c)* állítása mutatja, hogy a megoldások határhalmaza meghatározó a megoldások aszimptotikus viselkedésének leírása szempontjából, hiszen $\Omega(\phi)$ a legszűkebb olyan zárt halmaz R^k-ban, amelyhez a megoldások konvergálnak.

Most felidézzük a különböző stabilitási fogalmak definícióját (l. pl. [46]). Tegyük fel, hogy $0 \in G$ és $X(0,t) \equiv 0$ $(t \in R_+)$, vagyis $x = 0$ megoldása az (1.1) egyenletnek, továbbá bármely $t_0 \in R_+$ számhoz létezik olyan $\rho(t_0) > 0$, hogy ha $|x_0| < \rho(t_0)$, akkor $x(t; x_0, t_0)$ értelmezve van a $[t_0, \infty)$ intervallumon.

1.3 *Definíció.* Azt mondjuk, hogy az (1.1) egyenlet 0-megoldása:

(*a*) *y-stabilis*, ha bármely $\varepsilon > 0$, $t_0 \in R_+$ számokhoz létezik olyan $\delta = \delta(\varepsilon, t_0) > 0$, hogy ha $|x_0| < \delta$ és $t \geq t_0$, akkor az (1.2) probléma tetszőleges $x(t; x_0, t_0) = (y(t; x_0, t_0), z(t; x_0, t_0))^T$ megoldása esetén $|y(t; x_0, t_0)| < \varepsilon$ teljesül.

(*b*) *egyenletesen y-stabilis*, ha az előző definícióban δ nem függ t_0-tól.

(*c*) *y–vonzó (y–attraktív)*, ha bármely $t_0 \in R_+$ számhoz létezik olyan $\sigma = \sigma(t_0) > 0$, hogy ha $|x_0| < \sigma$, akkor $\lim_{t\to\infty} |y(t; x_0, t_0)| = 0$;

(*d*) *aszimptotikusan y–stabilis*, ha *y–stabilis* és *y–vonzó*;

(*e*) *ekvi–aszimptotikusan y–stabilis*, ha aszimptotikusan *y–stabilis*, és az *y–vonzás* definíciójában szereplő konvergencia $x_0 \in B_k(0, \sigma(t_0))$ –ban egyenletes, vagyis bármely $\eta > 0$ számhoz létezik olyan $T = T(\eta, t_0)$, hogy ha $|x_0| < \sigma(t_0)$, $t \geq t_0 + T$, akkor $|y(t; x_0, t_0)| < \eta$.

(*f*) *egyenletesen aszimptotikusan y–stabilis*, ha egyenletesen *y–stabilis*, ekvi–aszimptotikusan *y–stabilis*, továbbá σ és T nem függnek t_0–tól.

Ha az összes koordinátára vonatkozó stabilitási tulajdonságokról beszélünk ($y = x$, vagyis minden koordináta ellenőrzött), akkor *x–stabilitás, egyenletes x–stabilitás*, ... helyett egyszerűen *stabilitást, egyenletes stabilitást*, ... mondunk.

Az egyenletes aszimptotikus stabilitás fogalmát különösen fontossá teszi az a tény, hogy belőle a gyakorlati szempontból oly fontos *totális stabilitás* következik.

Ez a stabilitásfogalom nemcsak azt veszi figyelembe, hogy a kezdeti értékekben előfordulhatnak pontatlanságok, hanem azt is, hogy magát a modellt a mozgások során állandóan érik „kicsiny" zavarok (részletesen l. [46]).

A fent definiált stabilitási tulajdonságok feltételeinek tanulmányozására A. M. LJAPUNOV [65] adott meg egy módszert, amelynek lényege, hogy nem közvetlenül az $x(t)$ megoldás $|x(t)|$ normájának, hanem egy ügyesen választott $V(x,t)$ segédfüggvény és a megoldás $V(x(t), t)$ összetett függvényének a viselkedését vizsgálja a t idő nagy értékeire.

Adott $V : \Gamma \to R$ folytonos függvényhez az (1.1) egyenlet birtokában rendeljük hozzá a $\dot{V} : \Gamma \to R$ függvényt a következő definícióval :

$$\dot{V}(x,t) := \limsup_{h\to 0+0} (1/h)[V(x + hX(x,t), t + h) - V(x,t)].$$

Ezt a függvényt a V függvény (1.1) *rendszer szerinti deriváltjának* nevezzük. Ha hangsúlyozni akarjuk az (1.1) rendszerrel való kapcsolatát, akkor a $\dot{V}_{(1.1)}$ jelölést használjuk. Az elnevezést a következő lemma indokolja :

1.4.* LEMMA. (T. YOSHIZAWA [55], 3.o.). *Ha* $V \in Lip_x(\Gamma; R)$, *és* $\phi : [t_0, \omega) \to R^n$ *megoldása (1.1)-nek, akkor* $D^+V(\phi(t), t) \equiv \dot{V}(\phi(t), t)$ $(t \in [t_0, \omega))$.

A $\dot{V}(x,t)$ függvény meghatározásához nem kell ismerni a megoldásokat, birtokában mégis értékes információkat tudhatunk meg a megoldásokról. Például, ha $\dot{V}(x,t) \leq 0$ a Γ'_m halmazon, akkor $V(x,t)$ a megoldások mentén nem növekszik, amíg azok grafikonja Γ'_m–ben halad.

Jelölje $\mathcal{K}$ az $a : R_+ \to R_+$ folytonos, szigorúan monoton növekvő, $a(0) = 0$ feltételnek is eleget tevő függvények osztályát.

Most egy olyan fogalmat definiálunk, amely azt fejezi ki, hogy ha $V(y,z,t)$ kicsi, akkor $|y|$ is kicsi, éspedig (z,t)–re vonatkozóan egyenletesen.

1.5 *Definíció.* Azt mondjuk, hogy a $V : \Gamma_m \to R$ függvény *pozitív y–definit*, ha $V(0,t) \equiv 0$ $(t \in R_+)$ és létezik olyan $a \in \mathcal{K}$, hogy az $a(|y|) \leq V(y,z,t)$ egyenlőtlenség teljesül a Γ'_m halmazon.

A *Ljapunov-féle direkt módszerben* a legtöbbször azt bizonyítjuk be, hogy ha $x(t)$ (1.1)-nek megoldása, akkor $V(x(t),t) \to 0$ $(t \to \infty)$. Ha V pozitív y–definit; akkor ebből a 0–megoldás y–attraktivitása már azonnal következik.

1.6. *Megjegyzés.* A *Dini-féle konvergencia-tételhez* hasonlóan bizonyítható a következő állítás : Tegyük fel, hogy $V : \Gamma \to R$ pozitív y–definit, és bármely $t_0 \in R_+$ számhoz létezik olyan $\delta(t_0) > 0$, hogy ha $|x_0| < \delta(t_0)$, akkor $V(x(t;x_0,t_0),t)$ a t változóban monoton csökkenve 0-hoz tart, ha $t \to \infty$. Ekkor az (1.1) egyenlet 0–megoldása ekvi–aszimptotikusan y–stabilis [73].

Mint ahogyan az 1.3 definíció előtt a $\rho(t_0) > 0$ számok létezésére vonatkozó feltételből is látszik, a 0–megoldás stabilitási tulajdonságait csak akkor tudjuk vizsgálni, ha az origó kicsiny környezetéből induló megoldások jobbra akármeddig folytathatók. Mint kiderült, *Ljapunov direkt módszere* az összehasonlítási módszerrel [46] kombinálva ezen tulajdonság feltételeinek vizsgálatára is alkalmas. Ezt illusztrálja a következő tétel, amit itt bizonyítás nélkül idézünk.

A tétel kimondásához szükségünk van néhány új fogalomra. Ha $u,v \in R^r$, akkor $u \le v$ azt jelenti, hogy $u^i \le v^i$ $(i = 1,2,\dots,r)$. Ha $V : \Gamma \to R^r$ vektorértékü segédfüggvény, akkor a $\dot V$ műveletet komponensenként kell végrehajtani. Azt mondjuk, hogy egy $W : R^r \to R^r$ függvény kvázi-monoton nem–csökkenő [46], ha W^j nem–csökkenő $u^1, u^2, \dots, u^{j-1}, u^{j+1}, \dots, u^r$-ben $(j = 1,2,\dots,r)$.

1.7. TÉTEL. [22]. *Tegyük fel, hogy $\Gamma = R^k \times R_+$, és léteznek olyan $V : \Gamma \to R^r$, $w : R^r \times \Gamma \to R^r$ folytonos függvények, hogy V lokálisan Lipschitz tulajdonságú x–ben, $w(u,x,t)$ kvázimonoton nem–csökkenő u–ban és teljesül az*

$$\dot V(x,t) \le w(V(x,t),x,t) \tag{1.6}$$

egyenlőtlenség a Γ halmazon.

Legyen $x : [t_0,\omega) \to R^k$ (1.1)-nek megoldása. Ha az $\dot u = w(u,x(t),t)$, $u(t_0) = V(x(t_0),t_0)$ kezdetiérték-probléma $u^*(t)$ maximális megoldása definiálva van a $[t_0,\omega)$ intervallumon, $u^*(\omega - 0) := \lim_{t\to\omega-0} u^*(t)$ létezik és véges, és van olyan nem–csökkenő $p : R^r \to R$ funkcionál, hogy

$$\liminf_{\substack{|x|\to\infty \\ t\to\omega-0}} p(V(x,t)) > p(u^*(\omega - 0)),$$

akkor az $x(t)$ megoldás folytatható a $[t_0,T]$ intervallumra.

Kiegészítő megjegyzések

A megoldások folytathatóságával a kvalitatív elméletben sok dolgozat foglalkozik. T. A BURTON [10] adott egy olyan tételt, amely A. WINTNER, R. CONTI és T. YOSHIZAWA eredményeit kiterjesztve igen jól alkalmazható egészen általános egyenletekre is, viszont az egyik segédfüggvényről monotonitást tételez fel. Ez a feltétel WINTNERnél és CONTInál nem szerepelt, és BURTONnek az volt a sejtése, hogy tételének állítása igaz marad akkor is, ha ezt a feltételt elhagyjuk. [22] dolgozatomban egy példával megmutattam, hogy ez a sejtés nem igaz. Az ugyanitt közölt 1.7 tétel alkalmazásával az is megállapítható, hogy mi az a monotonitásnál lényegesen gyengébb feltétel, amellyel *Burton tételében* a monotonitás helyettesíthető.

Az 1.7 tétel a differenciálegyenlőtlenségek elméletén alapuló összehasonlítási módszert használja [46, 59]. Érdemes azonban megjegyezni, hogy a tételben az alapvető (1.6) differenciálegyenlőtlenség nem zárt (csak a V függvényt tartalmazó) függvényegyenlőtlenség, hanem az x függő változót is tartalmazza explicit módon. Ez komoly nehézségeket okoz, hiszen az egyenlőtlenséget nem lehet V–re megoldani, mint a korábbi tételekben. Az ilyen típusú egyenlőtlenségek alkalmazása [80] dolgozatommal kezdődött, amelynek alapvető ötlete az

$$\dot{u} = w(u, x, t), \qquad \dot{x} = X(x, t)$$

rendszer bevezetése és az $(u, x)^T$ megoldások parciális u–stabilitási tulajdonságainak vizsgálata, amelyekből először a $V(x(t), t)$ függvények, majd az $\dot{x} = X(x, t)$ rendszer $x(t)$ megoldásainak aszimptotikus tulajdonságaira következtethetünk.

Ezt a módszert — amit A. A. MARTYNYUK [37] beágyazásnak nevezett el — több irányba továbbfejlesztették [37, 68].

2. fejezet

Az invarianciaelv egy általánosítása nem-autonóm rendszerekre. Lokalizációs tételek.

A Bevezetésben már idéztük BARBASIN és KRASZOVSZKIJ alapvető tételét a

$$\dot{x} = X(x) \quad \left(x \in G \subset R^k;\ X(0) = 0\right) \tag{2.1}$$

autonóm differenciálegyenletrendszer 0–megoldásának aszimptotikus stabilitásáról. A tételnek a megoldások pozitív határhalmaza fogalmán alapuló bizonyításában a következő két lépés a leglényegesebb:

1. Tetszőleges $x = \varphi(t)$ megoldás $\Omega(\varphi)$ pozitív határhalmaza a V *Ljapunov-függvény* valamely szintfelületén fekszik, vagyis létezik olyan $c \in R$, hogy $\Omega(\varphi) \subset V^{-1}(c)$.

2. $\Omega(\varphi)$ invariáns, tehát tetszőleges pontján áthaladó megoldás trajektóriája is a szintfelületen halad, ahonnan $\Omega(\varphi) \subset \dot{V}^{-1}(0)$.

J. P. LASALLE ismerte fel 1968–ban, hogy a bizonyításnak ez a része nem csak a stabilitáselméletben, de a kvalitatív vizsgálatokban általában, a leghasznosabb eszközök egyike. Így született meg az *invarianciaelv:*

2.1.* TÉTEL. (J. P. LASALLE [35]). *Legyen $x = \varphi(t)$ a (2.1) egyenletnek egy megoldása, és tegyük fel, hogy létezik olyan lokálisan Lipschitz-tulajdonságú $V : G \to R$ függvény, amelyre $\dot{V}(x) \leq 0$ teljesül a φ megoldás $\gamma(\varphi)$ trajektóriájának pontjaiban. Ekkor $G \cap \Omega(\varphi) \subset M$, ahol M azon teljes trajektóriák egyesítése, amelyeknek mindegyike a $\dot{V}^{-1}(0)$ halmazban fekszik.*

Az invarianciaelv jelentősége abban áll, hogy lokalizálja a megoldás aszimptotikus viselkedését meghatározó pozitív határhalmazt. Pédául, ha tudjuk, hogy $\gamma(\varphi)$ korlátos és $\Omega(\varphi) \subset G$, akkor a pozitív határhalmaz vonzási tulajdonsága (l. 1. fejezet) miatt az invarianciaelv egy következményeként adódik a $\varphi(t) \to M (t \to \infty)$ tulajdonság.

A jelen fejezet célja ilyen lokalizációs tételek megalkotása a nem-autonóm

$$\dot{x} = X(x,t) \quad (X : \Gamma \to R^k) \tag{2.2}$$

differenciálegyenletrendszerre. Mint látni fogjuk, a fent említett első lépés (2.2)-re különösebb nehézségek nélkül általánosítható, mindössze a $V^{-1}(c)$ szintfelület megfelelőjét kell megtalálni. A második lépést viszont lehetetlenné teszi a nem-autonóm esetben az a tény, hogy a pozitív határhalmaz invarianciájának még csak értelme sincs. Igy az $\Omega(\varphi)$ és a $\dot{V}^{-1}(0)$ megfelelőjének kölcsönös helyzetét teljesen új módszerekkel kell tanulmányozni.

2.1. Lokalizáció Ljapunov-függvény szinthalmazával

A (2.2) rendszer nem autonóm, így egy $x : [t_0, \omega) \to R^k$ megoldását geometriailag $t \to (t, x(t))$ grafikonjával reprezentálhatjuk a kibővített R^{k+1} fázistérben. Azt mondjuk, hogy az x megoldás a $\Lambda \subset \Gamma$ halmazban halad, ha grafikonja Λ-ban fekszik.

2.1.1. *Definíció.* Legyen $\Lambda \subset \Gamma$ adott. Azt mondjuk, hogy $V \in \mathrm{Lip}(\Lambda; R)$ *a* (2.2) *rendszerhez tartozó Ljapunov-függvény* a Λ halmazon, ha létezik olyan $\eta : R_+ \to R_+$ folytonos függvény, amelyre az

$$\int_0^\infty \eta(t)dt < \infty, \quad \dot{V}(x,t) \leq \eta(t) \quad ((x,t) \in \Lambda)$$

egyenlőtlenségek teljesülnek.

A Λ halmazra vonatkozóan használni fogjuk a következő jelöléseket:

$$L(t) := \{x : (x,t) \in \Lambda\}, \quad L := \bigcup_{t \geq 0} L(t); \; L^* := \bigcap_{T \geq 0} \overline{\bigcup_{t \geq T} L(t)}$$

Ha egy $\varphi : [t_0, \omega) \to R^k$ megoldás Λ-ban halad, akkor $\Omega(\varphi) \subset L^*$.

A V függvénynek az ω, c ($0 < \omega \leq \infty$, $c \in R$) számokhoz tartozó *határszinthalmazán* [82] azt a $V^{-1}[c,\omega] \subset L^*$ halmazt értjük, amelynek bármely p pontjához létezik olyan $(x_i, t_i) \in \Lambda$ $(i = 1, 2, \dots)$ sorozat, hogy

$$\lim_{i\to\infty} x_i = p, \quad \lim_{i\to\infty} t_i = \omega - 0, \quad \lim_{i\to\infty} V(x_i, t_i) = c.$$

Ha V nem függ t-től, akkor nyilvánvalóan $V^{-1}[c,\omega] = V^{-1}(c)$. Könnyű belátni, hogy ha a $V(\cdot, t) : L(t) \to R$ $(t \in [0,\omega))$ függvénycsalád egyenlő mértékben folytonos, és

$$\lim_{t\to\omega-0} V(x,t) = V^*(x) \quad (x \in V^{-1}[c,\omega]),$$

akkor $V^{-1}[c,\omega] = (V^*)^{-1}(c)$.

2.1.2. LEMMA. *Legyen V a (2.2) egyenlethez tartozó Ljapunov-függvény a Λ halmazon, és legyen $\varphi : [t_0, \omega) \to R^k$ (2.2)-nek tetszőleges, Λ-ban haladó megoldása, amelyre $\Omega(\varphi) \neq \emptyset$. Tegyük fel, hogy az L^* halmaz minden p pontjához létezik olyan $\delta(p) > 0$ szám, hogy $V(x,t)$ alulról korlátos a $\{(x,t) \in \Lambda : |x - p| < \delta\}$ halmazon. Ekkor a*

$$\lim_{t\to\omega-0} V(\varphi(t), t) =: c$$

határérték létezik és véges, következésképpen

$$\Omega(\varphi) \subset V^{-1}[c,\omega].$$

Bizonyítás. Legyen $p \in \Omega(\varphi)$ és $\{t_i\}_{i=1}^{\infty}$ olyan sorozat, hogy

$$\lim_{i\to\infty} t_i = \omega - 0, \quad \lim_{i\to\infty} \varphi(t_i) = p.$$

Vezessük be a

$$w(t) := V(\varphi(t), t) + \int_t^{\omega} \eta(s) ds \quad (t_0 \leq t < \omega)$$

jelölést. A *Ljapunov-függvény* definíciója szerint

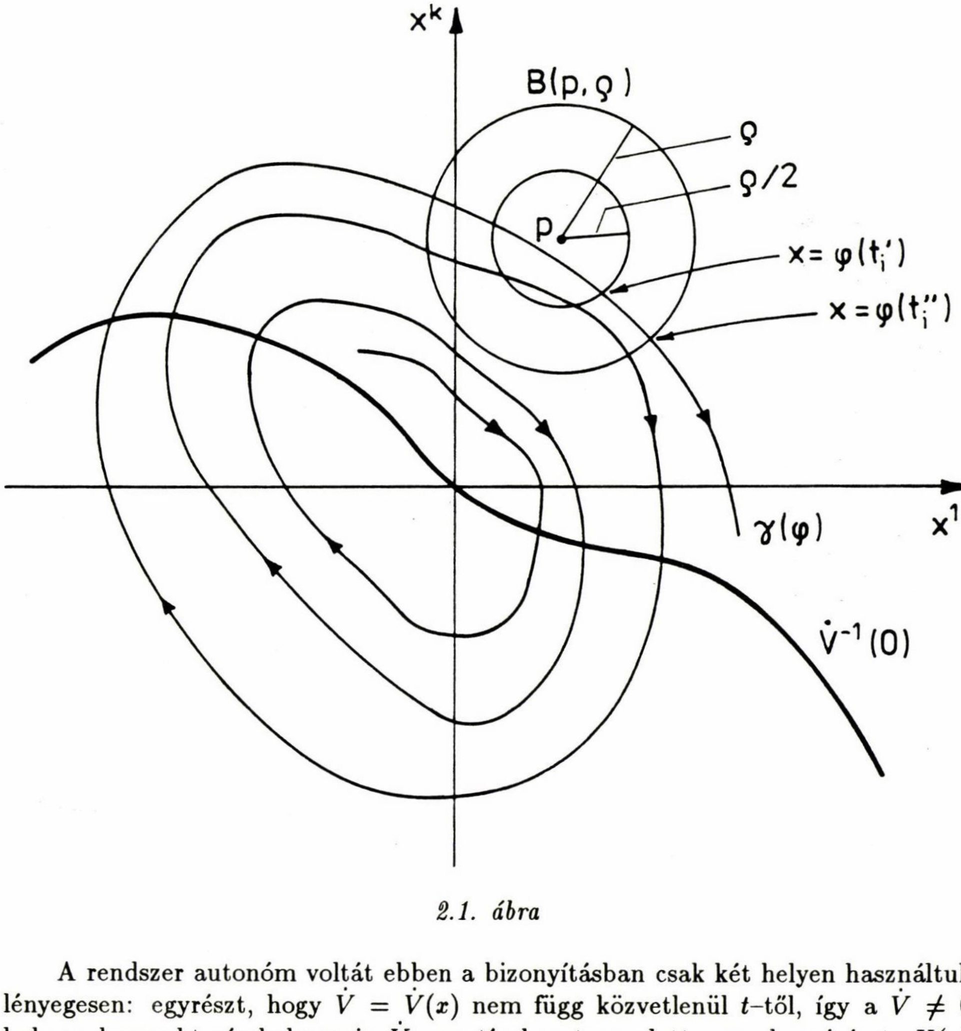
$$D^+ w(t) = \dot{V}(\varphi(t), t) - \eta(t) \leq 0,$$

tehát w csökkenő függvény. Másrészt, a lemma feltétele szerint a $\{w(t_i)\}$ sorozat alulról korlátos, így létezik a $\lim_{t\to\omega-0} w(t) =: c$ véges határérték, ahonnan a lemma állítása azonnal következik. □

2.2 Lokalizáció Ljapunov-függvény deriváltjának 0–halmazával

Az invarianciaelvnek nem-autonóm rendszerekre való kiterjesztésénél a fő nehézséget az okozza, hogy a megoldások pozitív határhalmazának invarianciája megszűnik. Most vázoljuk az invarianciaelvnek (1.2.1.* tétel) egy olyan bizonyítását az autonóm esetre, amelyben a határhalmaz invarianciáját nem használjuk ki – ez a bizonyítás kijelöli a kiterjesztés fő irányait.

Tegyük fel, hogy az állítás nem igaz. Ekkor létezik olyan $p \in \Omega(\varphi)$, $\varrho > 0$, hogy $B(p,\varrho) \cap \dot{V}^{-1}(0) = \emptyset$. Egyrészt, a $\gamma(\varphi)$ trajektóriának állandóan vissza kell térnie a p pont tetszőleges közelébe. Másrészt, $B(p,\varrho)$-ban $\dot{V}$-nak negatív maximuma van, így amíg $\gamma(\varphi)$ itt tartózkodik, $V(\varphi(t))$ gyorsan csökken; következésképpen $\gamma(\varphi)$-nek ki kell lépni $B(p,\varrho)$-ból. Tehát létezik időpontoknak $\dots < t_i' < t_i'' < t_{i+1}' < \dots$ végtelenbe tartó sorozata, hogy $\gamma(\varphi)$ a $[t_i', t_i'']$ időszakban "átmetszi" a $B(p,\varrho) \setminus B(p,\varrho/2)$ halmazt. Nyilván, $|\varphi(t_i'') - \varphi(t_i')| \geq \varrho/2$, ugyanakkor $\dot{\varphi}(t) = X(\varphi(t))$-nek a $[t_i', t_i'']$ intervallumon i-től független korlátja van, ezért $|t_i'' - t_i'| \geq \delta > 0$ $(i = 1,2,\dots)$, ahonnan $V(\varphi(t)) \to -\infty$, $t \to \infty$, ami ellentmondás.



2.1. ábra

A rendszer autonóm voltát ebben a bizonyításban csak két helyen használtuk lényegesen: egyrészt, hogy $\dot{V} = \dot{V}(x)$ nem függ közvetlenül t-től, így a $\dot{V} \neq 0$ halmaz kompakt részhalmazain $\dot{V}$ nagatív konstans alatt marad; márészt, $X(x)$ korlátos minden kompakt halmazon. Ennek köszönhető, hogy az invarianciaelv általánosítható a következő irányba: tegyük fel, hogy valamely U folytonos függvénnyel.

$$\dot{V}(x,t) \leq U(x) \leq 0 \quad ((t,x) \in \Gamma), \tag{2.3}$$

és adjuk át a $\dot{V}^{-1}(0)$ szerepét az $U^{-1}(0)$ halmaznak, továbbá tegyük fel, hogy $X(x,t)$ valamilyen értelemben korlátos a $K \times R^k$ halmazokon, ahol $K \subset R_+$ kompakt (l. LaSalle [35,36]).

Mint az alkalmazások során látni fogjuk, sok esetben mindkét feltétel túl szigorú. Alábbi kiterjesztésünkkel kezelni tudjuk majd azokat az eseteket is, amikor a *Ljapunov-függvény* deriváltjára (2.3) helyett csak egy

$$\dot{V}(x,t) \leq \alpha(t)U(x) \quad (\alpha : R_+ \to R_+, \quad U : G \to R_-)$$

alakú becslés van birtokunkban, ahol $\alpha(t)$ t-nek bármilyen nagy értékeire akár a zéró-értéket is felveheti, de „átlagban" nem kicsi:

2.2.1. *Definíció.* A mérhető $\alpha : R_+ \to R_+$ függvényt *integrálisan pozitívnak* nevezzük, ha bármely

$$S := \bigcup_{i=1}^{\infty} [a_i, b_i] \quad (a_i < b_i < a_{i+1}, \quad b_i - a_i \geq \delta > 0, \quad i = 1,2,\ldots)$$

halmaz esetén

$$\int_S \alpha = \infty .$$

A $\beta : R_+ \to R_-$ függvényt *integrálisan negatívnak* nevezzük, ha $-\beta$ integrálisan pozitív.

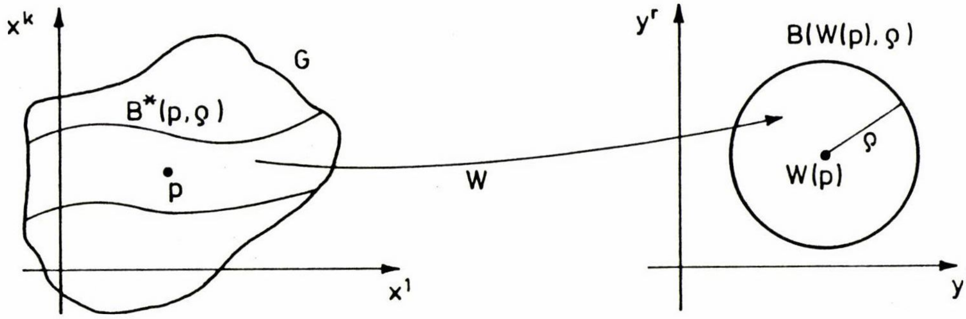
Az $\alpha(t) \equiv \alpha_0 > 0$ nyilvánvalóan integrálisan pozitív, de ugyanilyen tulajdonságú minden nem-negatív periódikus függvény is, amely egyetlen részintervallumon sem azonosan nulla. Könnyű bebizonyítani, hogy α akkor és csakis akkor integrálisan pozitív, ha bármely $\delta > 0$ szám esetén

$$\liminf_{t \to \infty} \int_t^{t+\delta} \alpha > 0 . \tag{2.4}$$

Az $X(x,t)$ korlátosságára vonatkozó feltétel azért túl szigorú, mert gyakran találkozunk olyan esetekkel (pl. a parciális stabilitásra vonatkozó problémákban), amikor az x állapothatározó helyett annak csak adott $W : R^k \to R^r$ függvényéről (pl. az állapothatározó-vektor koordinátáinak egy részéről) tudjuk, hogy a megoldások mentén nem változik „túl gyorsan". Ez abban jelentkezik, hogy $X(x,t)$ helyett a $\dot{W}(x,t)$ függvényre teljesül bizonyos korlátosság. Természetes gondolat ekkor az euklideszi $|x-p|$ távolság helyett a $|W(x)-W(p)|$ pszeudo-metrikát használni.

Legyen adva egy $W \in \mathrm{Lip}\,(G; R^r)$ függvény. Ennek felhasználásával vezessünk be egy új környezetrendszert G–ben az alábbi definícióval:

$B^*(p,\varrho)$ jelölje a $W^{-1}[B_r(W(p),\varrho)] = \{x \in G : |W(x)-W(p)| < \varrho\}$ halmaz p–t tartalmazó komponensét ($p \in G$, $\varrho > 0$).



2.2. ábra

Az invarianciaelv kiterjesztését előkészítendő most két olyan lemmát bizonyítunk, amelyek arra adnak elegendő feltételt, hogy egy p pont ne tartozzék hozzá egy adott φ megoldás pozitív határhalmazához.

2.2.2. LEMMA. *Legyen V a (2.2) egyenlethez tartozó Ljapunov-függvény Γ–n, és tekintsük az egyenlet egy tetszőleges $\Lambda \subset \Gamma$-ban haladó $\varphi : [t_0,\omega) \to R^k$ megoldását. Tegyük fel, hogy a $p \in G \cap L^*$ ponthoz létezik olyan $\varrho > 0$, $T > 0$ számpár, hogy*

$$H(p,\varrho) := \Lambda \cap (B^*(p,\varrho) \times (T,\infty))$$

összefüggő nyitott halmaz, és minden olyan $u : [T,\infty) \to R^k$ lokálisan abszolút folytonos függvényre, amyelynek grafikonja $H(p,\varrho)$–ban van, teljesülnek a következő feltételek:

(1) $\int_T^t \dot{W}(u(s),s)ds$ egyenletesen folytonos,

(2) $[\dot{V}(u(t),t)]_-$ integrálisan pozitív,

(3) $V(u(t),t)$ alulról korlátos

a $[T,\infty)$ intervallumon.

Ekkor $p \notin \Omega(\varphi)$.

2.2.3. LEMMA. *Az előző lemma állítása érvényben marad, ha az (1), (2) feltételt az alábbiakkal pótoljuk:*

(1') $\int_T^\infty |\dot{W}(u(t),t)|\,dt < \infty$;

(2') $\int\limits_T^{\infty} \dot V(u(t),t)dt = -\infty.$

Bizonyítás. A két lemmát együtt bizonyítjuk.

Tegyük fel, hogy az állítás nem igaz, vagyis $p \in \Omega(\varphi)$. Ekkor az 1.1. lemma szerint $\omega = \infty$, és létezik olyan $\{t_i\}$ sorozat, hogy $t_i \to \infty$, $\varphi(t_i) \to p$ $(i \to \infty)$. Ez azt jelenti, hogy elég nagy i–re φ grafikonjának t_i–hez tartozó pontja $H(p,\varrho)$–ban van. Másrészt, (2) – (3) miatt φ grafikonja nem maradhat valamely időponttól kezdve véglegesen $H(p,\varrho)$–ban, tehát létezik két olyan $\{t_i'\}$, $\{t_i''\}$ sorozat, hogy

$$
\begin{aligned}
&T \le t_i' < t_i'' < \dots < t_i' < t_i'' < \dots; \quad \lim_{i\to\infty} t_i' = \infty;\\
(2.5)\qquad &|W(p) - W(\varphi(t))| < \varrho \ (t_i' \le t \le t_i'')\\
&|W(\varphi(t_i'')) - W(\varphi(t_i'))| = \frac{\varrho}{2} \quad (i = 1,2,\dots).
\end{aligned}
$$

Mivel $H(p,\varrho)$ összefüggő nyitott halmaz, létezik olyan $u : R_+ \to R^k$ lokálisan abszolút folytonos függvény, amely a $[t_i', t_i'']$ intervallumokon megegyezik a φ megoldással, és grafikonja $H(p,\varrho)$–ban van. Nyilvánvalóan,

$$
(2.6)\qquad \left| \int\limits_{t_i'}^{t_i''} \dot W(u(t),t) \right| = \frac{\varrho}{2},
$$

és ez ellentmond az (1') feltételnek, mely szerint $W(u(t))$ teljes változása R_+–on véges. Ezzel a 2.6. lemmát bebizonyítottuk.

Bebizonyítjuk, hogy (2.6) az (1) – (3) feltételekkel is ellentmondásban van. Az (1) feltétel következtében létezik olyan $\delta > 0$, hogy ha $|t' - t''| < \delta$, akkor $\left|\int_{t'}^{t''} \dot W(u(s),s)ds\right| < \varrho/2$. (2.6) miatt tehát $t_i'' - t_i' \ge \delta > 0$ $(i = 1,2,\dots)$. (2) szerint $-[\dot V(u(t),t)]_-$ integrálisan negatív, így

$$
V(u(t_i''),t_i'') \le V(u(t_1'),t_1') + \int\limits_0^{\infty} \eta - \sum_{j=1}^{i} \int\limits_{t_j'}^{t_j''} [\dot V(u(t),t)]_- dt \longrightarrow -\infty \quad (i \to \infty),
$$

és ez ellentmondásban van $V(u(t),t)$ alulról való korlátosságával. □

Valójában már a fenti két lemma is közvetlenül használható határhalmaz lokalizálására, de a feltételek — különösen (2), illetve (2') — nehezen ellenőrízhetők. Ezt megkönnyíti az a körülmény, hogy a problémákban — mint már korábban említettük — a *Ljapunov-függvény* deriváltjára gyakran adható egy szeparált változójú becslés:

$$
(2.7)\qquad \dot V(x,t) \le \alpha(t)U(x) + \eta(t) \quad ((x,t) \in \Lambda),
$$

ahol $\alpha, \eta : R_+ \to R_+$, $U : \bar{L} \cap G \to R_-$ mérhetők, η integrálható R_+-on. Jelölje F az U függvény 0-halmazát:

$$F := \{x \in \bar{L} \cap G : U(x) = 0\}.$$

Ezt a halmazt sokszor „veszélyes halmaz"-nak szokták nevezni [45,46]. Ennek az az oka, hogy ha egy egyensúlyi helyzet attraktivitását szeretnénk bizonyítani (ami annyit jelent, hogy a megoldások pozitív határhalmaza egyetlen pont, éspedig az egyensúlyi helyzet), akkor egyedül F pontjai „veszélyesek", mivel — mint ahogyan az alábbi tételek mutatják — az F halmaz F^c komplementerének pontjai nem lehetnek megoldások határpontjai.

2.2.4. TÉTEL. *Legyen V a (2.2) egyenlethez tartozó Ljapunov-függvény Γ-n, legyen $\Lambda \subset \Gamma$ adott, és tegyük fel, hogy a (2.7) becslésben szereplő függvények rendelkeznek a következő tulajdonságokkal:*

(1) α integrálisan pozitív;

(2) bármely $p \in F^c$ ponthoz léteznek olyan $\varrho(p) > 0$, $T(p)$ számok, hogy

$$\sup\{U(x) : x \in S(p, \varrho) := L \cap B^*(p, \varrho)\} < 0;$$

továbbá minden olyan $u : [T, \infty) \to R^k$ lokálisan abszolút folytonos függvényre, amelynek grafikonja a

$$H(p, \varrho) := \Lambda \cap (B^*(p, \varrho) \times (T, \infty))$$

összefüggő nyitott halmazban van, teljesül, hogy

(3) $\int_T^t \dot{W}(u(s), s)ds$ egyenletesen folytonos, és

(4) $V(u(s), s)$ alulról korlátos

a $[T, \infty)$ intervallumon.

Ekkor bármely, Λ-ban haladó $\varphi : [t_0, \omega) \to R^k$ megoldásra $\Omega(\varphi) \cap G \subset F$.

Ha $L^* \subset G$ is teljesül, akkor $\varphi(t) \to F_\infty$ $(t \to \omega - 0)$. Ha φ még korlátos is, akkor $\omega = \infty$ és $\varphi(t) \to F$ $(t \to \infty)$.

2.2.5. TÉTEL. *Az előző tétel állításai érvényben maradnak, ha az (1) és (3) feltételt a következőképpen módosítjuk:*

(1') $\int_0^\infty \alpha = \infty$;

(2') $\int_T^\infty |\dot{W}(u(t), t)|\, dt < \infty$.

Bizonyítás. A két tételt együtt bizonyítjuk.

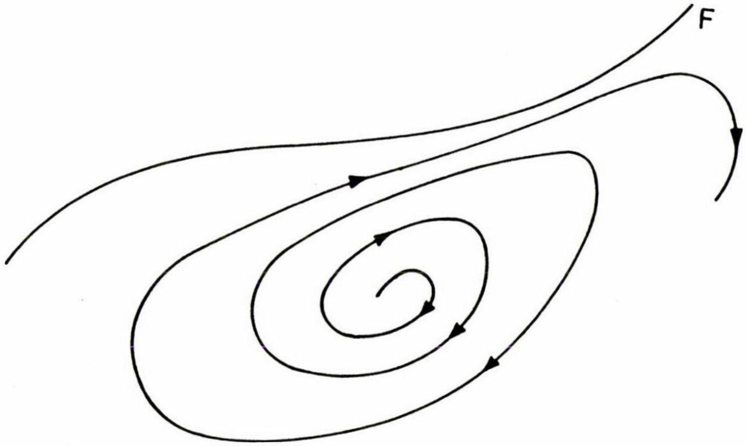
Tegyük fel, hogy az állítás nem igaz. Ekkor létezik olyan φ megoldás és $p \in \Omega(\varphi) \cap G$ pont, amely nem tartozik F-hez. A feltételek szerint léteznek olyan $\varrho(p) > 0$, $\delta(p) > 0$, $T(p)$ számok, hogy

$$\dot{V}(x, t) \leq -\delta\alpha(t) + \eta(t) \quad ((x, t) \in H(p, \varrho)).$$

Így teljesülnek a 2.5 (illetve 2.6.) lemma feltételei, ezért $p \notin \Omega(\varphi)$, ami ellentmondás.

Ha $L^* \subset G$, akkor $\Omega(\varphi) \subset G$, és ezért $\varphi(t) \to F_\infty$ $(t \to \omega - 0)$. Ha φ még korlátos is, akkor $\Omega(\varphi) \neq \emptyset$, és $\varphi(t) \to F$ $(t \to \infty)$. □

A $\varphi(t) \to F_\infty$ „konvergenciát" a 2.3. ábra szemlélteti.



2.3. ábra

A 2.2.5. tétel létjogosultságát az támasztja alá, hogy (1) $\Rightarrow$ (1'), (1') $\not\Leftarrow$ (1), és (3') $\Rightarrow$ (3), (3) $\not\Leftarrow$ (3').

2.3. Néhány következmény, példa

A 2.2.2., 2.2.3. lemmák bizonyítását analizálva könnyű látni, hogy ha W skaláris értékű függvény (azaz $r = 1$), akkor a (2.5) formulákban szereplő $\{t_i'\}$, $\{t_i''\}$ sorozatokról az utolsó sorban írott összefüggés helyett akár a

$$W(\varphi(t_i'')) - W(\varphi(t_i')) = \frac{\varrho}{2} \quad (i = 1, 2, \dots),$$

akár a

$$W(\varphi(t_i'')) - W(\varphi(t_i')) = -\frac{\varrho}{2} \quad (i = 1, 2, \dots)$$

egyenlőségek teljesülését is feltehetjük. Ennek megfelelően a kívánt ellentmondáshoz akkor is eljutunk, ha a lemmák (1), illetve (1') feltételében $\dot{W}$, illetve $|\dot{W}|$ helyébe akár a $[\dot{W}]_+$, akár a $[\dot{W}]_-$ függvényt írjuk. Ezért helytálló a következő

2.3.1. *Megjegyzés.* Ha W skaláris értékű, azaz $r = 1$, akkor a 2.2.4. illetve 2.2.5. tételben a (3), illetve (3') feltétel helyett az alábbi — gyengébb — feltétel követelhető meg: bármely $t \mapsto (u(t), t) \in H(p, \varrho)$ lokálisan abszolút folytonos függvény esetén

$$\int_T^t [\dot{W}(u(s), s)]_{+(-)} ds$$

egyenletesen folytonos R_+-on, ahol $[h(s)]_{+(-)}$ azt jelenti, hogy vagy $[h(s)]_+$ áll s minden értékére, vagy $[h(s)]_-$ áll s minden értékére. □

A tételekben szereplő feltételek szerepe, természete világosabb lesz, ha megfogalmazunk néhány következményt.

A Λ halmaz azért jelenik meg a feltételekben, hogy a lokalizálás során hasznosítani tudjuk a megoldásokról esetleg már birtokunkban lévő „á priori" ismereteket (ha ilyen információnk nincs, akkor $\Lambda = \Gamma$). Például, gyakran tudjuk, hogy minden megoldás korlátos (ennek megállapítására léteznek a lokalizációtól független módszerek [55]). Sokszor a lokalizálásra már csak akkor kerül sor, amikor az egyensúlyi helyzetről stabilitást tudunk, ahonnan a korlátosság következik. Ezekben az esetekben a feltételek az alábbi módon egyszerűsödnek.

2.3.2. KÖVETKEZMÉNY. *Legyen K pozitív szám, amelyre $\overline{B(O,K)} \subset G$, legyen V Ljapunov-függvény $B(O,K)\times R_+$-on, és tegyük fel, hogy a (2.7) becslésben szereplő függvények eleget tesznek a következő feltételeknek:*

(1) α integrálisan pozitív;

(2) bármely $p \in F^c$ ponthoz létezik olyan $\varrho(p) > 0$, hogy

$$S(p,\varrho) := B(O,K) \cap B^*(p,\varrho) \subset F^c;$$

(3) $\int\limits_0^t \sup\{|\dot{W}(x,s)| : x \in S(p,\varrho)\}ds$ egyenletesen folytonos R_+-on;

(4) V alulról korlátos az $S(p,\varrho) \times R_+$ halmazon.

Ekkor bármely $\varphi : [t_0,\infty) \to B(O,K)$ megoldásra $\varphi(t) \to F$ teljesül, ha $t \to \infty$. □

Tekinsük most a 2.2.4. tételnek azt a speciális esetét, amikor $W(x) \equiv x$. Mivel ekkor $B^*(p,\varrho) = B(p,\varrho)$, a Λ halmaz szerepe lényegtelenné válik, hiszen a (2) feltétel $\Lambda = \Gamma$ mellett is mindig teljesül.

2.2.3. KÖVETKEZMÉNY. *Tegyük fel, hogy V Ljapunov-függvény Γ-n, és a (2.7) becslésben α integrálisan pozitív. Továbbá tegyük fel, hogy ha $H \subset G$ kompakt, akkor bármely $u : R_+ \to H$ folytonos függvény esetén az $\int_0^t X(u(s),s)ds$ egyenletesen folytonos és $V(u(t),t)$ alulról korlátos R_+-on. Ekkor bármely $\varphi : [t_0,\omega) \to R^k$ megoldásra $\Omega(\varphi) \cap G \subset F$. Speciálisan, ha a φ megoldás trajektóriája prekompakt G-ben, akkor $\varphi(t) \to F$ $(t \to \infty)$.*

Ennek a következménynek $\alpha(t) \equiv 1$ speciális esete az invarianciaelv nemautonóm rendszerekre való *LaSalle-féle kiterjesztése* [36]. Annak illusztrálására, hogy a nem-állandó $\alpha(t)$ esete fontos lehet, elegendő elképzelnünk egy olyan rendszert és *Ljapunov-függvényt*, amelyekre

$$\dot{V}(x,t) = (\sin^2 t)U(x) \quad (U : G \to R_-)\,. \tag{2.8}$$

Mivel ebben az esetben *LaSalle-típusú* (2.3) becslés csak az $U(x) \equiv 0$ függvénnyel áll, a *LaSalle-féle általánosítás* nem alkalmazható. Ugyanakkor a 2.3.3. következmény igen, hiszen az $\alpha(t) = \sin^2 t$ függvény integrálisan pozitív.

A 2.2.4.–2.2.5. tételeknek a *LaSalle-féle kiterjesztésekkel* és más általánosításokkal [82,41,51,57] való összehasonlításában fontos szerepet játszik a W függvény megjelenése is. Az eddigi általánosításokban — mint ahogyan a 2.3.3. következményben is — az $\int_0^t X(u(s),s)ds$ függvény egyenletes folytonossága azt hivatott biztosítani, hogy az $x(t)$ megoldás ne változhasson ugyanakkorát bármilyen rövid idő alatt. Viszont gyakran előfordul, hogy egy aszimptotikus tulajdonság meglétét nem befolyásolja, ha $x(t)$ bizonyos irányokban gyorsan vátozik. Például, ha az $x = 0$ attraktivitása érdekel bennünket, akkor várhatóan nem zavaró, hogy $x(t)$ "gyorsan forog" az $x = 0$ körül, vagyis $X(x,t)$-nek az $x(t)$ helyzetvektorra merőleges összetevője nagy; elegendő az $(|x|^2) = 2X^T(x,t)x$ függvény korlátozása. A $W(x) = x^T x$ választással — a 2.3.1. megjegyzést is figyelembe véve — adódik az alábbi

2.3.4. KÖVETKEZMÉNY. *Tegyük fel, hogy $\Gamma = R^k \times R_+$, és V olyan Ljapunov-függvény Γ-n, amely korlátos, valahányszor x egy kompakt halmazon változik. Tegyük fel továbbá, hogy léteznek olyan folytonos α, $a : R_+ \to R_+$ függvények, amelyekre $a(0) = 0$; $a(r) > 0$, ha $r > 0$; α integrálisan pozitív, és*

$$\dot V(x,t) \le -\alpha(t)a(|x|) \quad (x \in R^k,\, t \in R_+).$$

Ha bármely r_1, r_2 $(0 < r_1 < r_2)$ számok esetén

$$\int_0^t \sup\{[X^T(x,s)x]_{+(-)} : r_1 \le |x| \le r_2\}ds$$

egyenletesen folytonos R_+-on, akkor (2.2) bármely $\varphi : [t_0,\omega) \to R^k$ megoldására vagy $\varphi(t) \to 0$, ha $t \to \infty$, vagy $\varphi(t) \to \infty$, ha $t \to \omega - 0$. □

Visszatérve a $\dot V(x,t) \le U(x) \le 0$ becslés gyengítésének kérdésére meg kell jegyeznünk, hogy LASALLE [36] és ONUCHIC [41] ebben az irányban a következőt érték el: csupán a $\dot V(x,t) \le 0$ becslést feltéve $\Omega(\varphi) \cap G \subset H$ tartalmazást bizonyítottak, ahol

$$H := \{x \in G : (\exists\{(x_i,t_i)\}_{i=1}^{\infty})\ (x_i \to x,\ t_i \to \infty,\\ \dot V(x_i,t_i) \to 0\ (i \to \infty))\}.$$

A (2.8) tulajdonsággal rendelkező *Ljapunov-függvényre* azonban ez sem mond semmit, hiszen $H = G$. Megvan viszont az az előnye, hogy a (2.7) becslés nélkül is alkalmazható. Módszerünkkel azonban ez az eredmény is élesíthető. Vezessük be az

$$F := \{x \in G : (\exists\{(x_i,t_i)\}_{i=1}^{\infty})\ (W(x_i) \to W(x),\ t_i \to \infty,\\ \dot V(x_i,t_i) - \eta(t_i) \to 0 \quad (i \to \infty))\}$$

jelölést.

A 2.2.2 lemmából a 2.2.4. tételhez hasonlóan vezethető le az alábbi

2.3.5. KÖVETKEZMÉNY. *Legyen V Ljapunov-függvény a Γ halmazon és legyen $\Lambda \subset \Gamma$ adott. Tegyük fel, hogy bármely $p \in F^c$ ponthoz léteznek olyan $\varrho(p) > 0$, $T(p)$ számok, hogy $L \cap B^*(p,\varrho)$ és F idegenek, valamint minden olyan $u : [T,\infty) \to R^r$ lokálisan abszolút függvényre, amelynek grafikonja a*

$$H(p,\varrho) := \Lambda \cap (B^*(p,\varrho) \times (T,\infty))$$

összefüggő nyitott halmazban van, teljesül, hogy $\int_T^t \dot{W}(u(s),s)ds$ egyenletesen folytonos, és $V(u(s),s)$ alulról korlátos a $[T,\infty)$ intervallumon.

Ekkor bármely, Λ-ban haladó $\varphi : [t_0,\omega) \to R^k$ megoldásra $\Omega(\varphi) \cap G \subset F$.

2.3.6. *Példa.* Tekintsük az

$$\ddot{x} + a(t)\dot{x} + b(t)f(x) = 0 \quad (x \in R) \tag{2.9}$$

egyenletet, ahol $a : R_+ \to R$, $f : R \to R$ folytonos, $b : R_+ \to R$ folytonosan differenciálható. Az $y = \dot{x}$ változó bevezetésével egyenletünk elsőrendű rendszerré írható át:

$$\dot{x} = y, \; \dot{y} = -b(t)f(x) - a(t)y\,. \tag{2.10}$$

Tekintsük a

$$V(x,y,t) := \frac{y^2}{b(t)} + 2F(x); \quad F(x) := \int_0^x f; \; W(x,y) = \left(\frac{x^2}{2}, \frac{y^2}{2}\right).$$

segédfüggvényeket. A V és W függvénynek a rendszer szerinti deriváltja:

$$\dot{V} = -\alpha(t)y^2, \; \alpha(t) := \frac{2a(t)}{b(t)} + \frac{\dot{b}(t)}{b^2(t)};$$
$$\dot{W} = (xy, -b(t)yf(x) - a(t)y^2)\,.$$

A 2.2.4. tétel következményeként az alábbi eredmény adódik:

2.3.7. KÖVETKEZMÉNY. *Tegyük fel, hogy $\int_0^t(|a|+|b|)$ egyenletesen folytonos R_+-on. Ha $b(t) > 0$, vagy valamely $\gamma > 0$ számmal $b(t) \le -\gamma$ teljesül t elég nagy értékeire, valamint α integrálisan pozitív, akkor (2.9) bármely $x : [t_0,\omega) \to R$ megoldására vagy a) $|x(t)| + |\dot{x}(t)| \to \infty$, ha $t \to \omega - 0$, vagy b) $\dot{x}(t) \to 0$, ha $t \to \infty$.* □

Ha $a(t) \ge 0$, akkor ebben az állításban szemet szúr az $\int_0^t a$ függvényre vonatkozó feltétel, hiszen mechanikai tapasztalataink azt sugallják, hogy „nagyobb" surlódás mellett a sebesség méginkább 0-hoz tart. Ezen tapasztalatainkkal összhangban álló eredményt kapunk, ha ebben az esetben segédfüggvénynek a $W(y) = y^2/2$ skalár értékű függvényt választjuk. Ennek deriváltjára érvényes a

$$[\dot{W}(x,y,t)]_+ = [-b(t)f(x)y - a(t)y^2]_+ \le [-b(t)f(x)y]_+$$

becslés. Ha alkalmas feltétellel biztosítjuk $x(t)$ korlátosságát, akkor alkalmazhatjuk a 2.2.4. tétel 2.3.1. megjegyzéssel finomított formáját. Így adódik az alábbi

2.3.8. KÖVETKEZMÉNY. *Tegyük fel, hogy* $a(t)$, $b(t) \geq 0$, *és*

$$\lim_{|x|\to\infty} F(x) = \infty$$

Ha $\int_0^t b$ *egyenletesen folytonos* R_+*-on, és* α *integrálisan pozitív, akkor (2.9) bármely* x *megoldására* $\dot{x}(t) \to 0$, *ha* $t \to \infty$. □

A lokalizációs tételek legfontosabb alkalmazását a harmadik fejezetben látjuk, ahol nem-autonóm rendszer egyensúlyi helyzetének aszimptotikus stabilitására vezetünk le elegendő feltételeket. A jelen fejezet végén az olvasó talál még egy alkalmazást 2x2 típusú lineáris rendszerekre, amelyre azért nem itt kerül sor, mert ugyanaz a rendszer a következő pontokban megfogalmazandó lokalizációs tételek illusztrálására is kiválóan alkalmas.

2.4. A 2.2.3. lemma egy általánosítása

Az eddigiekből kitűnik, hogy egy módszer nem-autonóm rendszerekre való kiterjesztése során annál jobb eredményt kapunk, minél jobban figyelembe tudjuk venni a rendszer jobboldalának az időtől való explicit függését (a *Ljapunov-függvény* is függ az időtől, a deriváltjára vonatkozó becslés is függ az időtől). Éppen ezért az olvasónak biztosan feltűnt, hogy az eredményekben oly fontos szerepet játszó W függvény nem függ az időtől. Ennek az az oka, hogy az időtől is explicite függő $W(x,t)$ szerepeltetése a kifejtést technikailag bonyolultabbá tette volna. Ugyanakkor az általunk bemutatandó módszer alapgondolata már a stacionárius $W(x)$ esetében is csonkítatlanul megjelenik. Másrészt viszont, sok esetben csak az általános eredmények alkalmazhatók, így ezekről is szólnunk kell. Utalunk azokra a pontokra, ahol lényeges változtatásokra van szükség — az eredmények pontos megfogalmazását és bizonyítását az értekezés terjedelmének korlátozottsága miatt itt mellőzzük (a részletekre vonatkozóan l. a [26] dolgozatot). Megfogalmazunk továbbá egy olyan eredményt, amely mutatja, hogy $W(x,t)$-ben a t megjelenése lényeges, és ezt a 2.3.6. példában tárgyalt egyenletre alkalmazzuk.

Legyen adva egy $W \in C^1(\Gamma; R^r)$ függvény, amelyről azt is tudjuk, hogy a $\{W(\cdot,t)\}_{t\in R_1}$ függvénycsalád elemei G-n egyenlő mértékben folytonosak. A $p \in G$ pontnak ezen W függvény segítségével definiált környezete az időben változik; a 2.2. ábra helyett most a 2.4. ábra érvényes.

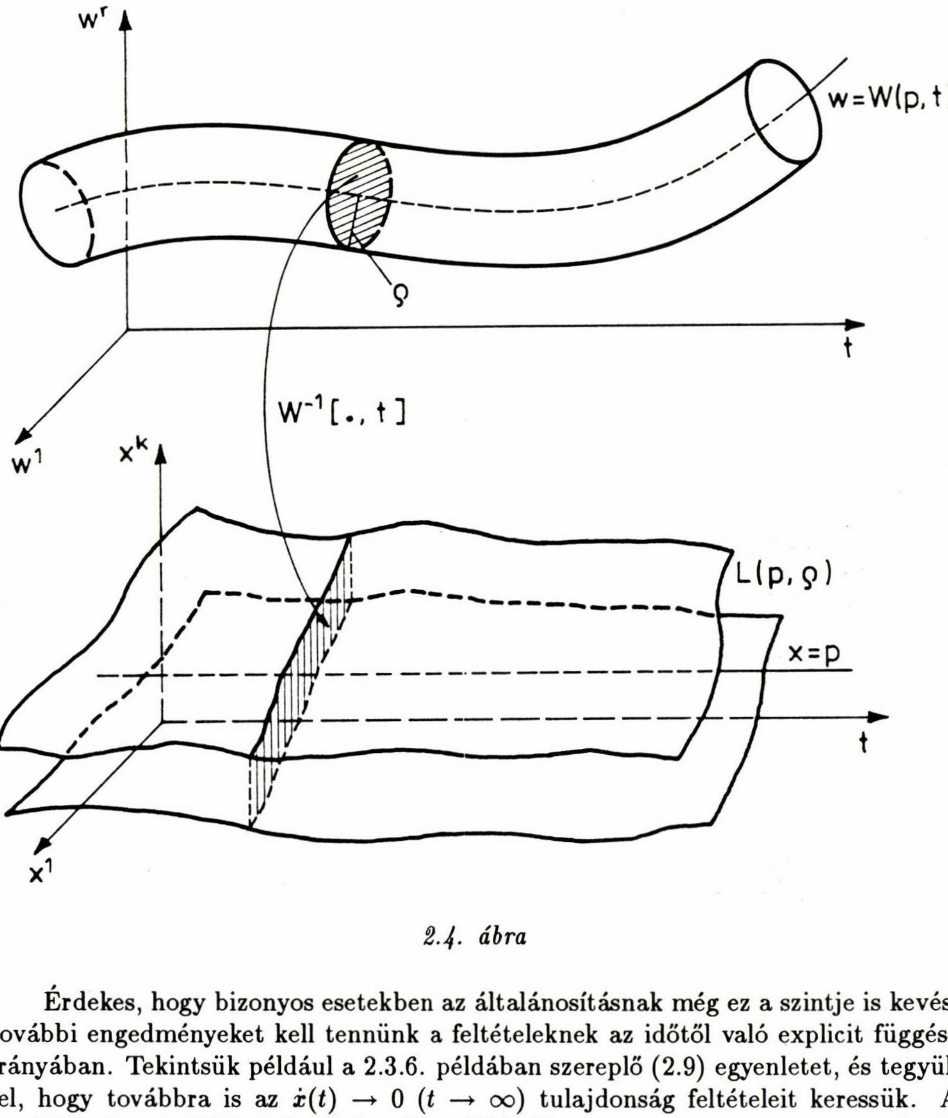
A tételek megfogalmazásában történő változások:

1) a $H(p,\varrho)$ halmaznak a 2.2.2. lemmában található definíciójában a $B^*(p,\varrho) \times (T,\infty)$ halmaz az

$$\{(x,t) \in \Gamma : |W(x,t) - W(p,t)| < \varrho\} \quad (p \in G;\ 0 < \varrho = \text{konst.}) \tag{2.11}$$

halmaznak az $x = p$ egyenest tartalmazó komponensével helyettesítendő;

2) a $\dot{W}(u(s),s)$ helyére mindenütt $\dot{W}(u(s),s) - D_t W(p,s)$ írandó.



2.4. ábra

Érdekes, hogy bizonyos esetekben az általánosításnak még ez a szintje is kevés: további engedményeket kell tennünk a feltételeknek az időtől való explicit függése irányában. Tekintsük például a 2.3.6. példában szereplő (2.9) egyenletet, és tegyük fel, hogy továbbra is az $\dot{x}(t) \to 0$ $(t \to \infty)$ tulajdonság feltételeit keressük. A korábban kapott feltételek mindegyike a $|b(t)|$ függvényt valamely értelemben felülről korlátozza. Ha $b(t) > 0$, akkor ezt úgy mondhatjuk, hogy a rezgés során a rugalmassági együttható nem nőhet akárhogy (a rendszer nem merevedhet akárhogy). Pedig a mechanikai tapasztalat azt sugallja, hogy kellő surlódással ekkor is elérhető, hogy a sebesség 0–hoz tartson.

Tekintsük hát a következő *problémát*: Legyen $b : R_+ \to R$ adott differenciálható függvény. Tegyük fel, hogy $b(t) > 0$ $(t \in R_+)$, de felülről a függvényt nem korlátozzuk, és $\dot{b}(t)$ is tetszőleges. Keressünk olyan feltételt az $a : R_+ \to R$

függvényre, amely biztosítja, hogy a (2.9) egyenlet minden megoldásának deriváltja 0-hoz tart, ha $t \to \infty$.

Tekintsük a $W(x,y,t) := y^2/2b(t)$ függvényt. Ennek (2.10) szerinti deriváltja:

$$\dot{W}(x,y,t) = -f(x)y - \left(\frac{2a(t)}{b(t)} + \frac{\dot{b}(t)}{b^2(t)}\right)\frac{y^2}{2} = -f(x)y - \alpha(t)\frac{y^2}{2}. \tag{2.12}$$

Tehát

$$\dot{W}(x,y,s) - D_t W(x_0,y_0,s) = -f(x)y - \alpha(t)\frac{y^2}{2} + \frac{\dot{b}(t)}{b^2(t)}\frac{y_0^2}{2}, \tag{2.13}$$

ami azt mutatja, hogy W-nek ezen választása megengedi a $b(t) \to \infty$ esetet. Viszont ha az $L(p,\varrho)$ halmazt a (2.11) szerint definiáljuk, akkor azt kell megkövetelnünk, hogy a $\dot{V}(u(t),t) = -\alpha(t)w^2(t)$ függvény legyen integrálisan nagatív, valahányszor az $u(t) = (v(t), w(t))$ függvény grafikonja benne van az

$$\{(x,y,t) \in R^2 \times R_+ : \frac{1}{b(t)}|y^2 - y_0^2| < \varrho\} \quad (y_0 \neq 0,\ \varrho > 0)$$

halmazok valamelyikében. Ez viszont egyetlen y_0-ra és ϱ-ra sem teljesül, hiszen a $b(t) \to \infty$ miatt akár $w(t) \equiv 0$ is lehetséges elég nagy t-re. Ezen úgy segíthetünk, ha a (2.11)-ben azt is megengedjük, hogy ϱ függjön t-től.

Foglalkozzunk a továbbiakban az egyszerűség kedvéért csak azzal az esettel, amikor W skaláris függvény: $W \in C^1(\Gamma; R)$. Adott $P \in G$ ponthoz és $\varepsilon > 0$-hoz rendeljük hozzá a

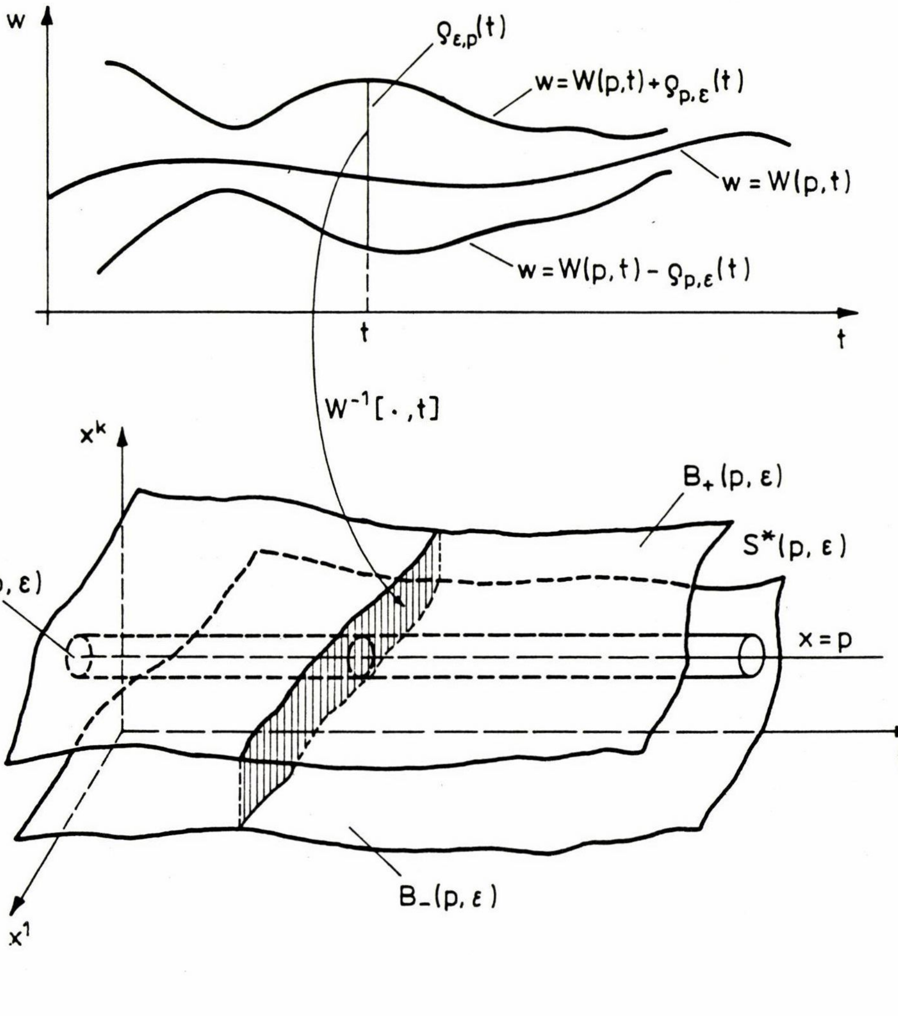
$$\varrho_{\varepsilon,p}(t) := \max\{|W(x,t) - W(p,t)| : |x - p| \leq \varepsilon\}$$

folytonos függvényt. Jelölje $S^*(p,\varepsilon)$ az

$$\{(x,t) \in \Gamma : |W(x,t) - W(p,t)| < \varrho_{\varepsilon,P}(t)\}$$

halmaz $x = p$ egyenest tartalmazó komponensét, és legyen

$$B_+(p,\varepsilon) := \{(x,t) \in \overline{S^*(p,\varepsilon)} : W(x,t) - W(p,t) = \varrho_{p,\varepsilon}(t)\},$$
$$B_-(p,\varepsilon) := \{(x,t) \in \overline{S^*(p,\varepsilon)} : W(x,t) - W(p,t) = -\varrho_{p,\varepsilon}(t)\}.$$



2.5. ábra

Az alábbi eredmény a 2.2.3. lemma általánosításának tekinthető.

2.4.1. LEMMA. *Legyen V a (2.2) egyenlethez tartozó Ljapunov-függvény Γ–n, legyen $\Lambda \subset \Gamma$ adott, és tekintsük az egyenlet egy tetszőleges $\varphi : [t_0, \infty) \to R^k$, Λ–ban haladó megoldását. Tegyük fel, hogy egy $p \in G \cap L^*$ ponthoz léteznek olyan $\mu > 0$, $T \in R_+$ számok, amelyekkel teljesülnek az alábbi feltételek: ha $u : R_+ \to R^k$ tetszőleges olyan lokálisan abszolút folytonos függvény, amelynek grafikonja a $H(p, \mu) := \Lambda \cap S^*(p, \mu)$ halmaz lezárásában halad, akkor* *(i)* $\int_0^\infty \dot{V}(u(t),t)dt = -\infty$;

(ii) $V(u(t),t)$ *alulról korlátos* R_+*-on ;*

(iii) *vagy*

$$\dot{W}(x,t) - D_t W(p,t) < D^- \varrho_{p,\mu}(t)$$

valahányszor $t \ge T$, $(x,t) \in B_+(p,\mu) \cap \Lambda$, *vagy*

$$\dot{W}(x,t) - D_t W(p,t) > D_- \varrho_{p,\mu}(t)$$

valahányszor $t \ge T$, $(x,t) \in B_+(p,\mu) \cap \Lambda$;

(iv) *vagy*

$$\dot{W}(x,t) - D_t W(p,t) < -D_- \varrho_{p,\mu})(t)$$

valahányszor $t \ge T$, $(x,t) \in B_-(p,\mu) \cap \Lambda$, *vagy*

$$\dot{W}(x,t) - D_t W(p,t) > D^- \varrho_{p,\mu}(t)$$

valahányszor $t \ge T$, $(x,t) \in B_-(p,\mu)\Lambda$.

Ekkor $p \notin \Omega(\phi)$.

A lemma bizonyítása hasonló a 2.2.2. és 2.2.3. lemma bizonyításához, csupán egyetlen lényeges pontban van eltérés. Az említett lemmákban a feltételek azt teszik lehetetlenné, hogy a megoldás grafikonja (a most érvényes jelölésekkel szólva) végtelen sokszor átmesse az $S^*(p,\mu) \setminus S^*(p,\mu/2)$ halmazt. A differenciálegyenlőtlenségek alaplemmájának ([59], 64. o.) felhasználásával be lehet látni: most a (iii)–(iv) feltétel azt biztosítja, hogy a megoldás grafikonja t nagy értékeire vagy csak beléphessen az $\overline{S^*(p,\mu)}$ halmazba, vagy csak elhagyhassa azt, ami ugyanúgy ellentmondáshoz vezet.

A részletes bizonyítást mellőzzük.

2.4.2. KÖVETKEZMÉNY. *Tegyük fel, hogy a* V *Ljapunov-függvény eleget tesz a (2.7) becslésnek, és* $\int_0^\infty \alpha = \infty$. *Legyen* $\Lambda \subset \Gamma$ *adott. Tegyük fel, hogy bármely* $p \in L^* \cap G \setminus F$ *ponthoz léteznek olyan* $\mu, \nu > 0$, $T \in R_+$ *számok, amelyekkel teljesülnek az alábbi feltételek:*

(i) *ha* $(x,t) \in H(p,\mu) := \Lambda \cap S^*(p,\mu)$ *és* $t \ge T$, *akkor* $U(x) \ge \nu$;

(ii) V *alulról korlátos a* $H(p,\mu)$ *halmazon;*

(iii)–(iv) *a 2.4.1. lemma (iii)–(iv) feltétele teljesül.*

Akkor a (2.2) egyenlet minden, Λ*-ban haladó megoldására* $\Omega(\varphi) \cap G \subset F$.

Ha még az $L^* \subset G$ *tartalmazás is igaz, akkor* $\varphi(t) \to F_\infty$ $(t \to \omega - 0)$. *Ha a* φ *megoldás korlátos, akkor* $\omega = \infty$ *és* $\varphi(t) \to F$ $(t \to \infty)$. □

Alkalmazzuk a 2.4.2. következményt a (2.9) egyenletre felvetett problémánk megoldására. Használjuk most is a 2.3.6. példában már bevezetett $V(x,y,t) := y^2/b(t) + 2F(x)$ *Ljapunov-függvényt,* amelynek (2.10) szerinti deriváltja:

$$\dot{V}(y,t) = -\alpha(t)y^2, \quad \alpha(t) := \frac{2a(t)}{b(t)} + \frac{\dot{b}(t)}{b^2(t)}.$$

Tekintsük a $W(y,t) := y^2/(2b(t))$ függvényt. Ekkor tetszőleges $p = (q,r)$ $(r \neq 0)$ pontra és μ $(0 < \mu < |r|)$ számra

$$\varrho_{p,\mu}(t) = \max\left\{\frac{1}{2b(t)}|y^2 - r^2| : |y - r| \leq \mu\right\} =$$
$$= \frac{2r\mu + \mu^2}{2b(t)} = \frac{(2\gamma + \gamma^2)r^2}{2b(t)} \quad (\gamma := \mu/|r|)\,.$$

Ha $0 < \gamma < \sqrt{2} - 1$, akkor

$$S^*(p,\mu) = \{(x,y,t) : x \in R,\ t \in R_+\,,$$
$$[2 - (1+\gamma)^2]^{1/2}|r| < |y| < (1+\gamma)|r|\,\}\,,$$

és a

$$\nu := [2 - (1+\gamma)^2]^{1/2}|r| = (1-\beta)|r|, \quad (\beta = 1 - [2 - (1+\gamma)^2]^{1/2})$$

választással a 2.4.2. következmény első két feltétele teljesül. A (iii)–(iv) feltétel biztosításához először vegyük észre, hogy

$$B_+(p,\mu) = \{(x,y,t) : y = (1+\gamma)r\}\,.$$

A (2.12)–(2.13) számolást folytatva a

$$(2.14)_+ \qquad [\dot W(x,y,t) - D_t W(p,t)]_{y=(1+\gamma)r} -$$
$$-\frac{d}{dt}\varrho_{p,\mu}(t) = -f(x)(1+\gamma)r + \frac{r^2}{2}(1+\gamma)^2\left\{-\frac{2a(t)}{b(t)} + \right.$$
$$\left. + \frac{\dot b(t)}{b^2(t)}\left[\frac{2\gamma + \gamma^2}{(1+\gamma)^2} - 1 + \frac{1}{(1+\gamma)^2}\right]\right\} \leq$$
$$\leq -r^2(1+\gamma)^2\left\{\frac{a(t)}{b(t)} - \frac{|f(x)|}{|r|(1+\gamma)}\right\}$$

becslést nyerjük.

Hasonlóan kapható, hogy

$$B_-(p,\mu) = \{(x,y,t) : y = (1-\beta)r\}$$

és

$$(2.14)_- \qquad [\dot W(x,y,t) - D_t W(p,t)]_{y=(1-\beta)r} +$$
$$+\frac{d}{dt}\varrho_{p,\mu}(t) \leq -r^2(1-\beta)^2\left\{\frac{a(t)}{b(t)} - \frac{|f(x)|}{|r|(1-\beta)}\right\}\,.$$

ÁLLÍTÁS. *Tegyük fel, hogy a (2.9) egyenletben szereplő függvényekre teljesülnek a következő feltételek:*

(i) $\lim\limits_{|x|\to\infty}\int\limits_0^x f(r)dr = \infty$;

(ii) $2a(t)b(t) + \dot{b}(t) \geq 0 \quad (t \geq 0)$;

(iii) $\lim\limits_{t\to\infty}\left[2\int\limits_0^t \frac{a(s)}{b(s)}ds - \frac{1}{b(t)}\right] = \infty$;

(iv) $\lim\limits_{t\to\infty} a(t)/b(t) = \infty$.

Ekkor (2.9) minden $x(t)$ megoldására

$$\lim_{t\to\infty}\dot{x}(t) = 0\,.$$

Bizonyítás.

Adott $K > 0$ számra definiáljuk a

$$\Lambda(K) := \{(x,y,t) : |x| < K\}$$

halmazt. Legyen $x : [t_0,\omega) \to R$ a (2.9) egyenletnek egy tetszőleges megoldása. Ekkor van olyan K, hogy ha $|x| \geq K$, akkor

$$F(x) > [\dot{x}(t_0)]^2/b(t_0) + F(x(t_0))\,.$$

Mivel $\dot{V}(x,y,t) \leq 0$, az $F(x(t))$ függvény nem vehet fel az ezen becslés jobb oldalán álló konstansnál nagyobb értéket, tehát a (2.10) egyenletrendszer megfelelő $x = x(t)$, $y = \dot{x}(t)$ megoldása $\Lambda(K)$-ban marad.

A $\Lambda = \Lambda(K)$, $F = \{(x,y) \in R^2 : y = 0\}$ választással a 2.4.2. következmény első két feltétele teljesül. Rögzített $p = (q,r)$ $(r \neq 0)$ ponthoz létezik olyan $T \in R_+$, hogy ha $t \geq T$, akkor

$$a(t)/b(t) > \max\{\,|f(x)| : |x| \leq K\}\,/\,|r|(1-\beta)\,.$$

A $(2.14)_+$ és $(2.14)_-$ becslések mutatják, hogy a (iii)–(iv) feltétel is teljesül.

A 2.4.2. következmény szerint tehát $(x(t),y(t)) \to M_\infty$ $(t \to \omega - 0)$. De $|y(t)| = |\dot{x}(t)| \not\to \infty$, mivel $x(t)$ korlátos, tehát $(x(t),y(t)) \to M$, vagyis $\dot{x}(t) \to 0$, ha $t \to \infty$. □

2.5. Lokalizáció a gyűrű–módszerrel

A 2.2. pontban ismertetett lokalizációs tételek valamilyen értelemben korlátozzák az $|X(x,t)|$ függvényt (általánosabban $|\dot{W}(x,t)| - t$). Most olyan tételeket ismertetünk, amelyekben ilyen közvetlen korlátozás nincs, pontosabban, a közvetlen korlátozás helyett a *Ljapunov-függvény* $\dot{V}(x,t)$ deriváltja és $|X(x,t)|$ (általánosabban $|\dot{W}(x,t)|$) között tételezünk fel kapcsolatot. Ez a kapcsolat, durván szólva, azt fejezi ki, hogy ha a $\varphi(t)$ megoldás (általánosabban a $W(\varphi(t))$ függvény) a veszélyes halmaz komplementerében gyorsan változik, akkor $V(\varphi(t),t)$ gyorsan csökken. Ezt a becslést lokalizálásra a gyűrőmódszerrel hasznosíthatjuk, amelynek lényege: a $p \notin \Omega(\varphi)$ relációt úgy mutatjuk meg, hogy p–t körülvesszük egy gyűrűvel, és belátjuk, hogy egyrészt φ trajektóriája nem maradhat valahonnan kezdve örökre a gyűrű belsejében, másrészt viszont a gyűrűt végtelen sokszor nem metszheti át.

A gyűrűsmódszer gyökere az alábbi egyszerű

2.5.1. LEMMA. *Legyen $a,b,$: $R_+ \to R$ lokálisan abszolút folytonos, $\eta \in L_1(R_+)$, és tegyük fel, hogy*

$$\liminf_{t\to\infty} a(t) > -\infty, \quad \liminf_{t\to\infty} \inf |b(t)| = 0\,,$$
$$\dot a(t) \le \eta(t) \quad (t \in R_+)\,.$$

Ha bármely $\varepsilon > 0$-hoz létezik olyan $\delta(\varepsilon) > 0$, $T(\varepsilon) > 0$, hogy

$$[t \ge T,\ \varepsilon \le |b(t)| \le 2\varepsilon] \Longrightarrow \dot a(t) \le -\delta[\dot b(t)]_-$$

akkor $\lim\limits_{t\to\infty} b(t) = 0.$

Bizonyítás. Tegyük fel, hogy az állítás nem igaz, vagyis van olyan B szám, amelyre

$$\limsup_{t\to\infty} |b(t)| \ge B > 0\,.$$

Az általánosság megszorítása nélkül feltehetjük, hogy $\limsup_{t\to\infty} b(t) \ge B$. Ekkor léteznek olyan $\{t_i'\}$, $\{t_i''\}$ sorozatok, hogy

$$T\left(\frac{B}{3}\right) \le t_1' < t_1'' < \dots < t_i' < t_i'' < t_{i+1}' < \dots,\ \lim_{i\to\infty} t_i' = \infty\,,$$
$$b(t_i') = \frac{2B}{3},\ b(t_i'') = \frac{B}{3},\ [t_i' \le t \le t_i''] \Longrightarrow \frac{B}{3} \le b(t) \le \frac{2B}{3} \quad (i = 1,2,\dots)\,.$$

A feltételek miatt

$$a(t_j'') \le \int_{t_1'}^{t_j''} \dot a \le a(t_1') + \int_0^\infty \eta - \delta\left(\frac{B}{3}\right) \sum_{i=1}^{j} \int_{t_i'}^{t_i''} [\dot b]_- \le$$
$$\le \text{konst.} - \delta\left(\frac{B}{3}\right)\cdot j\frac{B}{3} \longrightarrow -\infty \ (j \to \infty)\,,$$

ami ellentmondás. $\square$

Most elegendő feltételt adunk arra, hogy egy pont ne tartozzon egy megoldás pozitív határhalmazához.

Legyen adva egy $W \in \mathrm{Lip}(G; R^r)$ függvény, és tekintsük ismét a 2.2. pontban már használt környezetrendszert G–ben: $B^*(p,\varrho)$ jelölje a $W^{-1}[B_r(W(p),\varrho)]$ halmaz p–t tartalmazó komponensét ($p \in G$, $\varrho > 0$) (l. a 2.2. ábrát).

2.5.2. LEMMA. *Legyen V a (2.2.) egyenlethez tartozó Ljapunov-függvény Γ–n, és $\varphi : [t_0, \omega) \to R^k$ az egyenletnek egy $\Lambda \subset \Gamma$–ban haladó megoldása. Tegyük fel, hogy a $p \in G \cap L^*$ ponthoz létezik olyan $\varrho > 0$, $\delta > 0$, $T > 0$, hogy*

$$\text{(2.15)} \qquad \begin{aligned} &(1) \quad \dot V(x,t) \le -\delta|\dot W(x,t)| + \eta(t) \\ &(2) \quad V(x,t) \text{ alulról korlátos} \end{aligned}$$

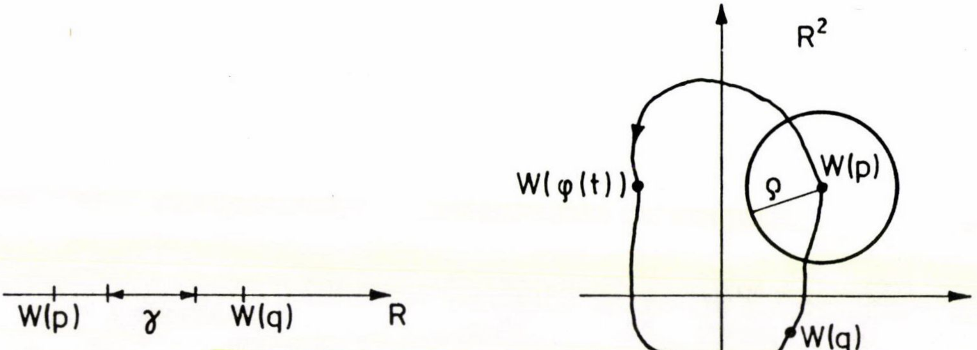
a $\Lambda \cap (B^*(p,\varrho) \times [T,\infty))$ halmazon.

Ekkor vagy a) $p \notin \Omega(\varphi)$, vagy b) $\omega = \infty$, és $\Omega(\varphi) \cap G \subset W^{-1}[W(p)]$.

Bizonyítás. Tegyük fel, hogy a) nem teljesül, vagyis $p \in \Omega(\varphi)$. Mivel $p \in G$ és G nyílt, a megoldások folytathatóságáról szóló tétel szerint $\omega = \infty$. Tegyük fel, hogy b) is hamis, vagyis létezik olyan $q \in \Omega(\varphi) \cap G$, amelyre $W(q) \ne W(p)$. Ekkor van olyan j ($1 \le j \le r$) természetes szám, hogy az

$$a(t) := V(\varphi(t), t), \quad b(t) := W^j(\varphi(t)) - W^j(p)$$

függvények egyrészt teljesítik a 2.5.1. lemma feltételeit, másrészt $\limsup_{t\to\infty} |b(t)| > 0$, ami ellentmondás. □



2.6. ábra

Ha W skaláris ($r = 1$), akkor a $W(\varphi(t))$ függvényérték csak úgy kerülhet közel $W(p)$–hez, majd $W(q) \ne W(p)$–hez, hogy közben a $W(\varphi(t))$ egy intervallumon egy pozitív γ–val nő, egy másikon pedig γ–val csökken úgy, hogy ezeken az intervallumokon $W(\varphi(t))$ közel van $W(p)$–hez ($\varphi(t) \in B^*(p,\varrho)$). (Ez az $r > 1$ vektoriális esetben a $W^j(\varphi(t))$ komponensekre nem igaz, l. a 2.6. ábrát: $W^2(\varphi(t))$ csak növekszik, amikor $W(\varphi(t))$ $W(p)$–hez ϱ–nál közelebb van.) Ezért igaz a következő

2.5.3. *Megjegyzés.* Ha a 2.5.2. lemmában $W(x)$ skaláris ($r = 1$), vagy pedig az (1)–(2) feltételeket az egész Γ–n megköveteljük, akkor az (1) feltételben $\dot{W}$ helyére $[\dot{W}]_+$ is írható. (Az utóbbi esetben az állítás úgy módosul, hogy vagy a) $\Omega(\varphi) \cap G$ üres, vagy b) $\omega = \infty$, és létezik olyan $p \in G \cap L^*$, hogy $\Omega(\varphi) \cap G \subset W^{-1}[W(p)]$.)

2.5.4. TÉTEL. *Legyen V a (2.2.) egyenlethez tartozó Ljapunov-függvény Γ–n, és legyen $\Lambda \subset \Gamma$, $H \subset G$ adott. Tegyük fel, hogy bármely $p \in H^c$–hez léteznek olyan $\varrho(p) > 0$, $\delta(p) > 0$, $T(p)$ számok, amelyekkel a 2.5.2. lemma feltételei teljesülnek a $\Lambda \cap (B^*(p,\varrho) \times [T,\infty))$ halmazon. Tekintsük a (2.2.) egyenlet egy $\varphi : [t_0,\omega) \to R^k$, Λ–ban haladó megoldását. Ekkor:*

1) vagy a) $\Omega(\varphi) \cap G \subset H$, vagy b) $\omega = \infty$, és létezik olyan $d \in R^r$, hogy $W^{-1}[d] \cap H^c$ nem üres és $\Omega(\varphi) \cap G \subset W^{-1}[d]$.

2) Ha $L^* \subset G$ is teljesül, akkor vagy a) $\varphi(t) \to H_\infty$ ($t \to \omega - 0$), vagy b) $\omega = \infty$, és létezik olyan $d \in R^r$, hogy $W^{-1}[d] \cap H^c$ nem üres, és $\varphi(t) \to W^{-1}[d]_\infty$ ($t \to \infty$). Nevezetesen, ha φ korlátos, akkor $\omega = \infty$, és vagy a) $\varphi(t) \to H$, vagy b) $W(\varphi(t))$–nek létezik véges határértéke, ha $t \to \infty$.

Skaláris W esetén ($r = 1$) az állítások igazak maradnak, ha a $\dot{W}$ függvényt a $[\dot{W}]_+$–szal pótoljuk.

Bizonyítás. Ha $\Omega(\varphi) \cap G$ üres, akkor a) teljesül. Tegyük fel, hogy ez a halmaz nem üres, és van olyan p pontja, amely H komplementerében van. A 2.5.2. lemma szerint ekkor b) teljesül (nevezetesen, $d = W(p)$).

Az utolsó állításra vonatkozóan l. a 2.5.3. megjegyzést. □

2.5.5. TÉTEL. *Legyen V a (2.2.) egyenlethez tartozó Ljapunov-függvény Γ–n, és tegyük fel, hogy*

(1) $\dot{V}(x,t) \le -\delta |[\dot{W}]_+| + \eta(t)$,

(2) $V(x,t)$ alulról korlátos

egy $\Lambda \subset \Gamma$ halmazon; $0 < \delta = konst.$

Ekkor a (2.2) egyenlet tetszőleges $\varphi : [t_0,\omega) \to R^k$, Λ–ban haladó megoldására vagy

a) $\Omega(\varphi) \cap G$ üres, vagy

b) $\omega = \infty$, és létezik olyan $d \in R^r$, hogy $\Omega(\varphi) \cap G \subset W^{-1}[d]$. □

Ez a tétel a 2.5.3. megjegyzés felhasználásával ugyanúgy bizonyítandó, mint a 2.5.4. tétel.

2.6. Alkalmazások, példák

A gyűrűmódszerrel való lokalizálás legfontosabb alkalmazására a mechanikai rendszerekkel kapcsolatban az aszimptotikus megállás feltételeinek tanulmányozásánál kerül sor, itt csak a módszert illusztráló példákra, az eredmények elhelyezését könnyítő speciális esetekre térünk ki.

I. Alkalmazzuk a 2.5.4. tételt a $\Gamma := \Lambda := R^k \times R_+$, $H := \{0\}$, $W(x) := |x|^2 = x^T x$ esetre.

2.6.1. KÖVETKEZMÉNY. *Legyen* $V : R^k \times R_+ \to R$ *a (2.2) egyenlethez tartozó Ljapunov–függvény, és tegyük fel, hogy* $V(x,t)$ *alulról korlátos bármely* $K \times R_+$ *hengeren, ahol* $K \subset R^k$ *kompakt. Ha bármely* $0 \neq p \in R^k$ *ponthoz létezik olyan* $\delta > 0$, $\varrho > 0$, T, *hogy*

$$\dot{V}(x,t) \leq -\delta[X^T(x,t)x]_{+(-)} + \eta(t)$$

az $\big||x| - |p|\big| < \varrho$, $t \geq T$ *halmazon, akkor (2.2) bármely* $\varphi : [t_0, \omega) \to R^k$ *megoldása normájának létezik véges vagy végtelen határértéke, ha* $t \to \omega - 0$.

II. A gyűrűmódszer alapján nyerhető tételek prototípusának V. MARACSKOV [67] azon nevezetes eredménye tekinthető, amelyben bebizonyítja, hogy LJAPUNOV-nak az aszimptotikus stabilitásról szóló, a bevezetésben idézett klasszikus tételében a $V(x,t) \rightrightarrows 0$ ($x \to 0$; $t \in R_+$) feltétel a következővel helyettesíthető: $|X(x,t)|$ korlátos minden $K \times R_+$ halmazon, ahol $K \subset R^r$ kompakt. Ezt az ötletet T. A. BURTON [9] és J. R. HADDOCK [14] fejlesztették tovább. Legélesebb eredményeik egyike a 2.5.4. tételből a $W(x) := x$ választással adódó

2.6.2.* KÖVETKEZMÉNY. *(J. R. HADDOCK [14]). Legyen* $V : \Gamma = R^k \times R_+ \to R$ *a (2.2) egyenlethez tartozó Ljapunov-függvény, amely alulról korlátos bármely* $K \times R_+$ *hengeren, ahol* $K \subset R^k$ *kompakt. Legyen* $H \subset R^k$ *adott zárt halmaz. Tegyük fel, hogy bármely* $\varepsilon > 0$*-hoz és* $K \subset R^k$ *kompakt halmazhoz létezik olyan* $\delta(\varepsilon, K) > 0$, $T(\varepsilon, K)$ *hogy a* $t \geq T$, $x \in B^c(H, \varepsilon) \cap K$ *halmazon teljesül a*

$$\dot{V}(x,t) \leq -\delta|X(x,t)| + \eta(t) \tag{2.16}$$

egyenlőtlenség.

Ha $\varphi : [t_0, \omega) \to R^k$ *megoldása (2.2)-nek, akkor vagy a)* $\varphi(t) \to H_\infty$, *vagy b)* $\varphi(t)$ *H komplementerének egy véges pontjához tart, ha* $t \to \infty$. □

Ebből a következményből könnyen levezethető *Maracskov tétele.* Az $x = 0$ stabilitása következik *Ljapunov tételéből* ([46], 21. o.), így az $x = 0$ attraktivitását kell csak belátni. Legyen $H = \{0\}$, $\varepsilon > 0$, $K \subset R^r$ kompakt. $\dot{V}$ negatív definit, ezért

$$m := \sup\{\dot{V}(x,t) : t \geq 0, \ \ x \in B^c(0,\varepsilon) \cap K\} < 0\,.$$

Másrészt, az X–re tett korlátossági feltétel miatt

$$M(K) := \sup\{|X(x,t)| : t \geq 0, \ \ x \in K\} < \infty\,,$$

így (2.16) teljesül az $\eta(t) \equiv 0$, $\delta(\varepsilon, K) := M(K)/(-m(\varepsilon, K))$ választással. A 2.6.2. következmény szerint minden elég kicsiny $|\varphi(t_0)|$ kezdeti értékkel induló $\varphi(t)$ megoldásnak létezik véges határértéke (az $x = 0$ stabilitása miatt $\varphi(t)$ korlátos!). Ez a határérték csak az $x = 0$ lehet, hiszen ha a φ megoldás trajektóriája egyenletesen távol maradna az origótól, akkor $\dot{V}$ negatív definit volta miatt $V(\varphi(t), t)$

negatív konstans alatt maradna, és így $V(\varphi(t), t) \to -\infty$ $(t \to \infty)$ teljesülne, ami lehetetlen.

Könnyű arra példát mutatni (l.még ebben a pontban), hogy (2.16) nem teljesül, de ha az $|X|$ függvényt $[X^i]_+$-vel vagy $[X^i]_-$-vel helyettesítjük, akkor igen $(1 \leq i \leq k)$. Azt várjuk, hogy ekkor a $\varphi(t) \not\to H_\infty$ esetben a φ megoldásnak csak a φ^i komponenséről van információnk. Ez valóban így van:

2.6.3. KÖVETKEZMÉNY. *Legyen* $V : R^k \times R_+ \to R$ *a (2.2) egyenlethez tartozó Ljapunov-függvény, és legyen* $H \subset R^k$ *adott. Tegyük fel, hogy bármely* $p \in H^c$*-hez létezik olyan* $\varrho(p) > 0$, $\delta(p) > 0$, $T(p)$, *hogy*

(1) $\dot{V}(x,t) \leq -\delta[X^i(x,t)]_{+(-)} + \eta(t)$,
(2) $V(x,t)$ *alulról korlátos*

az $\{(x,t) : |x^i - p^i| < \varrho,\ t \geq T\}$ *halmazon.*

Ha $\varphi : [t_0, \infty) \to R^k$ *megoldása (2.2)-nek, akkor vagy a)* $\varphi(t) \to H_\infty$, *vagy b) létezik olyan* $p \in H^c$, *hogy* $\varphi^i(t) \to p^i$ $(t \to \infty)$.

III. Az alábbi példa alkalmas a 2.2. és 2.4. pontban közölt módszerek együttes illusztrálására. Egyben mutatja azt is, hogy azok hogyan egészítik ki egymást az alkalmazások során, továbbá módot ad az eredményeknek korábbi tételekkel való összevetésére is.

Tekintsük a

$$\begin{aligned} \dot{x} &= -r(t)x + q(t)y \\ \dot{y} &= -q(t)x - p(t)y \quad ((x,y) \in R) \end{aligned} \tag{2.17}$$

lineáris nem-autonóm differenciálegyenlet-rendszert, ahol $p, r : R_+ \to R_+$, $q : R_+ \to R$ folytonos függvények, és válasszuk a $V(x,y) = (x^2 + y^2)/2$ *Ljapunov-függvényt.* Ennek deriváltja $\dot{V}(x,y,t) = -r(t)x^2 - p(t)y^2 \leq 0$, tehát minden megoldás korlátos, sőt azt is tudjuk, hogy minden megoldás (nem üres) pozitív határhalmaza egy origó körüli körön fekszik, (l. 2.1.1. lemma) ; más szóval, az $x^2(t) + y^2(t)$ függvénynek létezik véges határértéke.

J. P. LASALLE az invarianciaelv nem-autonóm rendszerekre való kiterjesztését felhasználva a (2.17) rendszer egy speciális esetére az alábbit bizonyította:

A) (J. P. LASALLE [35]: $r(t) \equiv 0$, $q(t) \equiv 1$). Ha alkalmas c, C konstansokkal $0 < c \leq p(t) \leq C$ $(t \in R_+)$ teljesül, akkor minden megoldásra $x(t) \to$ konst., $y(t) \to 0$, ha $t \to \infty$.

A (2.17) rendszer ebben a speciális esetben ekvivalens az

$$\ddot{x} + p(t)\dot{x} + x = 0$$

csillapított rezgőmozgással $(y = \dot{x})$. Az állítás úgy is fogalmazható, hogy a mozgó pont aszimptotikusan megáll a surlódás hatására. Ez az interpretáció azonnal sugallja, hogy a $p(t) \leq C$ korlátozás aligha szükséges, hiszen a nagyobb surlódás csak elősegíti ezen jelenség bekövetkeztét. Az alsó becslés már nem tűnik feleslegesnek, de úgy érezzük, hogy a $p(t)$ surlódási együtthatónak nem kell egyenletesen nagynak lenni — helyette valamilyen értelemben az összsurlódásnak kell jelentékenynek lenni.

J. R. HADDOCK egy valamivel általánosabb esetben a következőt bizonyította:

B) (J. R. HADDOCK [14]: $r(t) \equiv 0$). Ha létezik olyan $\beta > 0$, hogy

$$|q(t)| < \beta p(t) \quad (t \in R_+) \tag{2.18}$$

és $\int_0^\infty p = \infty$, akkor bármely megoldásra $x(t) \to$ konst., $y(t) \to 0$ $(t \to \infty)$.

Terjesszük ki a B) állítást az általános $r(t) \not\equiv 0$ esetre. Legyen $W(y) := y^2/2$; ennek a (2.17) rendszer szerinti deriváltja: $\dot W(x,y,t) = -q(t)xy - p(t)y^2$. Jelölje H az x tengely pontjait az (x,y)-síkon, és vezessük be a

$$\Lambda = \Lambda_c := \{(x,y,t) \in R^2 \times R_+ : x^2 + y^2 \le C^2\} \quad (0 < C = \text{konst.}).$$

A 2.5.4. tétel alkalmazásához azt kell belátni, hogy bármely ε-hoz $(0 < \varepsilon < c)$ létezik olyan $\delta > 0$, hogy ha $|y| \ge \varepsilon$, $(x,y,t) \in \Lambda_c$, akkor $\dot V(x,y,t) \le$ $\le -\delta[\dot W(x,y,t)]_+$. (2.18) felhasználásával a

$$[W(x,y,t)]_+ \le |q(t)|\frac{x^2+x^2}{2} \le |q(t)|\frac{C^2+1}{2\varepsilon^2} \le \beta\frac{C^2}{\varepsilon^2}p(t)$$

becslés adódik, tehát a $\delta := \varepsilon^2/(\beta C^2)$ választás megfelelő. Ezzel bebizonyítottuk a következő állítást:

2.6.4. KÖVETKEZMÉNY. *Tegyük fel, hogy (2.18) teljesül. Ekkor (2.17) minden megoldásának létezik véges határértéke, ha $t \to \infty$. Ha $\int_0^\infty p = \infty$, akkor $x(t) \to$ konst., $y(t) \to 0$ $(t \to \infty)$.* □

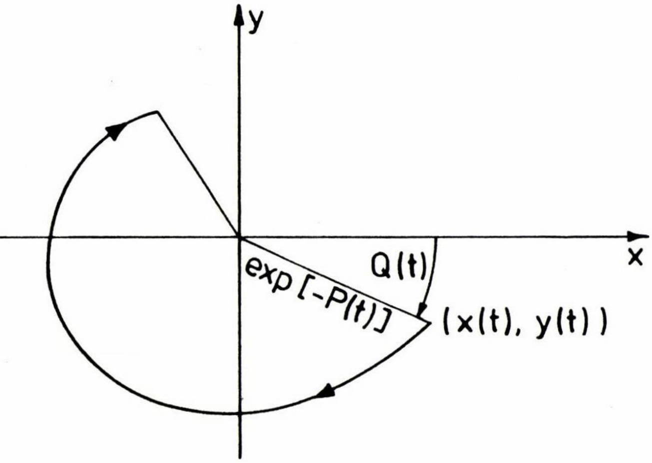
Ha a kiterjesztést HADDOCK tételével végeztük volna (azaz a $W(x,y) = (x,y)$ választással élünk), akkor az $r(t)$ függvényre is fel kellett volna tételezni a (2.18) egyenlőtlenség megfelelőjét — mint kiderült — feleslegesen.

Ismeretes [30, 534. o.] hogy a $p(t) \equiv r(t)$ $(t \in R_+)$ esetben a (2.17) egyenletrendszer általános megoldása:

$$\begin{aligned} x &= (c_1 \cos Q(t) + c_2 \sin Q(t)) \exp\,[-P(t)] \\ y &= (-c_1 \sin Q(t) + c_2 \cos Q(t)) \exp\,[-P(t)], \end{aligned} \tag{2.19}$$

ahol $Q(t) := \int_0^t q$, $P(t) := \int_0^t p$.

Ebből látszik, hogy a (2.18) feltétel nem engedhető el (az $r(t) \equiv p(t) \equiv 0$, $\int_0^\infty q = \infty$ esetben nem létezik a megoldások határértéke). Ugyanakkor úgy tűnik, hogy a (2.18) feltétel nem szükséges, hiszen az állítás a megoldások aszimptotikus viselkedéséről szól, a feltétel mégis egy egyenlőtlenséget követel meg *pontonként* valahonnan



2.7. ábra

kezdve. Az a sejtésem, hogy a 2.6.4. következményben (2.18) helyettesíthető az enyhébb

$$\int_0^t |q| < \beta \int_0^t p \quad (t \in R_+)$$

feltétellel.

A $p \equiv r$ esetben az 2.6.4. következmény az élesebb $x(t) \to 0$, $y(t) \to 0$ $(t \to \infty)$ konklúzióval is igaz, hiszen ekkor a 2.6.4. következményhez hasonlóan $x(t) \to 0$, $y(t) \to$ konst. is levezethető.

A 2.2.4. és 2.2.5. tétel alkalmazásával — a p, illetve q függvényre tett megszorítás szigorítása révén — teljesen meg tudunk szabadulni a (2.18) feltételtől. Valóban, legyen V, W, Λ_c ugyanaz, mint az előbb, továbbá $\alpha(t) := p(t)$, $U(y) := -y^2$. A Λ_c halmazon

$$[\dot{W}(x,y,t)]_+ \le [q(t)xy - p(t)y^2]_+ \le |q(t)|\frac{C^2}{2},$$

így az említett tételekből — a 2.3.1. megjegyzést is figyelembe véve — adódnak az alábbi állítások:

2.6.5. KÖVETKEZMÉNY. *Ha p integrálisan pozitív, és $\int_0^t |q|$ egyenletesen folytonos R_+-on, akkor (2.17) bármely megoldására $x(t) \to$ konst., $y(t) \to 0$ $(t \to \infty)$.* □

2.6.6. KÖVETKEZMÉNY. *Ha $\int_0^\infty p = \infty$ és $\int_0^\infty |q| < \infty$, akkor (2.17) bármely megoldására $x(t) \to$ konst., $y(t) \to 0$ $(t \to \infty)$.* □

A 2.6.6. következmény még a $p \equiv r$ speciális esetben is éles abban az értelemben, hogy ha $\int_0^\infty |q| < \infty$, akkor $\int_0^\infty p = \infty$ szükséges is az állítás teljesüléséhez.

Megjegyzés. Hasonlóan a 2.4. pont eredményeihez, a gyűrűmódszert is ki lehet terjeszteni úgy, hogy az időtől is explicit módon függő $W(x,t)$ függvények is megengedettek legyenek.

2.7 Kiegészítő megjegyzések

A fejezet 2.2.–2.3. és 2.5.–2.6. pontjai [20] dolgozatom eredményeit dolgozzák fel. Ezeket a lokalizációs tételeket azóta többen alkalmazták és fejlesztették tovább [40, 51, 56–58]. A gyűrűmódszer tételeit mechanikai rendszerekre alkalmazva TERJÉKI JÓZSEFfel közösen kidolgoztuk az aszimptotikus megállás elméletét [77,78]. (Azt mondjuk, hogy egy mechanikai rendszer aszimptotikusan megáll, ha a pozíciós koordinátáknak létezik véges határértéke, és a sebesség-koordináták 0–hoz tartanak, ha $t \to \infty$. A *Ljapunov-féle aszimptotikus stabilitás* ennek az a speciális esete, amikor a pozíciós koordináták határértéke minden mozgás mentén ugyanaz.) A kapott eredmények jó lehetőséget adnak a száraz és viszkózus surlódás összehasonlítására.

A 4. pont [26] dolgozatom eredményeinek felhasználásával készült.

3. fejezet

Aszimptotikus stabilitást és instabilitást biztosító feltételek több Ljapunov–függvénnyel

Már a bevezetésből tudjuk, hogy az invariancia-elv azon *Barbasin-Kraszovszkij-tétel* bizonyításának analíziséből született, amely autonóm differenciálegyenletrendszer egyensúlyi helyzetének aszimptotikus stabilitására ad elegendő feltételt. Természetes tehát, hogy az invariancia-elvnek az előző fejezetben kapott, a nem-autonóm rendszerek megoldásai pozitív határhalmazát lokalizáló általánosításának felhasználásával aszimptotikus stabilitást, illetve instabilitást biztosító eredmények nyerhetők a nem-autonóm esetre. Észre kell azonban venni, hogy míg az autonóm esetben a határhalmazok lokalizálása az egyedüli lényeges momentum (a határhalmazok invarianciája miatt a $\dot{v}^{-1}(0)$ „veszélyes halmazról" elegendő kikötni, hogy nem tartalmaz teljes trajektóriát az $x = 0$ ponton kívül), addig a nem-autonóm esetben a veszélyes halmaz (l. az F halmazt a 2.1. tételben) az origó tetszőleges környezetén kívül eső részének környékéről a trajektóriákat ki kell még valahogyan „tiltani". Ehhez V. M. MATROSZOV [69]-ot követve egy további *Ljupanov-típusú segédfüggvényt* használunk fel.

Tételeinket parciális aszimptotikus stabilitásra és instabilitásra mondjuk ki, mivel a mechanikai alkalmazásokban ennek külön jelentősége lesz. Megjegyezzük azonban, hogy tételeink a korábbi eredmények élesítéseit és általánosításait adják akkor is, ha az összes változókra vonatkozó tulajdonságokra (a *Ljupanov-tulajdonságokra*) szorítkozunk, így ezzel a speciális esettel néhányszor külön következmény formájában foglalkozunk. Ezt illusztrálja az első alkalmazás is, amelyben másodrendű nem-lineáris differenciálegyenlet triviális megoldásának aszimptotikus stabilitására keresünk feltételeket.

További alkalmazásként több szabadsági fokú, surlódási erő hatása alatt álló holonóm mechanikai rendszer egyensúlyi helyzetének parciális stabilitási tulajdonságait vizsgáljuk.

3.1. Alaptételek

Ismét szükségünk lesz egy W segédfüggvény által generált környezetrendszerre, de most W csak az ellenőrzött változóktól függ, amit a jelölésben is hangsúlyoznunk kell. Legyen $h' > 0$ és $W \in \mathrm{Lip}\,(B_m(0,h');\ R^j)$ adott. Ha $K \subset B_m(0,h')$ és $\rho > 0$, akkor $B_m^*(K,\rho)$ jelölje a $W^{-1}[B_j(W(K),\rho)]$ halmaz azon komponenseinek egyesítését, amelynek K-val van közös pontja.

A bevezetésben mondottakkal összhangban, differenciálegyenlet-rendszerünk particionált alakja:

$$\dot{y} = Y(y,z,t),\ \dot{z} = Z(y,z,t), \tag{3.1}$$

és feltesszük, hogy $Y(0,0,t) \equiv 0$, $Z(0,0,t) \equiv 0$ $(t \in R_+)$, vagyis $y = 0$, $z = 0$ megoldása (3.1)-nek.

Rögzített h' $(0 < h' < h)$ számra vezessük be az

$$M_z(t;t_0) := \bigcup_{|x_0|<h'} \{z(t;x_0,t_0)\} \quad (0 \le t_0 \le t)$$

$$M_z(t) := \bigcup_{t_0 \in [0,t]} M_z(t;t_0)$$

$$\Gamma_m'' := \{(x,t) : t \in R_+,\ |y| \le h',\ z \in M_z(t)\}$$

jelöléseket. Szükségünk lesz egy, az egyenletes y-stabilitásnál valamivel erősebb tulajdonságra is.

3.1.1. *Definíció.* Azt mondjuk, hogy (3.1) 0-megoldása *erősen y-stabilis,* ha bármely $\epsilon > 0$ esetén létezik olyan $\delta(\epsilon) > 0$, hogy ha $t_0 \in R_+$, $|y_0| < \delta(\epsilon)$, $z_0 \in M_z(t_0)$, akkor $|y(t;x_0,t_0)| < \epsilon$ a $[t_0,\infty)$ intervallumon.

Ennek a tulajdonságnak adja egy elegendő feltételét az alábbi

3.1.2* LEMMA. ([73], 1.3. TÉTEL). *Tegyük fel, hogy létezik a (3.1) rendszerhez olyan $V \in Lip_x(\Gamma_m''; R)$ függvény, amely rendelkezik a következő tulajdonságokkal:*

(1) alkalmas $a, b \in \mathcal{K}$ függvényekkel teljesül az

$$a(|y|) \le V(y,z,t) \le b(|y|)$$

egyenlőtlenségpár a Γ_m'' halmazon;

(2) $\dot{V}(x,t) \leq 0, \quad ((x,t) \in \Gamma''_m)$.

Ekkor a (3.1) rendszer 0-megoldása erősen y-stabilis.

3.1.3. LEMMA. *Ha (3.1) 0-megoldása erősen y-stabilis, és bármely* $t_0 \in R_+$*-hoz létezik olyan* $\sigma(t_0) > 0$*, hogy tetszőleges* $x_0 \in \bar{B}_k(0, \sigma(t_0))$ *pontra az* $|y(t; x_0, t_0)|$ *függvény tetszőlegesen kicsiny pozitív értékeket is felvesz a* $[t_0, \infty)$ *intervallumon, akkor a 0-megoldás ekvi-aszimptotikusan y-stabilis.*

Bizonyítás. Tekintsük egy rögzített $\epsilon > 0$ -hoz az erős y-stabilitás értelmében tartozó $\delta(\epsilon) > 0$ számot. A lemma feltételei szerint minden $t_0 \in R_+$-hoz létezik $\sigma(t_0) > 0$ úgy, hogy ha $|x_0| \leq \sigma(t_0)$, akkor egy alkalmas $T = T(\epsilon, t_0, x_0)$ számmal teljesül az $y|(t_0 + T; x_0, t_0)| < \delta(\epsilon)/2$ egyenlőtlenség. A megoldások a kezdeti értékektől folytonosan függnek, tehát a $\bar{B}_k(0, \sigma(t_0))$ gömb minden pontjának van olyan $\gamma(x_0) > 0$ sugarú gömbkörnyezete, amelynek tetszőleges x_* pontjából induló megoldásra teljesül az $|y(t_0 + T(\epsilon, t_0, x_0); x_*, t_0)| < \delta(\epsilon)$ egyenlőtlenség. A $\bar{B}_k\ (0, \sigma(t_0))$ kompakt halmazt pontjainak említett környezetei közül véges sok is lefedi, tehát létezik olyan $T_1 = T_1(\epsilon, t_0)$ szám, hogy bármely $x_0 \in \bar{B}_k(0, \sigma(t_0))$ pontra

$$\min\{|y(t; x_0, t_0)| : t \in [t_0, t_0 + T_1]\} < \delta(\epsilon).$$

Az erős y-stabilitás definíciója miatt ebből már következik az $|y(t; x_0,\ t_0)| < \epsilon$ egyenlőtlenség minden $t \geq t_0 + T_1(\epsilon, t_0)$ értékre, vagyis a 0-megoldás ekvi-aszimptotikusan y-stabilis. □

Tekintsük a (3.1) egyenlet egy $\phi(t) = (\psi(t), \chi(t))^T$ megoldását, amely értelmezve van egy $[t_0, \infty)$ intevallumon. Mivel csak a változók egy részére, az ellenőrzött koordinátákra vonatkozóan kívánunk aszimptotikus stabilitást biztosítani, be kell vezetnünk az ellenőrzött koordinátákra vonatkozó pozitív határhalmazt. Jelölje $\Omega_m = \Omega_m(\phi)$ azon $q \in R^m$ pontok halmazát, amelyekhez létezik egy $\{t_i\}_{i=1}^{\infty}$ sorozat úgy, hogy $t_i \to \infty$ és $\Psi(t_i) \to q$ $(i \to \infty)$. A 0-megoldás y-attraktivitásához nyilván azt kell bizonyítani, hogy $\Omega_m = \{0\}$. (Az Ω_m halmazzal az autonóm rendszereknél a következő fejezetben részletesen foglalkozunk.)

Most kimondjuk a tételt, és majd a feltételek birtokában, a bizonyítás elején vázoljuk a módszer alapvető ötleteit.

3.1.4. TÉTEL. *Tegyük fel, hogy a (3.1) egyenlethez léteznek olyan* $V, A \in Lip(\Gamma''_M; R)$, $W \in Lip(B_m(0, h'); R^j)$ *függvények, amelyek eleget tesznek az alábbi feltételeknek a* Γ''_m *halmazon:*

(1) létezik olyan $a, b \in \mathcal{K}$ *függvénypár, hogy*

$$a(|y|) \leq V(y, z, t) \leq b(|y|);$$

(2) $\dot{V}(y, z, t) \leq \alpha(t)U(y)$*, ahol az* $\alpha: R_+ \to R_+$, $U: B_m(0, h') \to R_-$ *függvények folytonosak, és* α *integrálisan pozitív,*

(3) ha F jelöli U zéróhelyeinek halmazát, akkor bármely $K \subset \bar{B}_m\ (0, h') \setminus F$ *zárt halmazhoz létezik olyan* $\rho > 0$ *szám, hogy* $\overline{B}^*_m(K, \rho) \cap F = \emptyset$;

(4) bármely $t_0 \in R_+$-ra és bármely $u : t \mapsto u(t) \in [B^*_m(K,\rho) \cap B_m(0,h')] \times M_z(t;t_0)$ lokálisan abszolút folytonos függvényre az $\int_{t_0}^{t} \dot{W}\ (s,u(s))ds$ egyenletesen folytonos a $[t_0,\infty)$ intevallumon;

(5) tetszőleges $t_0 \in R_+$ és $\eta > 0$ számhoz létezik olyan $\beta > 0$ szám és $\xi : R_+ \to R$ folytonos függvény ($\int_0^\infty \xi = \infty$), hogy minden $u : t \mapsto u(t) \in J \times M_z(t;t_0)$ lokálisan abszolút folytonos függvényre, ahol

$$J = \{y \in R^m : \eta \le |y| \le h', \quad y \in \bar{B}_m(F,\beta)\},$$

az $A(u(t),t)$ függvény felülről korlátos, és az $\dot{A}(u(t),t) \ge \xi(t)$ egyenlőtlenség teljesül a $[t_0,\infty)$ intervallumon.

Ekkor a (3.1) egyenlet 0-megoldása ekvi-aszimptotikusan y-stabilis.

Bizonyítás. Először vázoljuk a bizonyítás alapvető gondolatait.

Az első feltétel (a $\dot{V}(x,t) \le 0$ egyenlőtlenséggel együtt, ami (2)-ből következik) biztosítja a 0-megoldás erős y-stabilitását. Tehát már csak azt kell megmutatni, hogy tetszőleges megoldás ellenőrzött koordinátáiból álló vektor hossza tetszőlegesen kicsiny értékeket is felvesz. Ehhez előbb a (2)-(4) feltételeket használva megmutatjuk, hogy $\Omega_m \subset F$ (l. a 2.2.4. tétel feltételeit). Ebből következik, hogy az F halmaz tetszőleges környezete esetén a $H : t \mapsto y(t;x_0,t_0)$ görbe valahonnan kezdve ebben a környezetben helyezkedik el. Az (5) feltétel viszont azt mondja, hogy a $B_m(F,\beta)$ környezet J részhalmaza nem tartalmazhatja valamely pontjától kezdve az egész H görbét, mivel ekkor $\int_0^\infty \xi = \infty$ miatt $A(x(t),t)$ nem lenne felülről korlátos. Másrészt H a J halmazt az $|y| \ge \alpha_1$ feltételt megsértve hagyhatja el, amit éppen szeretnénk.

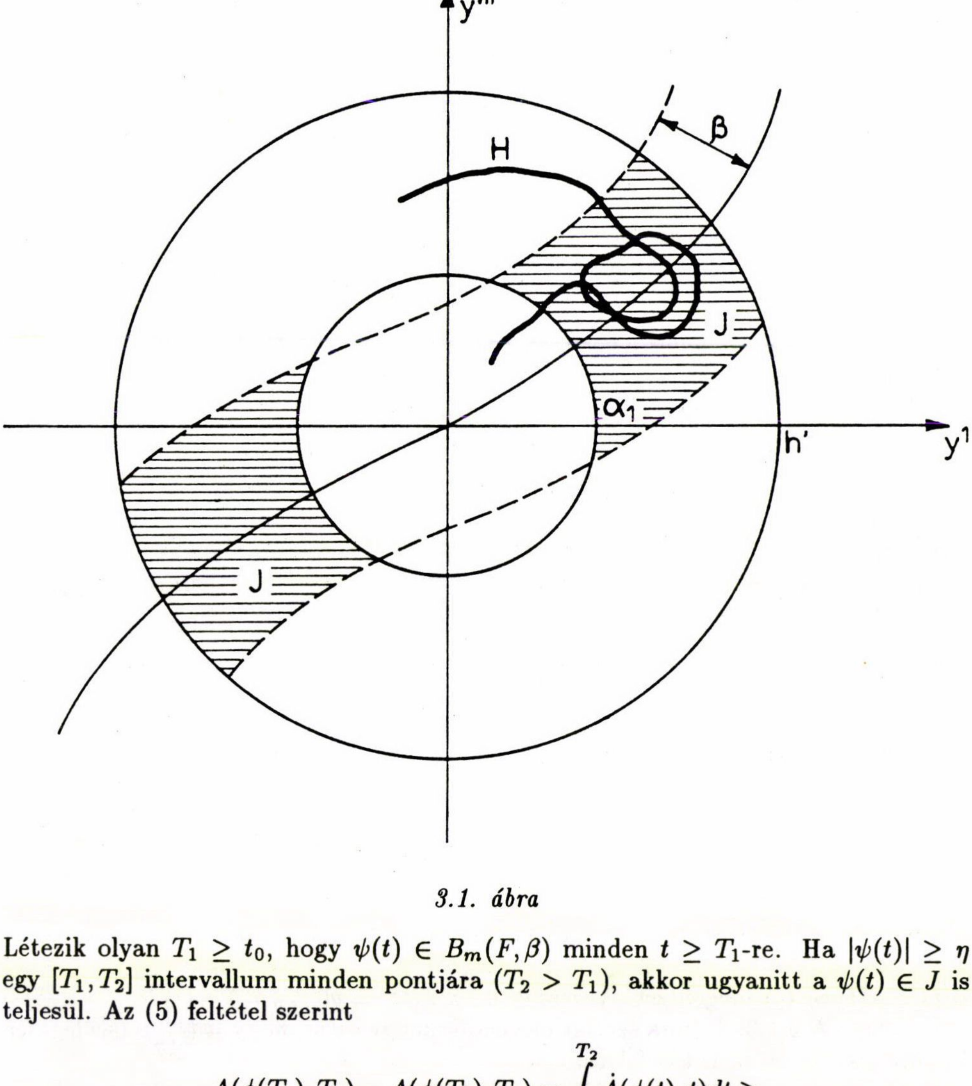
Nézzük mindezt részletesen. Az (1)-(2) feltételekből a 3.1.2. lemma szerint következik a 0-megoldás erős y-stabilitása, vagyis bármely $\epsilon > 0$ számhoz létezik olyan $\delta(\epsilon) > 0$, hogy ha $t_0 \in R_+$, $|y_0| < \delta(\epsilon)$, és $z_0 \in M_z(t_0)$, akkor $|y(t;x_0,t_0)| < \epsilon$ a $[t_0,\infty)$ intervallumon. Legyen $\sigma := \delta(h') > 0$ és tekintsünk egy tetszőleges $\phi(t) = x(t;x_0,t_0)$ megoldást. Vezessük be a $\psi(t) := y(t;x_0,t_0)$, $\chi(t) := z(t;x_0,t_0)$ jelöléseket. A 3.1.3. lemma szerint elegendő megmutatni, hogy $|\psi(t)|$ tetszőlegesen kicsiny pozitív értékeket is felvesz.

Az $y = \psi(t)$ függvény nyilván megoldása az

$$\dot{y} = Y(y,\chi(t),t) \tag{3.2}$$

differenciálegyenletnek, és ezen megoldás pozitív limeszhalmaza éppen $\Omega_m(\phi)$, azaz $\Omega(\psi) = \Omega_m(\phi)$. A $\bar{V}(y,t) := V(y,\chi(t),t)$ *Ljapunov-függvénye* a (3.2) egyenletnek. Könnyű látni, hogy a $\Lambda := B_m(0,h') \times R_+$ választással a 2.2.4. tétel minden feltétele teljesül a (3.2) rendszerre, tehát $\Omega_m(\phi) \subset F$, sőt $\psi(t) \to F$ $(t \to \infty)$.

Legyen $\eta > 0$ tetszőlegesen rögzített szám a $(0,h')$ intervallumon és tekintsük a (t_0,η) párhoz az (5) feltétel értelmében tartozó $\beta > 0$ számot és ξ függvényt.



3.1. ábra

Létezik olyan $T_1 \geq t_0$, hogy $\psi(t) \in B_m(F, \beta)$ minden $t \geq T_1$-re. Ha $|\psi(t)| \geq \eta$ egy $[T_1, T_2]$ intervallum minden pontjára ($T_2 > T_1$), akkor ugyanitt a $\psi(t) \in J$ is teljesül. Az (5) feltétel szerint

$$A(\phi(T_2), T_2) - A(\phi(T_1), T_1) = \int_{T_1}^{T_2} \dot{A}(\phi(t), t)dt \geq$$

$$\geq \int_{T_1}^{T_2} \xi(t)dt \to \infty \quad (T_2 \to \infty).$$

Másrészt, $A(\phi(t), t)$ felülről korlátos, tehát létezik olyan $T_2 \geq t_0$, hogy $|\psi(T_2)| < \eta$, amit bizonyítani akartunk. □

A módszer alkalmas instabilitást biztosító feltételek levezetésére is.

3.1.5. TÉTEL. *Tegyük fel, hogy a (3.1) egyenlethez léteznek olyan* $V, A \in Lip(\Gamma''_m; R)$, $W \in Lip(B_m(0,h'); R^j)$ *függvények, amelyek eleget tesznek a* Γ''_m *halmazon az alábbi feltételeknek:*

(1) bármely $t_0 \in R_+$*-ra és bármely* $\delta > 0$*-ra a* $V(\cdot, t_0) : B_k(0,\delta) \to R$ *függvény vesz fel negatív értékeket, továbbá létezik olyan* $b \in \mathcal{K}$, *hogy*

$$V(y,z,t) \geq -b(|y|);$$

(2)-(5) teljesülnek a 3.1.4. tétel (2)-(5) feltételei.

Ekkor (3.1) 0-megoldása y-instabilis, sőt bármely ϵ, δ, t_0 $(0 < \delta < \epsilon < h',\ t_0 \in R_+)$ *számhármashoz létezik olyan* x_0, T, *hogy* $|x_0| < \delta$, $T > t_0$ *és* $|y(T; x_0, t_0)| = \epsilon$.

Bizonyítás. Legyenek $t_0 \in R_+$, $\delta > 0$ rögzítve. Tekintsünk egy olyan $x_0 \in B_k(0,\delta)$ pontot, amelyre $V(x_0,t_0) < 0$, és tekintsük az egyenlet $x(t; x_0, t_0)$ megoldását, amely értelmezve van valamely $[t_0, \omega)$ intervallumon. Nyilván elegendő bebizonyítani, hogy ezen megoldás grafikonja elhagyja valamikor a Γ''_m halmazt, hiszen ez csak úgy lehetséges, hogy valamely $t_* \in (t_0, \omega)$ számra $|y(t_*; x_0, t_0)| > h'$.

Tegyük fel, hogy ez az állítás nem igaz. Akkor $\omega = \infty$, és a megoldás grafikonja Γ''_m-ben halad, amiből az (1) feltétel szerint az is következik, hogy a $V(x(t;x_0,t_0),t)$ függvény alulról korlátos. Ezen körülmények között a (2)–(4) feltételekből — ugyanúgy, mint a 3.1.4. tételben — következik az $\Omega_m \subset F$ tartalmazás. Másrészt, a $V(x,t)$ függvény a megoldás mentén csökkenő, ezért az (1) feltétel következtében

$$h' \geq |y(t;x_0,t_0)| \geq b^{-1}(-V(x(t;x_0,t_0),t) \geq$$
$$\geq b^{-1}(-V(x_0,t_0)) =: \eta > 0; \quad (t \geq t_0).$$

Ebből viszont az (5) feltétel szerint egyrészt $A(x(t),t) \to \infty$ $(t \to \infty)$, másrészt $A(x(t),t)$ korlátos a $[t_0, \infty)$ intervallumon, ami lehetetlen. □

3.2. Finomítások, általánosítások

1° Ha a 3.1.4. és 3.1.5. tétel (2) feltételében az α függvény nem feltétlenül integrálisan pozitív, csak $\int_0^\infty \alpha = \infty$ teljesül, viszont, egyidejűleg, a (4) feltételben az $\int_{t_0}^\infty |\dot W(u,(t),t)|dt < \infty$ egyenlőtlenséget követeljük meg, akkor a tétel állítása érvényben marad.

Valóban, a bizonyításban csupán annyi a különbség, hogy az Ω_m halmaz lokalizálására a 2.2.4. tétel helyett a 2.2.5. tételt kell használni.

2° Ha a 3.1.4. és 3.1.5. tétel (4) feltételében szereplő W függvény skaláris értékű $(j = 1)$, akkor a (4) feltételt $\dot W$ helyett elegendő a $[\dot W]_+$ függvénnyel megkövetelni (l. a 2.3.1. megjegyzést).

3° A tételek (2) feltételében szereplő α függvény integrális pozitivitása elég erős kikötés. Igaz ugyan, hogy $\alpha(t)$ még a 0 értéket is felveheti bármilyen nagy t-re, de az integrális pozitivitás (2.4) szükséges és elegendő feltétele mutatja, hogy átlagban már nem lehet akármilyen kicsiny. Ez a feltétel gyengíthető, ha egyidejűleg egy másik feltételt erősebbel pótolunk.

3.2.1. *Definíció.* A mérhető $\alpha : R_+ \to R_+$ függvényt ***gyengén integrálisan pozitívnak*** nevezzük, ha $\int_S \alpha = \infty$ teljesül minden olyan $S = \bigcup_{i=1}^{\infty}[a_i, b_i]$ halmazra, amelyhez van olyan $\delta > 0$ és $\gamma > 0$, hogy S komponensei eleget tesznek az

$$a_i < b_i < a_{i+1}, \quad b_i - a_i \geq \delta > 0, \quad a_{i+1} - b_i \leq \gamma$$

feltételeknek ($i = 1, 2, \dots$).

A továbbiakban az integrális pozitivitást IP-vel, a gyenge integrális pozitivitást GYIP-vel rövidítjük. A két fogalom definícióját összehasonlítva (2.2.1. definíció) látszik, hogy IP $\Rightarrow$ GYIP; nevezetesen, a GYIP-nál α integráljának csak olyan intervallumrendszer egyesítésén kell divergálnia, amelyben a szomszédos intervallumok távolsága korlát alatt marad. Másrészt viszont GYIP $\not\Rightarrow$ IP; például, az $1/(t+1)$, $(1/(t+1))\sin^2 t$ függvények GYIP-ok, de nem IP-ok.

Ha egy α függvény GYIP, akkor nyilván $\int_0^{\infty} \alpha = \infty$. Másrészt, könnyű belátni, hogy ha $\alpha : R_+ \to R_+$ monoton és $\int_0^{\infty} \alpha = \infty$, akkor α GYIP.

Ha egy α függvény IP, akkor $\int_T^{\infty} \alpha = \infty$ T-ben egyenletesen (amin azt értjük, hogy bármely M számhoz létezik olyan $S_0(M)$, hogy ha $S > S_0(M)$, akkor $\int_T^{T+S} \alpha > M$ minden $T \in R_+$-ra). Ennek bizonyításához szintén a (2.14) feltételhez kell fordulni. Az előző példákból látszik, hogy a GYIP-ból ez az egyenletes divergencia nem következik.

Néhány későbbi tétel feltételeinek összehasonlításánál hasznos az az észrevétel, hogy a GYIP és a T-ben egyenletes $\int_T^{\infty} \alpha = \infty$ együtt sem vonják maguk után, általában, az IP-t. Például, az

$$\alpha(t) := \begin{cases} \frac{1}{1+t}, & n \leq t \leq n + \frac{1}{2} \\ 1, & n + \frac{1}{2} < t < n + 1 \qquad (n = 1, 2, \dots) \end{cases}$$

függvény GYIP, $\int_T^{\infty} \alpha = \infty$ T-ben egyenletesen, de nem IP.

ÁLLÍTÁS. *Ha a (2) feltételben α csak GYIP, de az (5) feltételben szereplő A függvényre*

$$|A(x,t)| \leq c, \quad \dot{A}(x,t) \geq \xi(t) \quad (t \in R_+,\ y \in J,\ z \in \overline{M_z(t)})$$

teljesül valamely c konstanssal és olyan ξ függvénnyel, amelyre $\int_T^\infty \xi = \infty$ T-ben egyenletesen, akkor a 3.1.4. és 3.1.5. tétel állításai érvényben maradnak.

Bizonyítás. Mint a 3.1.4. tétel bizonyításában, most is azt kell kimutatni, hogy a $\psi(t) = |y(t; x_0, t_0)|$ függvény tetszőlegesen kicsiny értékeket is felvesz.

Legyen $\eta > 0$ tetszőleges, h'-nél kisebb szám, és tegyük fel, hogy $|\psi(t)| > \eta$ a $[t_0, \omega)$ intervallumon. Ekkor ezen az intevallumon

$$\psi(t) \in H(\eta, h') := \{y \in R^m : \eta \leq |y| \leq h'\}.$$

Tekintsük az η-hoz az (5) feltétel szerint tartozó $\beta > 0$ számot és ξ függvényt, és vezessük be a $K := H(\eta, h') \setminus \bar{B}_m(F, \beta)$ jelölést. A (3) feltétel szerint a $\overline{B^*_m(K,\rho)}$ halmaznak F-fel nincs közös pontja, és α GYIP ezért a $H : t \mapsto \psi(t)$ görbének el kell hagynia ezt a halmazt. Másrészt, az (5) feltétel miatt a H görbe a J halmazban sem tartózkodhat örökre, így létezik időpillanatoknak egy $t_0 < \ldots < a_i < s_i < t_i < b_i < a_{i+1} < \ldots < \omega$ sorozata úgy, hogy

$$|W(\psi(s_i)) - W(\psi(a_i))| \geq \rho, \quad |W(\psi(b_i)) - W(\psi(t_i))| \geq \rho$$
$$[t_i \leq t \leq s_{i+1}] \Longrightarrow \psi(t) \in J$$
$$[a_i < t < b_i] \Longrightarrow \psi(t) \in B^*_m(K, \rho), \qquad (i = 1, 2, \ldots).$$

A (4) feltétel miatt $b_i - a_i \geq \delta > 0$ $(i = 1, 2, \ldots)$ valamely alkalmas δ számmal. Másrészt,

$$2c \geq A(\psi(s_{i+1}), s_{i+1}) - A(\psi(t_i), t_i) \geq \int_{t_i}^{s_{i+1}} \xi(t)dt,$$

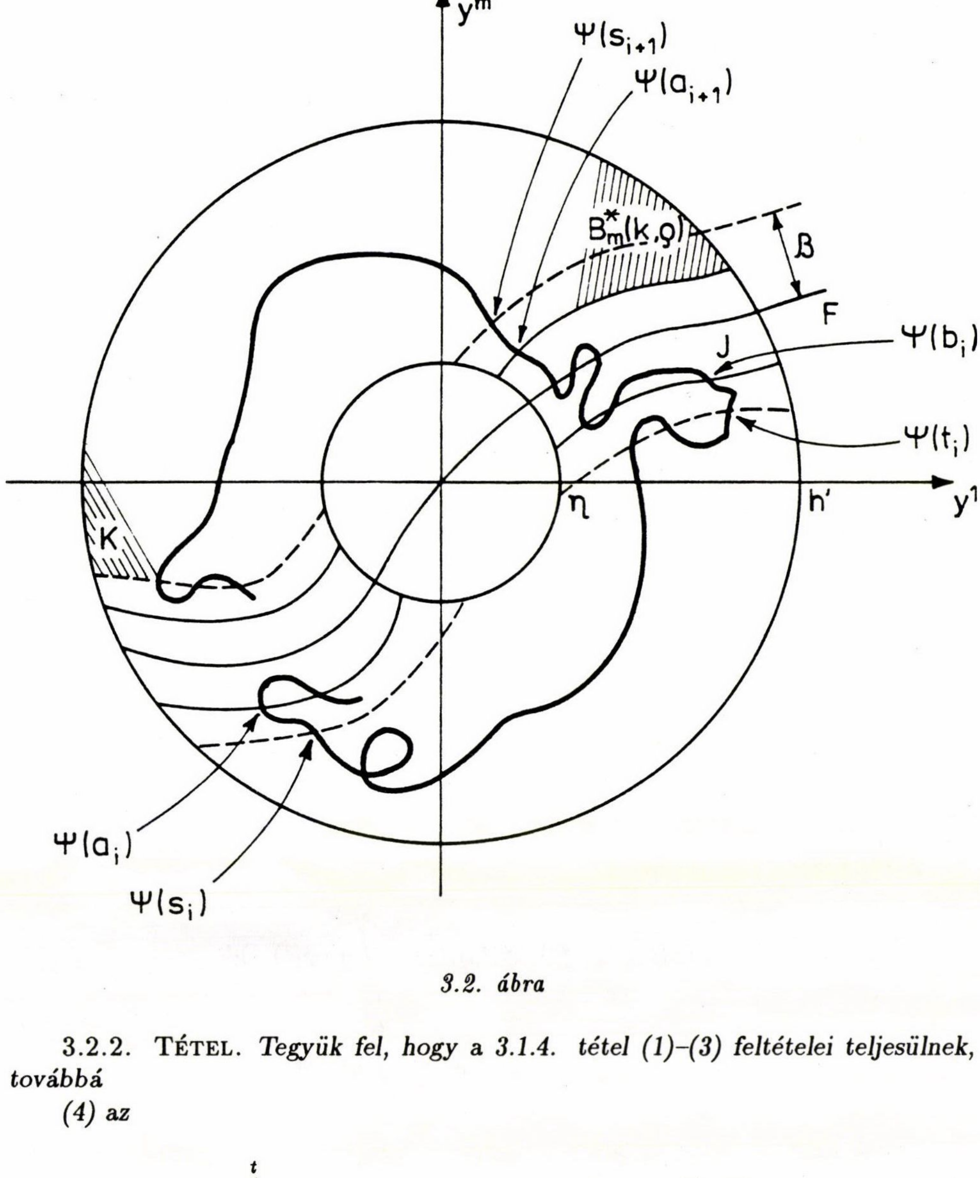
továbbá $\int_T^\infty \xi = \infty$ T-ben egyenletesen, ezért létezik olyan $\gamma = \gamma(\eta)$, hogy $a_{i+1} - b_i \leq s_{i+1} - t_i \leq \gamma$ $(i = 1, 2, \ldots)$.

A (2) feltétel következtében

$$0 \leq V(\phi(b_i), b_i) \leq V(\phi(t_0), t_0) + \int_{t_0}^{b_i} \dot{V}(\phi(t), t)dt$$
$$\leq b(\sigma) - c_1 \sum_i \int_{a_i}^{b_i} \alpha(t)dt.$$

Mivel α GYIP, és $b_i - a_i \geq \delta > 0$, $a_{i+1} - b_i \leq \gamma$ $(i = 1, 2, \ldots)$, az utolsó becslés az $\omega = \infty$ esetben ellentmondáshoz vezet. □

4° Mint ismeretes, az egyenletes aszimptotikus stabilitásnak a stabilitáselméletben a totális stabilitással való kapcsolata miatt különös jelentősége van. Feltételeink kis módosításával ezt a tulajdonságot is biztosítani tudjuk.



3.2. ábra

3.2.2. TÉTEL. *Tegyük fel, hogy a 3.1.4. tétel (1)–(3) feltételei teljesülnek, továbbá*

(4) az

$$\int_0^t \sup\{|\dot{W}(x,s)| : x \in B^*_m(K,\rho) \times M_z(s)\} ds$$

függvény egyenletesen folytonos R_+-on;

(5) tetszőleges $t_0 \in R_+$, $\eta > 0$ számokhoz létezik olyan c, $\beta > 0$ szám és $\xi : R_+ \to R$ folytonos függvény hogy

$$|A(x,t)| \le c, \quad \dot{A}(x,t) \ge \xi(t) \quad \left(t \in R_+,\ y \in J,\ z \in \overline{M_z(t)}\right),$$

és $\int_T^\infty \xi = \infty$ T-ben egyenletesen.

Ekkor a (3.1) egyenlet 0-megoldása egyenletesen aszimptotikusan y-stabilis.

Bizonyítás. A 3.1.2. lemma szerint a 0-megoldás erősen y-stabilis, vagyis bármely $\epsilon > 0$-hoz létezik olyan $\nu(\epsilon) > 0$, hogy ha $t_0 \in R_+$, $|y_0| < \nu(\epsilon)$, és $z_0 \in M_z(t_0)$, akkor $|y(t;x_0,t_0)| < \epsilon$ a $[t_0,\infty)$ intervallumon. Legyen $\sigma := \nu(h')$. Azt kell bebizonyítanunk, hogy tetszőleges $\zeta > 0$ számhoz létezik olyan $T(\zeta)$, hogy ha $|x_0| < \sigma$, $t_0 \in R_+$, $t \ge t_0 + T(\zeta)$ akkor $|y(t;x_0,t_0)| < \zeta$. Ehhez elegendő belátni, hogy létezik olyan $T(\zeta)$, hogy az origó σ sugarú környezetéből tetszőleges időpillanatban induló megoldásra $|y(t_*;x_0,t_0)| < \nu(\zeta)$ teljesül a $[t_0,\ t_0+T(\zeta)]$ intervallum valamely pontjában.

A 3° pontban adott bizonyítás jelöléseit használva, azt kell belátni, hogy az $\eta = \nu(\zeta)$-hoz létezik olyan $T(\zeta)$, hogy bármely $t_0 \in R_+$-ra, $x_0 \in B_k(0,\sigma)$-ra az $x(t;x_0,t)$ megoldáshoz tartozó ω kisebb, mint $t_0 + T(\zeta)$.

Tekintsük most is az $\{a_i, b_i\}$ sorozatot. A (4) feltétel szerint létezik olyan $\delta = \delta(\zeta)$, hogy $b_i - a_i \ge \delta(\zeta)$. Másrészt, α IP, és ennek a tulajdonságnak (2.4) feltétele szerint létezik olyan $T_1(\zeta)$ és $m(\zeta) > 0$, hogy ha $a_i > T_1(\zeta)$, akkor $\int_{a_i}^{b_i} \alpha(t)dt \ge m(\zeta)$. Ez azt jelenti, hogy a $V(\phi(t),t)$ függvény az $[a_i,b_i]$ intervallumon legalább $c_1 m(\zeta)$-val csökken. Mivel $V(\phi(t_0),t_0) \le b(\sigma)$, ez maga után vonja, hogy i nem vehet fel a $\left[\dfrac{b(\sigma)}{c_1 m(\zeta)}\right] + 1$ számnál nagyobb értéket.

Tudjuk azt is, hogy $\int_P^\infty \alpha = \infty$ P-ben egyenletesen, hiszen α IP. Tehát van olyan T_2, hogy $\int_P^{P+T_2} \alpha > b(\sigma)$ minden $P \in R_+$-ra. Ebből látszik, hogy az $\{a_i, b_i\}$ sorozatról feltehetjük a $b_i - a_i \le T_2$ tulajdonságot is (a_i-hez a lehető legközelebb kell választani b_i-t).

A fenti tényeket egymáshoz kapcsolva belátható, hogy

$$\omega \le T(\zeta) := T_1 + \left(2 + \frac{b(\sigma)}{c_1 m(\zeta)}\right)(T_2 + \gamma(\nu(\zeta))). \qquad \square$$

5° A 3.1.5. tétel bizonyításából kitűnik, hogy ha a tételben az (5) feltétel $\eta = 0$-val is teljesül, akkor az (1) feltételben a $b \in \mathcal{K}$ függvény létezése helyett elegendő feltenni a következőt: V alulról korlátos a Γ''_m halmazon.

3.3. Másodrendű nem-lineáris differenciálegyenletek megoldásainak és disszipatív mechanikai rendszerek egyensúlyi helyzetének aszimptotikus stabilitása

1° Első alkalmazásként ismét egy nem-lineáris másodrendű differenciálegyenletet vizsgálunk, de most olyan feltételeket keresünk, amelyek mind a kitérésre, mind a sebességre vonatkozóan biztosítják az egyensúlyi helyzet aszimptotikus stabilitását, illetve instabilitását.

Tekintsük a

$$(p(t)\dot{x})^{\cdot} + a(t)\dot{x} + q(t)f(x) = 0 \tag{3.3}$$

egyenletet, ahol $p : R_+ \to (0,\infty)$, $q : R_+ \to R \setminus \{0\}$ folytonosan differenciálható, $a : R_+ \to R$, $f : R \to R$ folytonos függvények, $f(0) = 0$. Vezessük be az $F(x) := \int_0^x f(r)dr$ jelölést.

A klasszikus mechanikából [7,12,70] ismeretes, hogy (3.3) egy olyan egy szabadsági fokú holonóm reonóm mechanikai rendszer mozgásegyenlete, amelynek kinetikai energiája $(1/2)\,p(t)(\dot{x})^2$, potenciális energiája $q(t)\;F(x)$, és amelyre még $a - a(t)\dot{x}$ nem-konzervatív erő is hat (ha $a(t) \geq 0$, akkor ezt surlódási erőnek hívjuk, ha $a(t) \leq 0$, akkor „gyorsító" nem-konzervatív erőnek (ez az elnevezés nem túl szerencsés, hiszen minden erő gyorsító, ennek ellenére használjuk, mivel elfogadott az irodalomban [70])).

A mechanikai rendszerek stabilitáselméletéből [46,81] ismeretes, hogy az $x = \dot{x} = 0$ egyensúlyi állapot stabilitása akkor várható, ha a potenciális energiának az $x = 0$-ban szigorú minimuma van (l. a *Lagrange-Dirichlet-tételt* a konzervatív rendszerekre [46]). Az is ismert tapasztalat, hogy a surlódási erőnek stabilizáló hatása van, és hogy aszimptotikus stabilitás nem érhető el megfelelő nagyságú surlódási erő nélkül [29]. A surlódási erő adagolásával az aszimptotikus stabilitás biztosításánál azért óvatosnak kell lennünk, mivel ha túl gyorsan fékezzük a rendszert, akkor az aszimptotikusan megállhat az egyensúlyi helyzettől távol. Ezt mutatja J. HALE [15] következő egyszerű példája: az $\ddot{x} + (2 + e^t)\dot{x} + x = 0$ egyenletnek az $x(t) = c[1 + e^{-t}]$ függvények megoldásai. Számos dolgozat foglalkozik azzal a problémával, hogy milyen surlódási erő mellett van aszimptotikus stabilitás [1,5,8,11,18,19,54,72], de általában vagy azt tételezik fel, hogy $a(t)$ korlátos, vagy hogy $a(t)$ pozitív konstans felett marad.

Ismeretes, hogy az egyensúlyi állapot instabilitása akkor várható, ha a potenciális erő instabilis típusú (a potenciális energiának nincs szigorú minimuma), illetve ha „gyorsító" nem-kozervatív erő hat $a(t) < 0$. Mindkét hatás kiolvasható lesz az instabilitást biztosító feltételeinkből.

Az $x_1 = x$, $x_2 = \dot{x}$ változók bevezetésével a (3.3) egyenlet ekvivalens az

$$\dot{x}_1 = x_2, \quad \dot{x}_2 = -\frac{q(t)}{p(t)}f(x_1) - \frac{\dot{p}(t) + a(t)}{p(t)}x_2 \tag{3.4}$$

rendszerrel. Válasszuk az alábbi segédfüggvényeket:

$$V(x_1,x_2,t)=\frac{p}{q}x_2^2+2F(x_1),\quad W(x_2)=x_2^2,$$
$$A_1(x_1,x_2,t)=-p(t)x_2f(x_1),\quad A_2(x_1,x_2)=-x_2f(x_1).$$

Ezeknek a (3.4) rendszerre vonatkozó deriváltjai:

$$\begin{aligned}
\dot V &= -\alpha(t)x_2^2,\quad \alpha(t):=\frac{(pq)^{\cdot}+2aq}{q^2},\quad (U(x_1,x_2)=-x_2^2);\\
\dot W &= -2x_2^2\left(\frac{\dot p}{p}+\frac{a}{p}\right)-2\frac{q}{p}x_2f(x_1)\\
\dot A_1 &= q[f^2(x_1)-\frac{p}{q}x_2^2f'(x_1)]+ax_2f(x_1)\\
\dot A_2 &= \frac{q}{p}\left[f^2(x_1)-\frac{p}{q}x^2f'(x_1)\right]+\frac{a+\dot p}{p}x_2f(x_1).
\end{aligned}$$

A 3.1.4. és 3.1.5. tétel következményeként adódik a

3.3.1. TÉTEL. *Tegyük fel, hogy a (3.3) egyenletben szereplő függvények eleget tesznek a következő feltételeknek:*

(i) $0<c_1\le\left|\dfrac{p(t)}{q(t)}\right|\le C_1,\quad (t\in R_+);$

(ii) $\displaystyle\int_0^t\frac{[\dot p(s)+a(s)]_{+(-)}}{p(s)}ds$ *egyenletesen folytonos* R_+*-on;*

(iii) $|\alpha(t)|$ *integrálisan pozitív;*

(iv) *vagy*

a) $\displaystyle\int_0^\infty|q|=\infty,\quad p(t)\le C_2,\quad \int_0^t|a|=O(\int_0^t|q|)\qquad(t\to\infty),$

vagy

b) $\displaystyle\limsup_{T\to\infty}\frac{1}{T}\int_0^T\left|\frac{a+\dot p}{p}\right|<\infty.$

Ekkor:

1) *ha* $xf(x)>0$ $(x\neq 0)$, $q(t)>0$ *és* $\alpha(t)\ge 0$ $(t\in R_+)$, *akkor az* $x=\dot x=0$ *egyensúlyi állapot ekvi–aszimptotikusan stabilis.*

2) *Tegyük fel, hogy bármely* $\sigma>0$ *számra* $F(x)\not\equiv 0$ *az* $|x|<\sigma$ *halmazon.*

   α) *Ha* $F(x)$ *vesz fel pozitív és negatív értéket is bármely* $|x|<\sigma$ *halmazon, vagy* $q(t)F(x)\le 0$ *valamely* $R_+\times B_1(0,\sigma_0)$ $(\sigma_0>0)$ *halmazon, továbbá* $\alpha(t)$ *jeltartó, akkor az* $x=\dot x=0$ *egyensúlyi állapot instabilis.*

   β) *Ha* $q(T)F(x)\ge 0$ *az* $R_+\times B_1(0,\sigma_0)$ *halmazon, és* $\alpha(t)\le 0$ $(t\in R_+)$, *akkor az* $x=\dot x=0$ *egyensúlyi állapot instabilis.*

3.3.2. *Megjegyzés.* A 3.3.1. tételben a (iii)–(iv) feltétel pótolható a következővel:

(iii') $|\alpha(t)|$ gyengén integrálisan pozitív;

(iv') vagy

a) $\int_{t_*}^{\infty} |q| = \infty$ t_*-ban egyenletesen, és létezik olyan K és T_0, hogy ha $t > T_0$, $T > t_0$, akkor

$$\int_t^{t+T} |a| \Big/ \int_t^{t+T} |q| \le K,$$

vagy

b) $\limsup\limits_{\substack{t\to\infty \\ T\to\infty}} \frac{1}{T} \int_t^{t+T} \left| \frac{a+\dot p}{p} \right| < \infty.$

Továbbá, ha $\alpha(t)$ integrálisan pozitív, és (iv') teljesül, akkor 1)-ben egyenletesen aszimptotikus stabilitást állíthatunk.

R. J. Ballieu és K. Peiffer [5] bebizonyították, hogy az $(x, \dot x) = (0,0)$ pont globális attraktora az

$$\ddot x + a(t)\dot x + f(x) = 0$$

egyenletnek, (azaz bármely megoldásra $x(t) \to 0$, $\dot x(t) \to 0$ $(t \to \infty)$), ha $a : R_+ \to R_+$ nem–növekvő, $\int^{\infty} a = \infty$, $xf(x) > 0$ $(x \ne 0)$, és $F(x) \to \infty$ $(|x| \to \infty)$.

Ebben a speciális esetben $p(t) \equiv q(t) \equiv 1$, és az a függvényre tett feltételek teljesülése esetén $\alpha(t) = 2a(t)$ GYIP, vagyis a 3.3.2. megjegyzés (iii'), (iv')b) feltételei teljesülnek. Tehát a 3.3.1. tétel – 3.3.2. megjegyzés általánosítása, de egyben élesítése is Ballieu és Peiffer eredményének, mutatva, hogy a 3.1.4. tétel újdonsága nemcsak abban áll, hogy parciális stabilitási tulajdonságra vonatkozik, hanem abban is, hogy alkalmas a *klasszikus Ljapunov-féle* (az összes változóra vonatkozó) *stabilitási tulajdonságok* megállapítására is.

2° Disszipatív giroszkópikus rendszer. Most kerül sor annak a mechanikai rendszernek a bevezetésére, amelynek vizsgálata a jelen dolgozat legtöbb eredményét inspirálta.

Tekintsünk egy r szabadsági fokú holonóm, szkleronóm mechanikai rendszert, amelyre potenciális, disszipatív és giroszkópikus erők hatnak. Egy ilyen rendszer mozgását leíró *Lagrange-féle másodfajú mozgásegyenlet* a következő alakú ([46], II. függelék):

$$\frac{d}{dt}\frac{\partial T}{\partial \dot q} - \frac{\partial T}{\partial q} = -\frac{\partial \pi}{\partial q} - B\dot q + G\dot q, \tag{3.5}$$

ahol q, $\dot q \in R^r$ az általánosított koordinátákból, illetve általánosított sebességekből álló vektor; $\pi : q \mapsto \pi(q) \in R$ a potenciális energia ($\pi(0) = 0$), amely kétszer folytonosan differenciálható; $T = T(q,\dot q) = (1/2)\dot q^T A(q)\dot q$ a kinetikai energia, amelyben az $A : q \mapsto A(q) \in R^{r\times r}$ mátrixfüggvény szimmetrikus, minden q-ra pozitív definit [46], és kétszer folytonosan differenciálható; a $B : (q,t) \mapsto B(q,t) \in R^{r\times r}$,

$G : (q,t) \mapsto G(q,t) \in R^{r \times r}$ mátrixfüggvények folytonosak, B szimmetrikus és pozitív szemidefinit (a surlódási együtthatók mátrixa), G ferdén szimmetrikus (a giroszkópikus együtthatók mátrixa). Szerencsére már magyar nyelven is rendelkezésre áll N. ROUCHE, P. HABETS és M. LALOY igen értékes [46] könyve, amelyben az olvasó nagyon sok érdekes konkrét példát találhat a (3.5) rendszerre a legkülönfélébb alkalmazási területekről (mechanika, repüléselmélet, elektromosságtan, kémia, közgazdaságtan, stb.).

A surlódási erő okozta fékezés mérésére a $\dot{q}^T B(q,t)\dot{q}$ *Rayleigh-féle függvényt* szokás használni [46,12] (l. később a (3.7) formulát). Tegyük fel, hogy létezik olyan folytonos $\beta : R_+ \to R_+$ függvény, hogy

$$\dot{q}^T B(q,t)\dot{q} \geq \beta(t)|\dot{q}|^2 \quad (q \in R^r,\ t \in R_+).$$

3.3.3. *Definíció.* Azt mondjuk, hogy a disszipáció

a) *teljes*, ha $\beta(t) = \beta_0 =$ konstans > 0;
b) *integrálisan teljes*, ha β integrálisan pozitív;
c) *gyengén integrálisan teljes*, ha β gyengén integrálisan pozitív.

Tegyük fel, hogy $\dfrac{\partial \pi}{\partial q}(0) = 0$, vagyis $q = 0$ egyensúlyi helyzet.

A (3.5) rendszer egyensúlyi állapotának stabilitási tulajdonságaira vonatkozó vizsgálatok hosszú múltra tekintenek vissza. *Lagrange tétele* a $B(t,q) \equiv G(t,q) \equiv 0$ konzervatív esetre vonatkozik, és azt mondja, hogy ha konzervatív mechanikai rendszerben a π potenciális energiának a $q = 0$ helyen szigorú minimuma van, akkor a $q = \dot{q} = 0$ egyensúlyi állapot stabilis. A tételt LAGRANGE csak speciális esetekben tudta bizonyítani, az általános esetre DIRICHLET adott egy bizonyítást, amivel a stabilitási vizsgálatok ma is használatos leghatékonyabb módszerének, a direkt módszernek vetette meg az alapjait. A. M. LJAPUNOV-nak, a direkt módszer feltalálójának ugyanis „csupán" az a zseniális felismerés az érdeme, hogy DIRICHLET módszere általános dinamikus rendszerekre is átvihető, ha tudunk hozzájuk valamilyen „energiaszerű" segédfüggvényt találni. A *Lagrange-Dirichlet tétel* még ma is a kutatások középpontjában áll: a hosszú erőfeszítések ellenére még ma is nyitott kérdés, hogy igaz-e a tétel megfordítása [46].

Az energiaintegrál létezése (az energia megmaradásának tétele) miatt konzervatív esetben az egyensúlyi állapot aszimptotikus stabilitása nem is várható. Viszont általános tapasztalat, hogy ha teljes disszipációjú surlódási erő hat, akkor az egyensúlyi állapot aszimptotikusan stabilissá válik. Érdekes, hogy ezt a tényt elméletileg nagyon sokáig csak speciális esetekre [53] (pl. lineáris rendszerekre) sikerült bebizonyítani. Ennek az az oka, hogy a lineáris közelítés alapján való vizsgálatok módszere itt általában nem működik, mivel a lineáris közelítések a kritikus esetekhez tartoznak [63]. A probléma megoldásához a stacionárius esetben a *Barbasin–Kraszovszkij-tételek* alkalmazása vezetett el; így született L. SALVADORI [47] híres tétele:

3.3.4* TÉTEL. *Tegyük fel, hogy*

(1) a π potenciális energiának a $q = 0$ helyen minimuma van;

(2) a $q = 0$ egyensúlyi helyzet izolált;

(3) a disszipatív és giroszkópikus erő nem függ az időtől, és a disszipáció teljes.

Ekkor a $q = \dot{q} = 0$ egyensúlyi állapot aszimptotikusan stabilis.

Ha (1) helyett azt tesszük fel, hogy

(1') a π potenciális energiának a $q = 0$-ban nincs minimuma, akkor a $q = \dot{q} = 0$ instabilis.

Nyilván fontos hasonló feltételek kidolgozása az instacionárius esetre is, amikor a surlódási és giroszkópikus erők az időtől explicite is függnek (pl. szakaszos fékezés, járművek, gépek megállítása, stb.). Az első ilyen eredmény V. M. MATROSZOV [69] nevéhez fűződik, és speciális (bizonyos homogenitási tulajdonságokkal rendelkező) potenciális energia esetére ad feltételt. Ehhez a vizsgálathoz többen csatlakoztak [45–48, 60], és ma is intenzív kutatások folynak ebben az irányban [39,52, 77–79]. Itt most mi olyan feltételeket adunk meg, amelyek — a korábbi eredményektől eltérően — nem követelik meg a disszipáció teljes voltát, és a *Salvadori-tételt* még a stacionárius esetre szorítkozva is általánosítják, amennyiben nem követelik meg az egyensúlyi helyzettől, hogy az izolált legyen. Ez az utolsó engedmény azzal jár, hogy csak parciális aszimptotikus stabilitásra számíthatunk. Eközben azonban egy új probléma merül fel: a nem-ellenőrzött koordináták a mozgások során nem-korlátosak is lehetnek. Ezt az esetet a vizsgálatokból sokszor kizárják (pl. [44,72]). Az alábbi tételek az általános esetre vonatkoznak.

Vezessük be a

$$\gamma_L(t) := \sup\{|G(q,t) - B(q,t)| : q \in L \subset R^r\}$$

jelölést.

3.3.5. TÉTEL. *Legyen $q = (\tilde{q}, \hat{q})^T$ az általánosított koordináták vektorának egy partíciója ($\tilde{q} \in R^s$, $\hat{q} \in R^{r-s}$, $0 \le s \le r$), és tegyük fel, hogy valamely $h' > 0$-ra az alábbi feltételek teljesülnek a $K \times R^{r-s}$ halmazon ($K = B_s(0, h')$):*

(1) létezik olyan $a_1, b_1 \in \mathcal{K}$, hogy

$$a_1(|\tilde{q}|) \le \pi(\tilde{q}, \hat{q}) \le b_1(|\tilde{q}|);$$

(2) minden α_1, α_2 ($0 < \alpha_1 < \alpha_2 < h'$) számpárhoz létezik olyan $\eta > 0$, hogy $|\operatorname{grad} \pi(\tilde{q}, \hat{q})| \ge \eta \quad (\alpha_1 \le |\tilde{q}| \le \alpha_2)$;

(3) léteznek olyan $\lambda, \Lambda > 0$ konstansok, amelyekkel $\lambda |p|^2 \le p^T A^{-1}(q)\, p \le \Lambda |p|^2 \quad (p \in R^r)$;

(4) az $A^{-1}(q)$ elsőrendű parciális deriváltjai és a $\pi(q)$ első és másodrendű parciális deriváltjai korlátosak;

(5) a disszipáció gyengén integrálisan teljes;

(6) az $\int_0^t \gamma_{K \times R^{r-s}}(\tau) d\tau$ függvény egyenletesen folytonos R_+-on.

Ezen feltételek mellett a $q = \dot{q} = 0$ egyensúlyi állapot ekviaszimptotikusan $(\tilde{q}, \dot{q})^T$-stabilis. Ha a disszipáció még integrálisan is teljes, akkor az aszimptotikus stabilitás egyenletes.

3.3.6. TÉTEL. *Tegyük fel, hogy a potenciális energiának nincs helyi minimuma a $q = 0$ pontban, és létezik egy $b_1 \in \mathcal{K}$ függvény, amellyel a $-\pi(\tilde{q}, \hat{q}) \leq b_1(|\tilde{q}|)$ egyenlőtlenség teljesül valamely $B_s(0, h') \times R^{r-s}$ halmazon $(h' > 0)$. Tegyük fel, továbbá, hogy teljesülnek a 3.12. tétel (2)–(6) feltételei.*

Akkor a $q = \dot{q} = 0$ egyensúlyi állapot $(\tilde{q}, \dot{q})^T$-instabilis.

A 3.3.5. és 3.3.6. tétel bizonyítása. A q, $p := A(q)\dot{q}$ *Hamilton-változók* bevezetésével a (3.5) rendszer a

$$\dot{q} = \frac{\partial H}{\partial p}, \quad \dot{p} = -\frac{\partial H}{\partial q} + (G - B)\frac{\partial H}{\partial p} \tag{3.6}$$

kanonikus alakba írható, ahol $H = H(q, p) = T + \pi$ a teljes mechanikai energia. A (3) feltétel következtében a (3.6) rendszer $q = p = 0$ megoldásának stabilitási tulajdonságai megegyeznek a (3.5) rendszer $q = \dot{q} = 0$ állapotának stabilitási tulajdonságaival.

Tekintsük a (3.6) rendszerhez a

$$V(q, p) = H(q, p), \quad W(\tilde{q}, p) = p, \quad A_1(q, p) = -p^T \operatorname{grad} \pi(q)$$

segédfüggvényeket, amelyeknek (3.6) szerinti deriváltjait kiszámítva kapjuk, hogy ha $K_1 \subset R^s$ egy tesz̋oleges kompakt halmaz és $h' > 0$ rögzített, akkor $\tilde{q} \in K_1$, $|p| \leq h'$ esetén

$$\begin{aligned} \dot{H} &= -(A^{-1}p)^T B(A^{-1}p) \leq -\beta(t)|A^{-1}p|^2; \\ |\dot{W}| &= \left| -\frac{\partial H}{\partial q} + (G - B)A^{-1}p \right| \leq c_1 + c_2 \gamma_{K_1 \times R^{r-s}}(t); \\ \dot{A}_1 &= \left(\frac{\partial T}{\partial q} + \operatorname{grad} \pi - (G - B)A^{-1}p \right) \operatorname{grad} \pi - p^T (\operatorname{grad} \pi)^{\cdot}. \end{aligned} \tag{3.7}$$

Továbbá, bármely $\eta > 0$ számhoz létezik olyan $\delta(\eta)$ $(0 < \delta(\eta) < \eta/2)$, hogy ha $\eta \leq |(\tilde{q}, p)| \leq h'$ és $|p| \leq \delta$, akkor

$$\dot{A}_1(q, p, t) \geq |\operatorname{grad} \pi(q)|^2 - c_3 \delta^2 - c_4 \delta_{K_1 \times R^{r-s}}(t) =: \xi(t),$$

$(c_1, \dots, c_4$ pozitív konstansok). Könnyű belátni, hogy a (6) feltétel maga után vonja a

$$\limsup_{\substack{t \to \infty \\ T \to \infty}} \frac{1}{T} \int_t^{t+T} \gamma_{K \times R^{r-s}}(\tau) d\tau < \infty$$

egyenlőtlenség teljesülését. Ha δ-t elég kicsire választjuk, akkor nyilván $\int_T^\infty \xi = \infty$ T-ben egyenletesen.

A fenti becslések mutatják, hogy a 3.1.4.–3.1.5. és 3.2.2. tétel feltételei teljesülnek az $x = (q,p)^T$, $y = (\tilde{q},p)^T$, $\alpha(t) = \beta(t)$, $U(\tilde{q},p) = -c_5|p|^2$ $(0 < c_5 = \text{konst.})$, $F = \{(\tilde{q},p) : |\tilde{q}| \le h',\ p = 0\}$ választás mellett. Az említett tételek éppen jelen tételünk állítását adják.

3.3.7. KÖVETKEZMÉNY. *Tegyük fel, hogy*

(i) *a π potenciális energiának a $q = 0$ pontban helyi minimuma van;*
(ii) *a (3.5) rendszer $q = 0$ egyensúlyi helyzete izolált;*
(iii) *a disszipáció gyengén integrálisan teljes;*
(iv) *valamely $h' > 0$-ra az*

$$\int_0^t \max\{|G(q,s) - B(q,s)| : |q| \le h'\} ds$$

függvény egyenletesen folytonos R_+-on.

Akkor a (3.5) rendszer $q = \dot{q} = 0$ egyensúlyi állapota ekviaszimptotikusan stabilis. Ha a disszipáció még integrálisan is teljes, akkor az aszimptotikus stabilitás egyenletes.

Ha (i) helyett azt tesszük fel, hogy a π potenciális energiának a $q = 0$-ban nincs helyi minimuma, akkor a $q = \dot{q} = 0$ egyensúlyi állapot instabilis.

Bizonyítás. A 3.3.5.–3.3.6. tételt alkalmazzuk a $q = q$ $(s = r)$ triviális partícióra.

Először bebizonyítjuk, hogy a potenciális energiának a $q = 0$ pontban szigorú helyi minimuma van, vagyis létezik olyan $a_1 \in \mathcal{K}$, hogy $a_1(|q|) \le \pi(q)$ az origó egy kis környezetében. Ha ez nem lenne így, akkor a $q = 0$ ponthoz akármilyen közel lenne olyan q_0, ahol $\pi(q_0) = 0$. De ekkor q_0 is minimumhelye π-nek, tehát grad $\pi(q_0) = 0$, vagyis q_0 egyensúlyi helyzet, ami ellentmond (ii)-nek.

Mivel $\pi(0) = 0$ és π folytonos a $q = 0$-ban, a $b_1 \in \mathcal{K}$ függvény létezése mindkét esetben nyilvánvaló.

A $q = 0$ egyensúlyi helyzet izolált, tehát az origóhoz elég közel grad $\pi(q) \ne 0$, ahonnan (2) következik, ha h' elég kicsi.

A mechanikából ismeretes, hogy $A^{-1}(q)$ minden rögzített q-ra pozitív definit. Mivel most $|q| \le h'$, ebből a (3) feltétel teljesülése következik. □

3.3.8. KÖVETKEZMÉNY. *Tegyük fel, hogy a nem-ellenőrzött koordináták a mozgások során egyenletesen korlátosak, azaz léteznek olyan $\delta > 0$ és M számok, hogy ha $|(q_0, \dot{q}_0)| < \delta$ és $0 \le t_0 \le t$, akkor $|\hat{q}(t; q_0, \dot{q}_0, t_0)| \le M$.*

Tegyük fel, továbbá, hogy

(i) *a π potenciális energiának a $q = 0$ pontban helyi minimuma van;*
(ii) *ha a $(\tilde{q}, \hat{q})^T$ pont az origó egy elég kis környezetében lévő egyensúlyi helyzet, akkor $\tilde{q} = 0$;*

(iii)-(iv) *teljesül a 3.3.7. következmény (iii)-(iv) feltétele.*

Akkor a (3.5) rendszer $q = \dot{q} = 0$ egyensúlyi állapota ekviaszimptotikusan $(\tilde{q}, \dot{q})^T$-stabilis. Ha a disszipáció még integrálisan is teljes, akkor az aszimptotikus stabilitás egyenletes.

Ha (i) helyett azt tesszük fel, hogy a π potenciális energiának a $q = 0$ pontban nincs helyi minimuma, akkor az egyensúlyi állapot $(\tilde{q}, \dot{q})^T$-instabilis.

3.4. Kiegészítő megjegyzések

Ez a fejezet [21] dolgozatomra épül.

A (3.3.) egyenlet hosszú idők óta a stabilitási vizsgálatok gyakori tárgya. A korábbi eredményekről jó áttekintést ad a [11] monográfia. A legújabb eredmények között meg kell említeni KARSAI JÁNOS két dolgozatát. [32]-ben — továbbfejlesztve egy korábbi dolgozatomban kidolgozott módszeremet — olyan tételeket bizonyít, amelyek figyelembe veszik az $a(t)$ surlódási együtthatót alulról és felülről korlátozó függvények kölcsönhatását. A [31]-ben lévő feltételek megengedik $a(t)$-nek „hosszú" intervallumokon való eltűnését is (szakaszos fékezés).

A (3.5) rendszer egyensúlyi helyzetei stabilitási tulajdonságainak meghatározása a klasszikus mechanika egyik központi kérdése [46]. Egészen a közelmúltig azonban szinte kizárólag a stacionárius esetet vizsgálták [70]. [21] dolgozatomhoz kapcsolódva S. MURAKAMI ad érdekes eredményeket a (3.5) rendszerre 1984-ben megjelent [39] dolgozatában. A [79] dolgozatban TERJÉKI JÓZSEF-fel közösen már olyan rendszereket tárgyalunk, amelyekben a potenciális energia is függhet az időtől.

4. fejezet

Parciális stabilitási tulajdonságok autonóm rendszerekre

Ebben a fejezetben olyan módszert dolgozunk ki, amely kifejezetten parciális aszimptotikus stabilitás, illetve parciális instabilitás megállapítására alkalmas olyan *Ljapunov-függvény* birtokában, amelynek a rendszer szerinti deriváltja csak negatív szemidefinit. Ilyen feltételeket autonóm rendszerekre célszerűnek látszik a *Barbasin-Kraszovszkij módszer* általánosításával levezetni. Ez sikerrel meg is történt V.V. RUMJANCEV [76], C. RISITO [44] és A. S. OZIRANER [72] dolgozataiban. Az általánosítás során azonban egy új probléma merült fel: annak érdekében, hogy a megoldások határhalmazán annak invarianciáján alapuló technikát átmentsék, mindannyian feltették, hogy a nem-ellenőrzött koordináták a mozgások során korlátosak. Ez a feltétel két szempontból is szerencsétlennek, természetellenesnek tűnik. Egyrészt, a nem-ellenőrzött koordinátákra vonatkozik, pedig azokról a legtöbbször nincs információnk. Másrészt, a direkt módszer alapelvének ellentmondva, „a priori" ismereteket tételez fel a megoldásokról, és így nem ellenőrizhető a megoldások ismerete nélkül.

4.1.* TÉTEL. (A. S. OZIRANER [72]). *Tegyük fel, hogy az*

$$\dot{x} = X(x) \quad (X(0) = 0) \tag{4.1}$$

autonóm rendszer minden, az origóhoz elég közelről induló megoldása z-korlátos (azaz a megoldások ellenőrizetlen koordinátái korlátos függvények R_+-on).

I. Tegyük fel, hogy létezik a (4.1) rendszerhez olyan $V : G'_m \to R_+$ Ljapunov-függvény, amely teljesíti a következő feltételeket:

(1) V pozitív y-definit;

(2) az $\{x : V(x) > 0\} \cap \dot{V}^{-1}(0)$ halmaz nem tartalmazza a (4.1) rendszer egyetlen teljes trajektóriáját sem.

Ezen feltételek mellett a (4.1) rendszer 0-megoldása egyenletesen aszimptotikusan y-stabilis.

II. Tegyük fel, hogy létezik olyan $V \in Lip(G'_m; R)$ függvény, amely teljesíti a következő feltételeket:

(1) $V(0) = 0$, és az $x = 0$ pont bármely környezetében van olyan pont, ahol V negatív értéket vesz fel;

(2) az $\{x : V(x) < 0\} \cap \dot{V}^{-1}(0)$ halmaz nem tartalmazza a (4.1) rendszer egyetlen teljes trajektóriáját sem.

Ezen feltételek mellett a (4.1) rendszer 0-megoldása y-instabilis. □

A korlátossági feltétel a bizonyításban igen lényeges szerepet kap: biztosítja, hogy a megoldások határhalmaza nem üres. A triviális $\dot{y} = -y$, $\dot{z} = z$ rendszer mutatja, hogy a 0-megoldás úgy is lehet egyenletesen aszimptotikusan y-stabilis, ha a megoldások nem z-korlátosak. Másrészt viszont, ha a 4.1* tételből a korlátossági feltételt elhagyjuk, akkor az állítás nem igaz, mint ahogyan a következő egyszerű példa ([72]) mutatja.

Tekintsük az

$$\dot{y} = -y \setminus (1 + z^2), \quad \dot{z} = z$$

rendszert, amelynek általános megoldása:

$$y(t) = y_0 \exp\left[-\int_0^t \frac{ds}{1 + z_0^2 \exp[2s]}\right], \quad z(t) = z_0 \exp[t].$$

Ha $y_0 \neq 0$, $z_0 \neq 0$, akkor $y(t) \to y_\infty \neq 0$ $(t \to \infty)$, tehát a 0-megoldás nem aszimptotikusan y-stabilis. Másrészt, viszont, a 4.1* tétel I. részének minden feltétele teljesül a $V(y,z) = y^2/2$ *Ljapunov-függvénnyel* a z-korlátossági feltétel kivételével. Valóban, $\dot{V}^{-1}(0) = \{(y,z) : y = 0\}$, tehát a (2) feltételben szereplő halmaz üres.

Hasonlóan látható be, hogy az

$$\dot{y} = y \setminus (1 + z^2), \quad \dot{z} = z$$

rendszer a $V(y,z) = -y^2$ függvénnyel teljesíti a 4.1* tétel II. részének feltételeit a z-korlátosság kivételével, a 0-megoldás mégis y-stabilis.

A probléma tehát: Milyen, közvetlenül ellenőrizhető feltételekkel lehet a 4.1.* tételben a z-korlátossági feltételt helyettesíteni ?

A probléma megoldását azzal kezdjük, hogy megadjuk az $\Omega_m(\phi)$ határhalmaz egy lokalizációját. Ennek felhasználásával igen egyszerűen osztályozni tudjuk a megoldásokat, ha a 4.1.* tétel feltételei teljesülnek a z-korlátossági feltétel kivételével. Ez lehetővé teszi, hogy a *Ljapunov-függvényre* tett további megszorítással aszimptotikus y-stabilitást, illetve y-instabilitást biztosító feltételeket nyerjünk. Eredményeinket K. PEIFFER és N. ROUCHE [43] egy problémájára alkalmazzuk, amely gravitációs térben adott felületen mozgó anyagi pont egyensúlyi helyzetének stabilitási tulajdonságaira vonatkozik.

4.1. Parciális határhalmazok. Egy alternatíva a megoldások viselkedésére

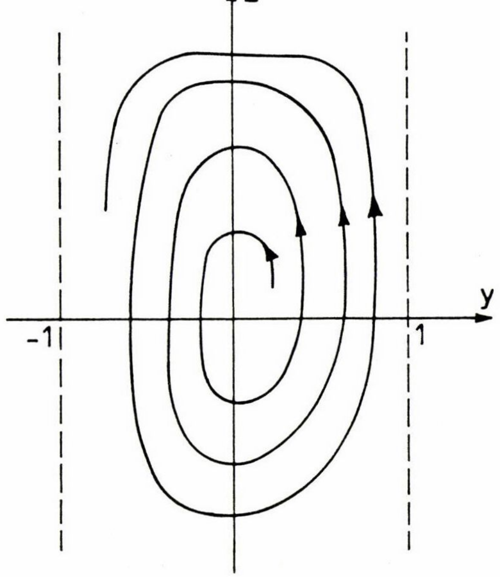
4.1.1. *Definició.* Tekintsük a (4.1) egyenlet egy tetszőleges $\phi : [t_0,\omega) \to R^k$ megoldását. Jelölje $\psi : [t_0,\omega) \to R^m$ a megoldás ellenőrzött, $\chi : [t_0,\omega) \to R^n$ pedig a nem-ellenőrzött koordinátáiból álló vektort. Az $x = \phi(t)$ megoldáshoz tartozó $\Omega_m(\phi)$ *parciális határhalmaz* azon $q \in R^m$ pontokból áll, amelyek mindegyikéhez létezik olyan $\{t_i\}_{i=1}^{\infty}$ sorozat, hogy $t_i \to \omega - 0$, $\psi(t_i) \to q \quad (i \to \infty)$.

Jelölje $P_m : R^k \to R^m$ az R^k térből az R^m altérre való merőleges vetítés operátorát. Nyilvánvalóan igaz az $\Omega_m(\phi) \supset P_m(\Omega(\phi))$ tartalmazás. A *Bolzano-Weierstrass-tétel* alkalmazásával könnyű látni, hogy ha $\chi : [t_0,\omega) \to R^n$ korlátos, akkor a két halmaz megegyezik. A 4.1. ábra egy olyan esetet mutat, amikor az előbbi tartalmazás valódi:

$$\Omega(\phi) = \{y,z) : y = 1, \text{ vagy } y = -1, \ z \in R\}$$
$$\Omega_m(\phi) = \{(y,z) : -1 \le y \le 1, \ z = 0\}$$
$$P_m(\Omega(\phi)) = \{(-1,0),(1,0)\}$$

A parciális határhalmaz sok tulajdonságot örököl az összes koordinátákra vonatkozó határhalmaztól. Nevezetesen, teljesen hasonlóan be lehet bizonyítani, hogy ha $\psi : [t_0,\omega) \to R^m$ korlátos, akkor $\omega = \infty$, $\Omega_m(\phi)$ nem üres ($\Omega(\phi)$ lehet üres !), kompakt, összefüggő halmaz, továbbá a legszűkebb azon zárt halmazok közül, amelyekhez $\psi(t)$ konvergál, ha $t \to \infty$. (Ez a legutóbbi tulajdonság mutatja, hogy az $\Omega_m(\phi)$ határhalmaz meghatározó az y-stabilitási tulajdonságok szempontjából.) Nem örökli viszont a parciális határhalmaz a legfontosabb tulajdonságot: $\Omega_m(\phi)$ általában nem invariáns a (4.1) rendszerre vonatkozóan. Ez azt jelenti, hogy az invariancia-elv itt nem érvényes; hasonlóan a *Barbasin-Kraszovszkij tétel* nem-autonóm rendszerekre való általánosításához, $\Omega_m(\phi)$ lokalizálásához új módszert kell keresni.

4.1.2. LEMMA. *Tegyük fel, hogy* $V : G'_m \to R$ *a (4.1) rendszerhez tartozó Ljapunov-függvény, és* $x = \phi(t) = (\psi(t),\chi(t))^T$ *olyan megoldása (4.1)-nek a* $[t_0,\infty)$



4.1 ábra

intervallumon, amelyre $|\psi(t)| \leq h'' < h'$ $(t \geq t_0)$*. Akkor vagy a)* $|\chi(t)| \to \infty$*, ha* $t \to \infty$*, vagy b)* $v(t) = V(\phi(t)) \to V_0 \in R$*, ha* $t \to \infty$*,* $\Omega(\phi)$ *nem üres és* $\Omega(\phi) \subset V^{-1}(v_0) \cap \dot{V}^{-1}(0)$.

Bizonyítás. Tegyük fel, hogy a) nem teljesül. Akkor ψ korlátosságát is felhasználva következtethetünk egy olyan $\{t_i\}_{i=1}^{\infty}$ sorozat és $q \in R^m$, $r \in R^n$, $v_0 \in R$ létezésére, amelyekre teljesülnek: $t_i \to \infty$, $\psi(t_i) \to q$, $\chi(t_i) \to r$ és $v(t_i) \to v_0$, ha $i \to \infty$. $\Omega(\phi)$ invariáns (4.1)-re vonatkozóan, tehát létezik (4.1)-nek olyan $x = x(t; q, r)$ megoldása, hogy $\gamma(x) \subset \Omega(\phi)$. Legyen $\bar{t} \in R_+$ rögzítve. Ekkor létezik olyan $\{\bar{t}_i\}$ sorozat, hogy $\bar{t}_i \to \infty$, $\phi(\bar{t}_i) \to x(\bar{t}; q, r)$ és $v(\bar{t}_i) \to V(x(\bar{t}; q, r)$:), ha $i \to \infty$. De $v(t)$ monoton csökkenő, tehát $V(x(\bar{t}; q, r)) = v_0$, ahonnan E tetszőleges volta miatt az állítás már következik. □

4.1.3. LEMMA. *Tegyük fel, hogy* V *és* ϕ *olyanok, mint az előző lemmában, továbbá legyen* V *alulról korlátos* G'_m*-n. Akkor*

$$\Omega_m(\phi) \subset P_m(\Omega(\phi)) \cup V_m^{-1}[v_0, \infty],$$

ahol $v_0 = \lim\limits_{t \to \infty} V(\phi(t))$.

Bizonyítás. Tegyük fel, hogy $q \in \Omega_m(\phi)$, de $q \notin P_m(\Omega(\phi))$. Ekkor létezik olyan $\{t_i\}$ sorozat, hogy $t_i \to \infty$, $\psi(t_i) \to q$, $|\chi(t_i)| \to \infty$, $v(t_i) \to v_0$, ha $i \to \infty$, azaz $q \in V_m^{-1}[v_0, \infty]$. □

4.1.4. TÉTEL. *Legyen* $V : G'_m \to R$ *a (4.1) egyenlethez tartozó olyan pozitív y-definit Ljapunov-függvény, hogy tetszőleges* $c > 0$ *számra a* $V^{-1}(c) \cap \dot{V}^{-1}(0)$ *halmaz nem tartalmazza az egyenlet egyetlen teljes trajektóriáját sem. Ekkor a (4.1) 0-megoldása y-stabilis, és tetszőleges, az origó elég kicsiny környezetéből induló* $x(t) = (y(t), z(t))$ *megoldásra vagy a)* $V(x(t)) \to 0$ *(és így* $|y(t)| \to 0$*), vagy b)* $|z(t)| \to \infty$*, ha* $t \to \infty$.

Bizonyítás. Mivel V pozitív y-definit és $\dot{V}(x) \leq 0$, a 0-megoldás y-stabilis (l. [46], 22. o.). Legyen $x = \phi(t) = (\psi(t), \chi(t))^T$ egy olyan megoldás, amelyre ψ korlátos. Tegyük fel, hogy $|\chi(t)| \not\to \infty$, ha $t \to \infty$. Ekkor a 4.1.2. lemma szerint $v(t) = V(\phi(t)) \to v_0 \geq 0$, és létezik olyan $p \in R^k$ pont, hogy $p \in \Omega(\phi) \subset V^{-1}(v_0) \cap \dot{V}^{-1}(0)$. De az $\Omega(\phi)$ halmaz invariáns, így (4.1)-nek létezik trajektóriája a $V^{-1}(v_0) \cap \dot{V}^{-1}(0)$ halmazban. A feltételek szerint ekkor $v_0 = 0$. □

Oziraner tétele (a 4.1.* tétel) a most bizonyított tételnek nyilvánvaló következménye (az egyenletességre vonatkozóan l. az első fejezet 1.6. megjegyzését).

4.2. Tételek az aszimptotikus stabilitásról és az instabilitásról

4.2.1. TÉTEL. *Tegyük fel, hogy a 4.1.4. tétel feltételei teljesülnek. Tegyük fel, továbbá, hogy* $V(y, z) \to 0$ *a* $B_m(0, h')$ *gömbben y-ra vonatkozóan egyenletesen, ha* $\dot{V}(y, z) \to 0$ *és* $|z| \to \infty$*. Ekkor a (4.1) egyenlet 0-megoldása egyenletesen aszimptotikusan y-stabilis.*

Bizonyítás. Az 1.6. megjegyzés szerint elegendő azt bizonyítani, hogy a V *Ljapunov-függvény* minden olyan $\phi(t) = (\psi(t), \chi(t))^T$ megoldás mentén 0-hoz tart, amelyre $|\psi(t)| \leq h'' < h'$ egy alkalmas h'' konstanssal. Mint tudjuk, $v(t) = V(\phi(t)) \to v_0 \geq 0$, ha $t \to \infty$, tehát létezik olyan $\{t_i\}$ sorozat, hogy $t_i \to \infty$ és $\dot{V}(\phi(t_i)) \to 0$, ha $i \to \infty$. A 4.1.4. tétel szerint vagy a $v_0 = 0$, vagy $\{z_i = \chi(t_i)\}$ sorozat normában ∞-hez divergál. A második esetben $\dot{V}(\psi(t_i), z_i) \to 0$, $|z_i| \to \infty$, és ugyanakkor $V(\psi(t_i), z_i) \to v_0$, ha $i \to \infty$. A tétel utolsó feltétele szerint ebből következik, hogy $v_0 = 0$. □

Hasonlítsuk össze a most bizonyított tételt K. PEIFFER és N. ROUCHE egy tételével ([43]), Th. IV), amely a (4.1) autonóm rendszerre a következőt mondja: Legyen $V : G'_m \to R$ pozitív y-definit *Ljapunov-függvény*, és $V(x) \to 0$ a G'_m halmazon egyenletesen x-re vonatkozóan, ha $\dot{V}(x) \to 0$. Ekkor (4.1) 0-megoldása egyenletesen aszimptotikusan stabilis. A 4.2.1. tétel tartalmazza ezt az eredményt, sőt — mint azt a következő példa mutatja — élesíti is.

Tekintsük az

$$\dot{y}_1 = -y_1(1 + z^2), \ \dot{y}_2 = -y_2, \ \dot{z} = z - z^3$$

rendszert a $V(y_1, y_2, z) = y_1^2 + y_2^2 + y_2^2 z^2$ *Ljapunov-függvénnyel.* Ez nyilván pozitív (y_1, y_2)-definit, és a rendszer szerinti deriváltja

$$\dot{V}(y_1, y_2, z) = -2y_1^2(1 + z^2) - 2y_2^2 - 2y_2^2 z^4.$$

Ha $\dot{V} \to 0$, akkor $y_1 \to 0$, $y_2 \to 0, y_2^2 z^4 \to 0$, és ezekből nem következik, hogy $V \to 0$, vagyis a *Peiffer–Rouche-tételt* erre az esetre nem lehet alkalmazni. Ugyanakkor $y_1 \to 0$, $y_2 \to 0$, $y_2^2 z^4 \to 0, |z| \to \infty$ együtt maga után vonja, hogy $V \to 0$, és a 4.2.1. tétel alkalmazható.

Az alkalmazások során látni fogjuk, hogy a 4.1.*, 4.1.4. és 4.2.1. tételben szereplő azon feltétel, mely szerint „bármely $c > 0$ számra a $V^{-1}(c) \cap \dot{V}^{-1}(0)$ halmaz nem tartalmazhat teljes trajektóriát", túl szigorúnak tűnik. Persze, már elméleti szempontból nézve is feltűnik, hogy a feltétel nem eléggé alkalmazkodik a feladathoz, hiszen nem tesz különbséget ellenőrzött és nem-ellenőrzött koordináták között. C. RISITO [44] — igaz, hogy meglehetősen bonyolult módon — gyengébb feltétel mellett is bizonyít aszimptotikus y-stabilitást, egyenletesség nélkül.

4.2.2.* TÉTEL. (C. RISITO [44]). *Tegyük fel, hogy a (4.1) rendszer minden, az origóhoz elegendően közelről induló megoldása z-korlátos, és az $\{(y,z) : y = 0\}$ halmaz invariáns a rendszerre nézve. Tegyük fel továbbá, hogy a rendszerhez létezik olyan $V : G'_m \to R_+$ Ljapunov-függvény, amely teljesíti a következő feltételeket:*

(1) V pozitív y-definit;

(2) a $\dot{V}^{-1}(0) \setminus \{(y,z) : y = 0\}$ halmaz nem tartalmaz teljes trajektóriát.

Ekkor a (4.1) rendszer 0-megoldása aszimptotikusan y-stabilis. □

A 4.1.4. tételben megfogalmazott alternatíva és a lemmák segítségével egyszerűen megkapható *Risito tételének* általánosítása olyan esetekre, amikor a z-korlátosságra és invarianciára vonatkozó feltétel nem teljesül.

4.2.3. TÉTEL. *Tegyük fel, hogy $V : G'_m \to R$ egy pozitív y–definit Ljapunov–függvény a (4.1) rendszerre vonatkozóan, amely bármely $c > 0$ számra rendelkezik a következő tulajdonságokkal:*

(1) a (4.1) rendszernek a $V^{-1}(c) \cap \dot{V}^{-1}(0)$ halmazban fekvő minden teljes trajektóriája részhalmaza az $\{(y,z) : y = 0\}$ altérnek is;

(2) $V_m^{-1}[c, \infty] \subset \{0\}$.

Ekkor a (4.1) rendszer 0-megoldása aszimptotikusan y-stabilis.

Bizonyítás. A 4.1.4. tétel bizonyításának elején már beláttuk, hogy a 0-megoldás y-stabilis. Legyen $\phi(t) = (\psi(t), \chi(t))^T$ egy olyan megoldás, amelyre ψ korlátos R_+-on. Azt fogjuk bebizonyítani, hogy $\Omega_m(\phi) = \{0\}$.

Mint tudjuk, $v(t) = V(\phi(t)) \to v_0 \geq 0$, ha $t \to \infty$. Ha $v_0 = 0$, akkor készen vagyunk, hiszen V pozitív y-definit. Tegyük fel, hogy $v_0 > 0$, és az állítás nem igaz, vagyis létezik az $\Omega_m(\phi)$ halmaznak egy $q \neq 0$ eleme. A 4.1.3. lemma szerint ekkor a (2) feltételből az következik, hogy van olyan $p \in \Omega(\phi)$, hogy $q = P_m p$. De $\Omega(\phi)$ invariáns a (4.1) rendszerre vonatkozóan, tehát (4.1)-nek létezik olyan $\xi : R \to R^k$ megoldása, amelyre $\xi(0) = p$, és $\gamma(\xi) \subset \Omega(\phi)$. A 4.1.2. lemma szerint $\gamma(\xi)$ részhalmaza a $V^{-1}(v_0) \cap \dot{V}^{-1}(0)$ halmaznak is. Másrészt, $q \neq 0$, tehát a $\gamma(\xi)$ trajektóriát az $\{(y,z) : y = 0\}$ halmaz nem tartalmazza, ami ellentmond az (1) feltételeknek. □

4.2.4. KÖVETKEZMÉNY. *Tegyük fel, hogy a 4.2.3. tétel (1) feltétele teljesül, és létezik olyan $\rho > 0$ szám, amellyel a $0 < |\bar{y}| < \rho$ halmazon a*

$$\lim_{\substack{y \to \bar{y} \\ |z| \to \infty}} V(y, z) = \infty$$

reláció érvényes.

Ekkor a (4.1) rendszer 0-megoldása aszimptotikusan y-stabilis.

A 4.1.2. lemmában bizonyított alternatíva alkalmas instabilitási feltételek levezetésére is.

4.2.5. TÉTEL. *Tegyük fel, hogy a (4.1) rendszerhez létezik olyan $V \in \mathrm{Lip}\,(G'_m; R)$ függvény, amely rendelkezik a következő tulajdonságokkal:*

(1) minden $\delta > 0$ számhoz létezik olyan $x_0(\delta) \in B_k(0, \delta)$, hogy $V(x_0(\delta)) < 0$;

(2) létezik olyan ε_0 $(0 < \varepsilon_0 < h')$, hogy bármely $c < 0$ számra a

$$V^{-1}(c) \cap \dot{V}^{-1}(0) \cap \bar{B}_m(0, \varepsilon_0) \times R^n) \tag{4.2}$$

halmaz nem tartalmazza a (4.1) rendszer egyetlen trajektóriáját sem.

Akkor bármely $\varepsilon > 0$-ra vagy a) minden $\gamma_m^+(x_0(\delta)) : t \mapsto y(t; x_0(\delta))$ $(t \in R_+)$ görbe véges idő alatt kilép a $B_m(0, \varepsilon_0)$ gömbből, vagy b) $|z(t; x_0(\delta))| \to \infty$, ha $t \to \infty$.

Bizonyítás. Ha az állítás nem igaz, akkor valamely δ_0 $(0 < \delta_0 < \varepsilon_0)$ számhoz létezik olyan

$$\phi(t) = (\psi(t), \chi(t))^T \quad (\phi(0) = x_0(\delta_0)) \tag{4.3}$$

megoldás, amelyre teljesülnek a következők:

$$|\psi(t)| \leq \varepsilon_0 \ (t \in R_+); \ v(t) = V(\phi(t)) \to v_0 < 0 \ (t \to \infty), \tag{4.4}$$

az $\Omega(\phi)$ határhalmaz nem üres és a (4.2) halmazban fekszik $c = v_0$-val. Másrészt $\Omega(\phi)$ invariáns, így (4.2) tartalmaz egy teljes trajektóriát, ami ellentmondás. □

4.2.6. KÖVETKEZMÉNY. *Tegyük fel, hogy a 4.2.5. tétel feltételei teljesülnek, továbbá*

$$\liminf_{|z| \to \infty} V(y, z) \geq 0 \tag{4.5}$$

y-ra vonatkozóan egyenletesen a $\bar{B}_m(0, \varepsilon_0)$ halmazon.

Ekkor a (4.1) rendszer 0-megoldása y-instabilis.

4.2.7. TÉTEL. *Tegyük fel, hogy a 4.2.5. tétel feltételei teljesülnek, továbbá a V függvény alulról korlátos a $\bar{B}_m(0, \varepsilon_0) \times R^n$ halmazon, és*

$$\liminf_{\substack{|z| \to \infty \\ \dot{V}(y,z) \to 0}} V(y, z) \geq 0$$

y-ra vonatkozóan egyenletesen a $\bar{B}_m(0,\varepsilon_0)$ halmazon.
Ekkor a (4.1) rendszer 0-megoldása y-instabilis.

Bizonyítás. Azt bizonyítjuk be, hogy bármely δ-ra a 4.2.5. tételben definiált tetszőleges $\gamma_m^+(x_0(\delta))$ görbe kilép a $B_m(0,\varepsilon_0)$ gömbből. Tegyük fel, hogy ez nem igaz. Akkor valamely δ_0-ra $(0 < \delta_0 < \varepsilon_0)$ a (4.2) megoldás rendelkezik a (4.4) tulajdonságokkal. A 4.2.5. tétel szerint ekkor $|\chi(t)| \to \infty$, ha $t \to \infty$. Hasonlóan, mint a 4.2.1. tétel bizonyításában, ebből be lehet bizonyítani, hogy a (4.5) feltétel miatt $v_0 = 0$, ami ellentmondás. □

4.3. Alkalmazások mechanikai rendszerekre

Tekintsük ismét a 3. fejezet 2. pontjában már tanulmányozott mechanikai rendszert, amelyre disszipatív és giroszkópikus erők hatnak (l. (3.5) egyenlet). Most azonban feltesszük, hogy ezek az erők sem függnek explicit módon az időtől, megengedjük viszont, hogy a sebességtől nem-lineárisan is függhetnek. Ekkor a mozgásegyenlet

$$\frac{d}{dt}\frac{\partial T}{\partial \dot{q}} - \frac{\partial T}{\partial q} = -\frac{\partial \pi}{\partial q} + Q, \quad (q,\dot{q} \in R^r) \tag{4.6}$$

ahol $Q : R^{2r} \to R^r$ a disszipatív és giroszkópikus erők eredője, azaz feltesszük, hogy $Q^T(q,\dot{q})\dot{q} \leq 0 \quad (q,\dot{q} \in R^r)$. Ilyen surlódási erők mellett akkor mondjuk, hogy a *disszipáció teljes*, ha egy alkalmas $c \in \mathcal{K}$ függvénnyel teljesül a $Q^T(q,\dot{q})\dot{q} \leq -c(|\dot{q}|)$ egyenlőtlenség az R^{2r} halmazon. Könnyű belátni, hogy a *Salvadori-tétel* (a 3.3.4.* tétel) ilyen surlódási erők mellett is változatlanul érvényes. Lényeges azonban a *Salvadori-tétel* (2) feltétele, amely azt kívánja, hogy a $q = 0$ egyensúlyi helyzet izolált legyen. Képzeljük el, például, egy golyónak gravitációs térben vízszintes sík lapon való mozgását surlódási erő hatása alatt. Nyilván egyetlen egyensúlyi helyzet sem lesz aszimptotikusan stabilis, pedig a *Salvadori-tétel* összes többi feltétele teljesül. Aszimptotikusan stabilis lesz viszont minden egyensúlyi helyzet a sebességre vonatkozóan. Vagy tekintsük az (y,z,u) koordinátarendszerben (az „u" tengely felfelé mutat) az $u = y^2$ felület mentén való mozgást. Ekkor a z-tengely minden pontja egyensúlyi helyzet, tehát az $y = 0$, $z = 0$ egyensúlyi helyzet nem izolált, mégis úgy érezzük, hogy y-ra vonatkozóan aszimptotikusan stabilis lesz. Világos tehát, hogy a (2) feltétel gyengítésekor parciális stabilitási tulajdonságok levezetésére kell törekednünk. Ugyanerre a következtetésre jutunk, ha a disszipáció teljességére vonatkozó feltételt gyengítjük, és megengedjük, hogy bizonyos irányokban nem hat disszipatív erő.

A sebességre vonatkozó aszimptotikus stabilitással egy külön pontban fogunk foglalkozni egy későbbi fejezetben, együtt az instacionárius (nem-autonóm) esettel. Most az általánosított koordináták egy részére vonatkozó aszimptotikus stabilitásra és instabilitásra adunk feltételeket. (Megjegyezzük, hogy *Salvadori tételében* az instabilitásról szóló részben is lényeges az egyensúlyi helyzet izoláltságának feltétele. K. PEIFFER [42] mutatott egy szellemes példát arra, hogy nem-izolált egyensúlyi helyzet teljes disszipációjú surlódás hatására még akkor is lehet stabilis, ha a potenciál C^∞ és az egyensúlyi helyzetben szigorú maximuma van!)

Legyen adva az általánosított koordinátáknak egy $q = (q_1, q_2)^T$ ($q_1 \in R^s$, $q_2 \in R^{r-s}$) partíciója.

4.3.1. TÉTEL. *Tegyük fel, hogy*

(1) π *pozitív* q_1*-definit;*

(2) létezik olyan $a \in \mathcal{K}$ *és* $\gamma, \rho > 0$, *amelyekkel a* $T(q,\dot q) \geq a(|\dot q_1|)$ *egyenlőtlenség teljesül a* $B_s(0,\gamma) \times R^{r-s} \times B_s(0,\rho) \times R^{r-s}$ *halmazon;*

(3) a disszipáció q_1*-teljes, azaz léteznek olyan* $b : R^r \to R_+$ *és* $c \in \mathcal{K}$ *függvények, hogy*

$$Q^T(q,\dot q)\dot q \leq -b(q_1,q_2)c(|\dot q|)$$

az előbbi halmazon, és ha $b(q_1,q_2) = 0$, *akkor* $q_1 = 0$.

(4) a $\{q = (q_1,q_2)^T : q_1 \neq 0\}$ *halmaz nem tartalmaz egyensúlyi helyzetet;*

(5) a $0 \in R^{2s}$ *egy elég kicsiny környezetében fekvő tetszőleges* $(q_1^0, \dot q_1^0)^T \neq 0$ *pontra*

$$\lim_{\substack{q_1 \to q_1^0,\, \dot q_1 \to \dot q_1^0 \\ |(q_2,\dot q_2)^T| \to \infty}} (T(q_i,\dot q_i) + \pi(q_i)) = \infty.$$

Ekkor a (4.6) rendszer $q = \dot q = 0$ *egyensúlyi állapota aszimptotikusan* $(q_1,\dot q_1)$*-stabilis.*

Bizonyítás. Alkalmazzuk a 4.2.4. következményt a $V = H(q,\dot q) = T + \pi$ *Ljapunov-függvénnyel.* A (3) feltétel szerint ennek a (4.6) rendszer szerinti deriváltja:

$$\dot H(q,\dot q) = Q^T(q,\dot q)\dot q \leq -b(q_1,q_2)c(|\dot q|),$$

tehát ha $\dot H(q,\dot q) = 0$, akkor $\dot q = 0$ vagy $q_1 = 0$, vagyis $\dot H^{-1}(0) \subset \{(q,\dot q) : q_1 = 0$ vagy $\dot q = 0\}$. Ha a (4.6) rendszer $t \mapsto (q(t),\dot q(t))^T$ mozgása az utóbbi halmazban fekszik, akkor az vagy egyensúlyi helyzet, vagy pedig teljesül a $q_1(t) \equiv 0$ ($t \in R_+$) azonosság, tehát a (4) feltétel miatt mindkét esetben a mozgás trajektóriája a $\{(q,\dot q)^T : q_1 = 0,\ \dot q_1 = 0\}$ halmazban fekszik. Ez azt jelenti, hogy a 4.2.4. következmény minden feltétele teljesül. □

A 4.2.4. következményből ugyanígy lehet levezetni aszimptotikus q_1-stabilitásra vonatkozó tételt is:

4.3.2. TÉTEL. *Tegyük fel, hogy*

(1) π *pozitív* q_1*-definit;*

(2) a disszipáció q_1*-teljes, vagyis bármely* $\rho > 0$ *számhoz léteznek olyan* $b : R^r \to R_+$ *és* $c \in \mathcal{K}$ *függvények, hogy*

$$Q^T(q,\dot q)\dot q \leq -b_\rho(q_1,q_2)c_\rho(|\dot q|) \quad (|q_1| \leq \rho,\ \rho_2 \in R^{r-s},\ \dot q \in R^r)$$

és a b_ρ *függvényre az is teljesül, hogy ha* $b_\rho(q_1,q_2) = 0$, *akkor* $q_1 = 0$;

(3) a $\{q = (q_1, q_2)^T : q_1 \neq 0\}$ *halmaz nem tartalmaz egyensúlyi helyzetet;*
(4) a $0 \in R^s$ *pontnak van olyan környezete, ahol minden* $q_1^0 \neq 0$ *pontra*

$$\lim_{\substack{q_1 \to q_1^0 \\ |q_2| \to \infty}} \pi(q_1, q_2) = \infty.$$

Ekkor a (4.6) rendszer $q = \dot{q} = 0$ *egyensúlyi állapota aszimptotikusan* q_1-*stabilis.* □

Az előző két tételben a surlódás részleges abban az értelemben, hogy a konfigurációs tér egy alterében nem lép fel surlódás. Természetes a következő kérdés: mit tudunk mondani az egyensúlyi helyzet stabilitásáról, ha a súrlódás abban az értelemben is részleges, hogy az általánosított sebességek egy alteréből származó sebesség–összetevők nem eredményeznek surlódást (vagyis bizonyos irányokban nincs súrlódás). Ekkor a (3) illetve (2) feltételekben szereplő becslés helyett a

$$Q^T(q, \dot{q})\dot{q} \leq -b(q_1, q_2)c(|\dot{q}_1|)$$

egyenlőtlenség teljesül.

Be lehet látni, hogy az állítások érvényben maradnak, ha a (4), illetve (3) feltétel helyett azt tesszük fel, hogy ha egy mozgás során $q_1(t) \equiv q_1^0$ $(t \in R_+)$, akkor $q_1^0 = 0$.

A tételekben szereplő utolsó feltételek igen erősnek tűnnek. Azzal az esettel, amikor az ott szereplő határértékek végesek, a következő fejezetben foglalkozunk, mert tárgyalásához egy új módszer szükséges.

Térjünk most rá arra a kérdésre, hogy milyen potenciális energia mellett lesz az egyensúlyi állapot instabilis. W. T. KOITER [34] bebizonyíotta, hogy ha a $q = 0$ egyensúlyi helyzet izolált a $\{q : \pi(q) < 0\}$ halmazon, és a π-nek a $q = 0$-ban nincs minimuma, akkor a $q = \dot{q} = 0$ instabilis. Alábbi tételünk (a $q = q_1$ speciális esetben) mutatja, hogy az egyensúlyi állapot valójában q-instabilis.

4.3.3. TÉTEL. *Tegyük fel, hogy alkalmas* $h', \varepsilon_0, \lambda_0$ $(0 < \varepsilon_0 < h',\ \lambda_0 > 0)$ *konstansokkal teljesülnek az alábbi feltételek:*

(1) bármely δ $(0 < \delta < \varepsilon_0)$ *számhoz létezik olyan* $q_0(\delta) \in B_r(0, \delta)$, *hogy* $P(q_0(\delta)) < 0$;

(2) $T(q, \dot{q}) \geq \lambda_0 |\dot{q}|^2$ $(|q_1| \leq h',\ q_2 \in R^{r-s},\ \dot{q} \in R^r)$;

(3) a $\{q : \Pi(q) < 0,\ |q_1| < \varepsilon_0\}$ *halmaz nem tartalmaz egyensúlyi helyzetet;*

(4) a *disszipáció teljes.*

Ekkor bármely δ $(0 < \delta < \varepsilon_0)$ *számra vagy a)* a $\gamma_{q_1}^+(q_0(\delta), 0) : t \mapsto q_1(t; q_0(\delta), 0)$ $(t \in R_+)$ *görbe véges idő alatt kilép a* $B_s(0, \varepsilon_0)$ *gömbből, vagy* *b)* $|q_2(t; q_0(\delta), 0)| \to \infty$ $(t \to \infty)$.

Bizonyítás. Alkalmazzuk a 4.2.5. tételt a $V = H = T + \pi$, $y = q_1$ választással. □

A 4.2.6. következményből az alábbi eredmény adódik.

4.3.4. KÖVETKEZMÉNY. *Tegyük fel, hogy a 4.3.3. tétel feltételei teljesülnek, továbbá*

$$\liminf_{|q_2|\to\infty} \pi(q_1, q_2) \geq 0 \tag{5}$$

egyenletesen a q_1-re nézve a $\bar{B}_s(0, h')$ halmazon.

Ekkor a (4.6) rendszer $q = \dot{q} = 0$ egyensúlyi állapota q_1-instabilis.

A jelen fejezetet *egy példával* zárjuk. K. PEIFFER és N. ROUCHE [43] vizsgálta egységnyi tömegű anyagi pontnak homogén gravitációs térben az $u = (1/2)y^2(1+z^2)$ egyenletű felület mentén való mozgását (az Oy, z, u inerciarendszerben az u tengely felfelé mutat). A rendszer kinetikai energiája és a potenciális energia:

$$T = \frac{1}{2}(\dot{y}^2 + \dot{z}^2) + \frac{1}{2}y^2(\dot{y}(1+z^2) + \dot{z}yz)^2, \quad \pi = \frac{g}{2}y^2(1+z^2),$$

ahol g a gravitációs gyorsulás nagysága. PEIFFER és ROUCHE feltételezte, hogy a disszipatív erőt a

$$Q_1 = -\alpha\frac{\partial T}{\partial \dot{y}}, \quad q_2 = -\alpha\frac{\partial T}{\partial \dot{z}} \quad (0 < \alpha = \text{konst.}) \tag{4.7}$$

képletek definiálják.

A $H = T + \pi$ teljes mechanikai energia $(y, \dot{y}, \dot{z})$-ra vonatkozóan pozitív definit, tehát az $y = z = \dot{y} = \dot{z} = 0$ egyensúlyi állapot ezekre a változókra nézve stabilis. PEIFFER és ROUCHE bebizonyították, hogy $\dot{y}$-ra vonatkozóan aszimptotikusan stabilis. (Meg kell említeni, hogy az első közelítésben való stabilitás–vizsgálatról szóló *Ljapunov–Malkin–tétel* ([66], 113. o.) alkalmazásával aszimptotikus $(y, \dot{y}, \dot{z})$-stabilitás adódik!)

Tekintsünk el most a meglehetősen erőltetettnek tűnő (4.7) feltételtől, és tegyük fel, hogy tetszőleges olyan (akár nem-lineáris) surlódási erő hat a pontra, amelyre nézve a disszipáció y-teljes. (Ekkor — az esetleges nem–linearitás miatt — a *Ljapunov–Malkin* tétel sem alkalmazható!). A 4.3.1. tételből azonnal adódik, hogy az egyensúlyi állapot aszimptotikusan $(y, \dot{y})$-stabilis.

Eredményeinkkel vizsgálható egy tetszőleges $u = f(y, z)$ felület mentén való mozgás is. A 4.3.2. tétel és 4.3.4. következmény felhasználásával kaphatók a $\pi = g \cdot f(y, z)$ potenciális energiára aszimptotikus stabilitást, illetve instabilitást biztosító feltételek.

Például, legyen

$$f(y,z) = \begin{cases} (1/2)y^2(1+z^2) + \exp[-1/|z|]\sin^2(1/z^2) & z \neq 0 \\ 0 & z = 0 \end{cases}.$$

A 4.3.2. tétel szerint az $y = z = \dot{y} = \dot{z} = 0$ egyensúlyi állapot aszimptotikusan y-stabilis. Ugyanakkor, *A. S. Oziraner tétele* (4.1.* tétel) még akkor sem lenne alkalmazható, ha tudnánk, hogy a mozgások a sebességekkel együtt korlátosak, hiszen az $\{(y, z) : \pi(y, z) > 0\}$ halmazban van egyensúlyi helyzet!

A 4.3.4. következmény szerint az $f(y, z) = y^3/(1 + z)$ esetben az egyensúlyi állapot y-instabilis.

4.4. Kiegészítő megjegyzések

1979-ben, az első szegedi differenciálegyenletes konferencián L. SALVADORI hívta fel a figyelmet olyan feltételrendszerek kidolgozásának fontosságára, amelyek parciális aszimptotikus stabilitást, illetve instabilitást biztosítanak a nem-ellenőrzött koordináták korlátosságának megkövetelése nélkül. Az ő biztatására kezdett ilyen irányú kutatásaim eredményeit [23, 24, 28] dolgozataimban foglaltam össze. A jelen fejezet [23] alapján készült.

Hasonló típusú, funkcionálegyenletekre vonatkozó aszimptotikus vizsgálatok találhatók J. KATO megjelenés alatt álló [33] dolgozatában. A fejezetnek a mechanikai alkalmazások során nyert eredményeit TERJÉKI JÓZSEF [52] egészítette ki és fejlesztette tovább.

5. fejezet

A határegyenlet módszere

Az előző fejezetben autonóm rendszerekre alkalmas *Ljapunov-függvény* birtokában egy alternatívát bizonyítottunk (4.14. tétel), amely azt mondja, hogy a rendszer tetszőleges megoldására vagy a) az ellenőrzött koordináták 0-hoz tartanak, vagy b) a nem-ellenőrzött koordinátákból álló vektorfüggvény normában ∞ -hez tart, ha $t \to \infty$. Az a) eset kedvező az aszimptotikus stabilitás szempontjából, a b) esetet pedig úgy intéztük el, hogy a Ljapunov–függvény limeszére tettünk feltételt, miközben a nem-ellenőrzött koordináták vektora ∞–hez tart (l. 4.2.4. következmény); nevezetesen, feltettük, hogy ez a limesz ∞. A jelen fejezetben azzal az esettel foglalkozunk, amikor ez a limesz véges. Kapcsolódva ismét K. PEIFFER és N. ROUCHE [43] példájához, tegyük fel, hogy a gravitációs térben egy anyagi pont az $u = (1/2)y^2[1 + 1/(1 + z^2)]$ egyenlettel adott felület mentén surlódva mozog (az Oy, z, u inerciarendszer u–tengelye függőlegesen felfelé mutat). Az előző fejezet eredményeit használva nem lehet az $y = z = \dot{y} = \dot{z} = 0$ egyensúlyi állapot aszimptotikus y–stabilitását bizonyítani, hiszen a potenciális energia nem tart végtelenbe, ha $|z| \to \infty$. Ugyanakkor joggal várhatjuk el ezt a tulajdonságot, hiszen ha a mozgások korlátosak, akkor már a 4.1* tételből is adódik, ha viszont egy mozgás mentén z nem korlátos, akkor az alternatíva miatt $|z(t)| \to \infty$, és a mozgás „aszimptotikusan közel van” az $x = (1/2)y^2$ felület mentén történő mozgáshoz; amelynek során $y(t) \to 0$, ha $t \to \infty$.

Ebben a fejezetben azt fogjuk feltenni, hogy a rendszer jobboldalán az ellenőrzött koordináták változását szabályozó vektornak valamilyen értelemben létezik a határértéke, ha a nem-ellenőrzött koordináták vektora normában végtelenbe tart. Ez lehetővé teszi, hogy a T. YOSHIZAWA [55], G. SELL [50] és Z. ARTSTEIN [2–4] által kidolgozott határegyenlet-módszert feladatunkra alkalmazzuk.

Az itt kifejtendő módszer — szemben az előző fejezet tételeiben leírtakkal — nem-autonóm rendszerekre is alkalmazható [24,27]. A jelen értekezésben mi mégis csak az autonóm esetet tárgyaljuk, mivel az alapvető gondolatok ezen is bemutathatók, továbbá a nem-autonóm eset teljes kifejtéséhez komoly technikai apparátus bevezetésére lenne szükség [2–4, 27], ezért az általános esetre csak utalni fogunk.

5.1. Autonóm rendszerekre vonatkozó tételek

Tekintsük ismét az

$$\dot{x} = X(x) \quad (x \in R^k, \quad X(0) = 0) \tag{5.1}$$

egyenletrendszert, és ennek a szokásos $x = (y,z)^T$ partíciónak megfelelő alakját:

$$\dot{y} = Y(y,z), \quad \dot{z} = Z(y,z) \,. \tag{5.2}$$

Ebben a pontban végig feltesszük, hogy $Y(y,z) \to Y_*(y)$, ha $|z| \to \infty$, y-ra vonatkozóan egyenletesen a $\overline{B}_m(0,h')$ halmazon.

Szükségünk lesz a következő jelölésre. Adott $W : \Gamma'_m \to R$ függvényre és $c \in R$ számra jelölje $W_m^{-1}[c,\infty]_0$ azon $y \in R^m$ pontok halmazát, amelyekhez létezik olyan $\{(y_i,z_i,t_i)\}$ sorozat, hogy $y_i \to y$, $|z_i| \to \infty$, $t_i \to \infty$, $W(y_i,z_i,t_i) \to c$, és $\dot{W}(y_i,z_i,t_i) \to 0$, ha $i \to \infty$.

5.1.1. TÉTEL. *Tegyük fel, hogy az (5.2) rendszerhez létezik olyan $V : G'_m \to R_+$ Ljapunov-függvény, amely eleget tesz a következő feltételeknek:*

(1) V pozitív y-definit;

(2) bármely c pozitív számra a $\dot{V}^{-1}(0) \cap V^{-1}(c)$ halmaz nem tartalmazza az (5.2) rendszer egyetlen teljes trajektóriáját sem, és

(3) a $V_m^{-1}[c,\infty]_0$ halmaz nem tartalmazza az

$$\dot{y} = Y_*(y) \tag{5.3}$$

határegyenlet egyetlen teljes trajektóriáját sem, kivéve esetleg az R^m tér origóját.

Ekkor az (5.2) rendszer 0-megoldása aszimptotikusan y-stabilis.

Bizonyítás. Az (5.2) rendszer 0-megoldása y-stabilis (l. [46], 22. o.), tehát bármely $\varepsilon > 0$-hoz létezik olyan $\delta(\varepsilon) > 0$, hogy ha $|x_0| < \delta(\varepsilon)$, akkor $|y(t;x_0)| < \varepsilon$ minden $t \geq 0$-ra. Legyen $\varepsilon_0 > 0$ rögzített, és definiáljuk a $\sigma = \delta(\varepsilon_0) > 0$ számot. Be fogjuk bizonyítani, hogy bármely $x_0 \in B_k(0,\sigma)$ esetén $|y(t;x_0)| \to 0$, ha $t \to \infty$.

Legyen $x = \phi(t) = (\Psi(t), \chi(t))^T$ tetszőleges olyan megoldása az (5.2) egyenletnek, amelyre $\phi(0) \in B_k(0,\sigma)$ teljesül. A $v(t) = V(\phi(t))$ függvény csökkenő, tehát $v(t) \to v_0 \geq 0$. Ha $v_0 = 0$, akkor (1) miatt készen vagyunk. Tegyük fel, hogy $v_0 > 0$. Az előző fejezet 4.1.4. tétele szerint a (2) feltételből

$$\lim_{t\to\infty} |\chi(t)| = \infty \tag{5.4}$$

következik. Tekintsük az

$$\dot{y} = Y(y, \chi(t)) \quad (y \in B_m(0, h'), t \in R_+) \tag{5.5}$$

rendszert és azt az $U : B_m(0, h') \times R_+ \to R$ függvényt, amelyet az $U(y, t) = V(y, \chi(t))$ definiál. Nyilvánvalóan, $\dot{U}_{(5.5)}(y, t) = \dot{V}_{(5.2)}(y, \chi(t)) \le 0$, tehát U az (5.5) rendszerhez tartozó *Ljapunov-függvény*, és $u(t) = U(\Psi(t), t) \to v_0$, ha $t \to \infty$. Azt állítjuk, hogy az (5.5) egyenlet $y = \Psi(t)$ megoldásának $\Omega(\Psi)$ határhalmazára $\Omega(\Psi) \subset U_m^{-1}[v_0, \infty]_0$.

Valóban, a 2.1.2. lemma szerint $\Omega(\Psi) \subset U_m^{-1}[v_0, \infty]$. Tegyük fel, hogy ennek ellenére az előbb állított tartalmazás nem igaz. Mivel $U_m^{-1}[v_0, \infty]_0$ zárt halmaz a $B_m(0, h')$ gömbben, ekkor létezik olyan $q \in \Omega(\Psi)$ és $\rho > 0$, hogy $\overline{B}_m(q, \rho) \cap U_m^{-1}[v_0, \infty]_0 = \emptyset$ ahonnan a

$$\limsup_{T \to \infty} \{\dot{U}_{(5.5)}(\Psi(t), t) : t \ge T,\ \Psi(t) \in \overline{B}_m(q, \rho)\} < 0$$

egyenlőtlenség adódik. Figyelembe véve azt is, hogy az (5.5) egyenlet jobboldala korlátos a $\overline{B}_m(0, h') \times R_+$ halamazon, a 2.2.2. lemmát az (5.5) egyenletre a $\Lambda = \overline{B}_m(0, h')$, $V = U$, $W(y) = y$ választással alkalmazva kapjuk a $q \notin \Omega(\Psi)$ relációt, ami ellentmondás.

Másrészt, az (5.4) miatt $U_m^{-1}[v_0, \infty]_0 \subset V_m^{-1}[v_0, \infty]_0$, tehát az előbb bizonyított tartalmazást felhasználva

$$\Omega(\phi) = \Omega(\Psi) \subset V_m^{-1}[v_0, \infty]_0 \tag{5.6}$$

adódik.

Az $Y(y, z) \to Y_*(y)$ ($|z| \to \infty$) és az (5.4) tulajdonság együtt azt eredményezik, hogy az (5.3) egyenlet az (5.5) egyenletnek határegyenlete. Ismeretes, hogy a megoldások határhalmaza invariáns a határegyenletre vonatkozóan (l. [46], 178. o., 5.7. tétel), amiből már következik az $\Omega_m(\phi) = \{0\}$ egyenlőség, vagyis hogy $|\Psi(t)| \to 0$, ha $t \to \infty$, amit bizonyítani akartunk. Valóban, hiszen ha $\Omega_m(\phi)$, és így $\Omega(\Psi)$ tartalmaz az R^m-ből az origótól különböző pontot is, akkor az invariancia miatt tartalmazná a (3.3) egyenletnek egy olyan trajektóriáját is, amely az origótól különbözik. (5.6) miatt ez a trajektória a $V_m^{-1}[v_0, \infty]_0$ halmazban is benne lenne, ami viszont $v_0 > 0$ és a (3) feltétel miatt lehetetlen. □

Talán érdemes megjegyezni, hogy a tétel autonóm rendszerekről szól, bizonyításában az egyik lényeges lépés mégis a nem-autonóm egyenletek elméletéből egy lokalizációs tétel alkalmazása.

Az előző fejezetben volt szó arról, hogy — szemben C. RISITO [44] tételeivel — a (2) feltétel nem tesz különbséget ellenőrzött és nem-ellenőrzött koordináták között (l. 4.2.2* tétel). A határegyenlet módszere alkalmas arra is, hogy ezt a feltételt finomítsuk.

5.1.2. LEMMA. *Tegyük fel, hogy az (5.2) rendszerhez létezik olyan* $V : G'_m \to R_+$ *Ljapunov-függvény, amely eleget tesz a következő feltételeknek:*

(1) V pozitív y-definit;

(2) bármely c pozitív számra teljesül, hogy ha a $\dot{V}^{-1}_{(5.2)}(0) \cap V^{-1}(c)$ *halmaz tartalmazza az (5.2) rendszer egy teljes trajektóriáját, akkor ez a trajektória az* $\{(y,z) : y = 0\}$ *altérben helyezkedik el.*

Ekkor az (5.2) rendszer 0-megoldása y-stabilis, és ha $\phi = (\Psi, \chi)^T : [t_0, \infty) \to R^k$ *(5.2)-nek tetszőleges olyan megoldása, amelyben* Ψ *korlátos, akkor valahányszor* $t_i \to \infty$, $\Psi(t_i) \to q \neq 0$, $V(\phi(t_i)) \to v_0 > 0$, *mindannyiszor* $|\chi(t_i)| \to \infty$, *ha* $i \to \infty$.

Bizonyítás. Az y-stabilitás a *Rumjancev-tételből* következik ([46], 22. o.).

Ha az állítás második része nem igaz, akkor feltehető, hogy $\chi(t_i) \to r \in R^n$ $(i \to \infty)$. A 4.1.2. lemma szerint $\Omega(\phi) \subset M(v_0) := \dot{V}^{-1}_{(5.2)}(0) \cap V^{-1}(v_0)$. Mivel $\Omega(\phi)$ invariáns az (5.2) rendszerre vonatkozóan, az $M(v_0)$ halmaz tartalmazza az (5.2) (q, r) ponton áthaladó trajektóriáját, ami $q \neq 0$ miatt ellentmond a (2) feltételnek. □

5.1.3. TÉTEL. *Tegyük fel, hogy az 5.1.2. lemma (1)–(2) feltétele teljesül, továbbá*

(3) bármely c pozitív számra a $V_m^{-1}[c, \infty]_0$ *halmaz nem tartalmazza az (5.3) határegyenlet egyetlen teljes trajektóriáját sem, kivéve esetleg az* R^m *tér origóját;*

(4) létezik olyan $b \in \mathcal{K}$ *függvény, amellyel*

$$V(y,z) \leq b(|y|) \qquad ((y,z) \in \Gamma'_m).$$

Ekkor az (5.2) egyenlet 0-megoldása egyenletesen aszimptotikusan y-stabilis.

Bizonyítás. Az 5.1.2. lemma szerint a 0-megoldás y-stabilis. Tekintsünk egy $\phi(t) = (\Psi(t), \chi(t))^T$ megoldást, amelyre $\Psi(t)$ korlátos. Két eset lehetséges:

$$\text{a) } |\chi(t)| \to \infty \qquad \text{b) } |\chi(t)| \not\to \infty \quad (t \to \infty).$$

Bebizonyítjuk, hogy mindkét esetben $v_0 = \lim_{t\to\infty} V(\phi(t)) = 0$.

Az a) esetben $\Psi(t) \to 0$, ha $t \to \infty$, amint azt láttuk az 5.1.1. tétel bizonyításában, ebből viszont a (4) feltétel miatt $v_0 = 0$ már következik.

A b) esetben létezik olyan $\{t_i\}$ sorozat, hogy $t_i \to \infty$, $\Psi(t_i) \to q$, $\chi(t_i) \to r$, ha $i \to \infty$. Az 5.1.2. lemma szerint $q = 0$. A (4) feltétel értelmében

$$0 \leq v_0 = \lim_{t\to\infty} V(\phi(t)) \leq \lim_{t\to\infty} b(|\Psi(t_i)|) = 0.$$

Tehát mindkét esetben $v_0 = 0$, amiből az 1.6. megjegyzés alapján a tétel állítása következik. □

Az 5.1.3. tétel (4) feltétele sok fontos esetben nem teljesíthető (pl. $V(y,z) = y^2 + 1/(1+z^2)$), ezért most egy olyan tételt bizonyítunk, amelyben a (4) feltételt az (5.2) rendszer jobboldalának egy korlátossági tulajdonsága helyettesíti.

5.1.4. TÉTEL. *Tegyük fel, hogy az 5.3. tétel (1)–(3) feltételei teljesülnek, továbbá a Z függvény korlátos a G'_m halmazon. Akkor az (5.2) egyenlet 0–megoldása aszimptotikusan y–stabilis.*

Bizonyítás. A bizonyítást ugyanúgy lehet kezdeni, mint az 5.1.1. tételnél, a $v_0 > 0$ esetet kell viszont máképpen kezelni, hiszen az (5.4) tulajdonság most nincs a birtokunkban.

Tegyük fel, hogy $v_0 > 0$. Be fogjuk bizonyítani, hogy $\Omega_m(\phi) = \{0\}$. Ha ez nem igaz, akkor először azt mutatjuk meg, hogy $\Omega_m(\phi) \subset N(v_0) := V_m^{-1}[v_0, \infty]_0$, majd azt, hogy a határegyenletnek halad $\Omega_m(\phi)$–ban $\{0\}$–tól különböző trajektóriája, ami ellentmond a (3) feltételnek.

Az $\Omega_m(\phi)$ halmaz kompakt és összefüggő, továbbá $N(v_0)$ zárt, ezért elegendő az $\Omega_m(\phi) \setminus \{0\} \subset N(v_0)$ tartalmazást megmutatni. Tegyük fel, hogy ez nem igaz. Akkor létezik olyan $q \in \Omega_m(\phi)$ $(q \neq 0)$ és $\varepsilon > 0$, hogy $\overline{B}_m(q, 2\varepsilon) \cap [N(v_0) \cup \{0\}] = \emptyset$. Bebizonyítjuk, hogy

$$\alpha = \limsup_{T \to \infty} \{\dot{V}(\Psi(t), \chi(t)) : t \geq T,\ \Psi(t) \in \overline{B}_m(q, 2\varepsilon)\} < 0 \tag{5.7}$$

Ha ez nincs így, akkor létezik egy olyan $\{t_i\}$ sorozat, amelyre $t_i \to \infty$, $\dot{V}(\phi(t_i)) \to 0$, $\Psi(t_i) \to q' \in \overline{B}_m(q, 2\varepsilon)$ teljesül. Az 5.1.2. lemma szerint akkor azt is tudjuk, hogy $|\chi(t_i)| \to \infty$, ha $i \to \infty$, vagyis $q' \in N(v_0)$, ami ellentmond ε definíciójának. (5.7) tehát igaz.

V alulról korlátos, tehát (5.7) miatt $\Psi(t) \in \overline{B}_m(q, 2\varepsilon)$ nem teljesülhet egyetlen $[T, \infty)$ intervallumon sem. Ezért létezik olyan $\{t'_i, t''_i\}$ sorozat, hogy

$$t'_i < t''_i < t'_{i+1}, \quad t'_i \to \infty; \quad |\Psi(t'_i) - q| = \varepsilon, \quad |\Psi(t''_i) - q| = 2\varepsilon,$$

$$\varepsilon \leq |\Psi(t) - q| \leq 2\varepsilon \qquad (t'_i \leq t \leq t''_i,\ i = 1, 2, \dots).$$

De $Y(\Psi(t), \chi(t))$ korlátos, Ψ nem változhat akármilyen gyorsan, így $t''_i - t'_i \geq \beta > 0$ teljesül minden i–re egy alkalmas konstanssal. Ekkor

$$V(t''_i) - V(t'_1) \leq \sum_{j=1}^{i} \int_{t'_j}^{t''_j} \dot{V}(\phi(t))dt \leq i\alpha\beta \longrightarrow -\infty \qquad (i \to \infty)$$

ami ellentmondás. Ezzel az $\Omega_m(\phi) \subset N(v_0)$ tartalmazást bebizonyítottuk.

Tegyük fel, hogy $q \in \Omega_m(\phi)$, $q \neq 0$ és $t_i \to \infty$, $q(t_i) \to q$, ha $i \to \infty$. Tekintsük a $\{\Psi^i(t) := \Psi(t_i + t)\}$ függvénysorozatot. Ennek i-edik eleme megoldása az

$$\dot{y} = Y(y, \chi(t_i + t)), \qquad y(0) = \Psi(t_i)$$

kezdetiérték-problémának. Z korlátos a G'_m halmazon, és az 5.1.2. lemma miatt $|\chi(t_i)| \to \infty$, tehát bármely kompakt $[a, b]$ intervallumon $|\chi(t_i + t)| \to \infty$ egyenletesen, ha $i \to \infty$. Ezért $Y(y, \chi(t_i + t)) \to Y_*(y)$ az (y, t) változókra vonatkozóan egyenletesen a $\overline{B}_m(0,h') \times [a,b]$ halmazon. A *Kamke-lemma* következtében ([46], 175. o. 5.3. Lemma) a $\{\Psi_i(t)\}$ sorozat egy részsorozata egyenletesen konvergál az $\dot{y} = Y_*(y)$, $y(0) = q$ kezdetiérték-probléma egy γ megoldásához az $[a,b]$ intervallumon, ha $i \to \infty$. Tehát bármely rögzített t-re $\gamma(t)$ a $\{\Psi(t+t_i)\}$ egy részsorozatának limesze, továbbá $t_i + t \to \infty$, így $\gamma(t) \in \Omega_m(\phi)$, vagyis $\Omega_m(\phi)$ tartalmazza az (5.3) határegyenlet egy teljes trajektóriáját, amely nem esik egybe az $\{0\}$ halmazzal. A korábban bizonyított $\Omega_m(\phi) \subset N(v_0)$ tartalmazás miatt ez a trajektória $N(v_0)$ -ból sem lép ki, ami ellentmond a (3) feltételnek. □

A határegyenlet módszere instabilitási tételek levezetésére is alkalmas.

5.1.5. TÉTEL. *Tegyük fel, hogy az (5.2) egyenlethez létezik olyan $V \in Lip(G'_m; R)$ függvény, amely rendelkezik a következő tulajdonságokkal:*

(1) V alulról korlátos;

(2) bármely pozitív δ -hoz létezik olyan $x_0 \in B_k(0,\delta)$, hogy $V(x_0) < 0$;

(3) bármely c negatív számra a $\dot{V}^{-1}_{(5.2)}(0) \cap V^{-1}(c)$ halmaz nem tartalmazza az (5.2) egyenlet egyetlen teljes trajektóriáját sem, és

(4) a $V_m^{-1}[c,\infty]_0$ halmaz nem tartalmazza az (5.3) határegyenlet egyetlen teljes trajektóriáját sem.

Akkor az (5.2) egyenlet 0–megoldása y–instabilis.

Bizonyítás. Legyenek ε_0, δ adottak ($0 < \varepsilon_0 < h'$, $0 < \delta < \varepsilon_0$) és vegyünk egy olyan $x_0 \in B_k(0,\delta)$ pontot, amelyre $V(x_0) < 0$. Tekintsük az (5.2) egyenlet $\phi(0) = x_0$ feltételnek eleget tevő $x = \phi(t) = (\Psi(t), \chi(t))^T$ megoldását. Bebizonyítjuk, hogy $|(\Psi(T)| > \varepsilon_0$ valamely $T > 0$ számra.

Tegyük fel, hogy ez utóbbi állítás nem igaz, vagyis $|\Psi(t)| \leq \varepsilon_0$ minden $t \geq 0$ értékre. Az előző fejezet 4.1.2. lemmája szerint ekkor a (3) feltételből $|\chi(t)| \to \infty$ $(t \to \infty)$ következik. Mint ahogyan az 5.1.1. tétel bizonyításában láttuk, ekkor a nem-üres $\Omega_m(\phi)$ halmaz, amely invariáns az (5.3) határegyenletre nézve, részhalmaza a $V_m^{-1}[v_0,\infty]_0$ halmaznak. Ez azt jelenti, hogy ezen utóbbi halmaz tartalmazza a határegyenlet teljes trajektóriáját, ami ellentmondásban van a (4) feltétellel. □

5.1.6. *Megjegyzés.* Legyen $y = (y_1, y_2)^T$ az $y \in R^m$ vektor egy partíciója ($y_1 \in R^{m_1}$, $y_2 \in R^{m_2}$, $1 \leq m_1 < m$, $m_1 + m_2 = m$), és tegyük fel, hogy valamely $\varepsilon_0 > 0$ számra teljesül a következő: ha $|y_1| \leq \varepsilon_0$ és $V(y_1, y_2, z) < 0$, akkor $|y_2| \leq h'$. Az 5.1.5. tétel bizonyításából könnyen látszik, hogy ebben az esetben a tétel feltételeiből y_1–instabilitás is következik.

5.2. Holonóm és anholonóm mechanikai rendszerek

Tekintsük ismét a 4. fejezetben már vizsgált

$$\frac{d}{dt}\frac{\partial T}{\partial \dot{q}} - \frac{\partial T}{\partial q} = -\frac{\partial \pi}{\partial q} + Q \quad (q, \dot{q} \in R^r) \tag{5.8}$$

mechanikai rendszert és az általánosított koordináták $q = (q_1, q_2)^T$ partícióját ($q_1 \in R^{r_1}$, $q_2 \in R^{r_2}$, $1 \le r_1 \le r$, $r_1 + r_2 = r$). A jelen fejezet eredményeit most az aszimptotikus $(q_1, \dot q)^T$-stabilitás és a q_1-instabilitás feltételeinek tanulmányozására használjuk abban az esetben, amikor az (5.8) rendszer aszimptotikusan q_2-független.

5.2.1. *Definíció* [24]. Az (5.8) rendszert *aszimptotikusan q_2-független*nek nevezzük, ha valamely $h' > 0$ konstanssal tetszőleges $K \subset R^r$ kompakt halmazra teljesülnek a következő feltételek:

(a) létezik olyan $\lambda > 0$ és $c \in \mathcal{K}$, hogy

$$\lambda |\dot q|^2 \le \frac{1}{2}\dot q^T A(q_1, q_2)\dot q, \qquad Q^T(q_1, q_2, \dot q)\dot q \le -c(|\dot q|),$$

feltéve,hogy $q_1 \in \overline{B}_{r_1}(0, h')$, $q_2 \in R^{r_2}$, $\dot q \in R^r$;

(b) A $(q_1, q_2) \to A_*(q_1)$, $\pi(q_1, q_2) \to \Pi_*(q_1)$, ha $|q_2| \to \infty$, továbbá $Q(q_1, q_2, \dot q) \to Q_*(q_1, \dot q)$ a $\overline{B}_{r_1}(0, h') \times K$ halmazon egyenletesen, ha $|q_2| \to \infty$, és $\partial A/\partial q$, $\partial \pi/\partial q$ egyenletesen konvergál a $\overline{B}_{r_1}(0, h')$ halmazon, ha $|q_2| \to \infty$.

Illusztrálásképpen tekintsük a jelen fejezet bevezetésében már említett példát: m tömegű anyagi pont gravitációs térben mozog az $u = (1/2)y^2[1 + 1/(1 + z^2)]$ egyenlettel adott felület mentén, a pontra $Q = -(\alpha\dot y, \alpha\dot z, \alpha\dot u)$ $(\alpha > 0)$ surlódási erő is hat. A potenciális és kinetikai energia, mint az y, z általánosított koordináták és az $\dot y, \dot z$ általánosított sebességek függvényei:

$$T = \frac{1}{2}m\left\{\dot y^2 + \dot z^2 + \left(y\dot y\left(1 + \frac{1}{1+z^2}\right) - y^2\frac{z\dot z}{1+z^2}\right)^2\right\}$$

$$\pi = \frac{mg}{2}y^2\left(1 + \frac{1}{1+z^2}\right)$$

Könnyű ellenőrízni, hogy ez a mechanikai rendszer aszimptotikusan z-független, és a megfelelő határfüggvények:

$$T_*(y, \dot y, \dot z) = \frac{1}{2}m\{(1 + y^2)\dot y^2 + \dot z^2\}$$

$$\pi_*(y) = \frac{1}{2}mgy^2, \qquad Q_*(y, \dot y, \dot z) = -(\alpha\dot y, \alpha\dot z, \alpha y\dot y)$$

Térjünk vissza az (5.8) mechanikai rendszerhez, és a kívánt feltételek meghatározására próbáljuk alkalmazni az 5.1.4. és 5.1.5. tételt. Bevezetve a q, $p = A(q)\dot q$ *Hamilton-féle változókat*, az (5.8) rendszer a

$$\text{(5.9)} \qquad \begin{cases} \dot p = -\dfrac{1}{2}p^T\left(\dfrac{\partial A^{-1}(q)}{\partial q}\right)p - \dfrac{\partial \pi}{\partial q} + Q(q, A^{-1}(q)p), \\ \dot q = A^{-1}(q)p \end{cases}$$

kanonikus alakot ölti. Ha (5.8) aszimptotikusan q_2–független, akkor (5.8) $q = \dot{q} = 0$ egyensúlyi helyzetének és (5.9) $q = p = 0$ megoldásának ugyanazok a stabilitási tulajdonságai.

Határozzuk meg az alkalmazni kívánt tételben előforduló függvényeket, halmazokat. Válasszuk segédfüggvénynek a $H = T + \pi$ teljes mechanikai energiát. Ha a disszipáció teljes, akkor létezik olyan $d \in \mathcal{K}$ függvény, hogy

$$\dot{H}_{(5.9)}(p, q_1, q_2) = Q^T(q, A^{-1}(q)p)A^{-1}(q)p \leq -d(|p|) \tag{5.10}$$

teljesül, ha $(p, q_1) \in \overline{B}_{r_1+r}(0, h')$, $q_2 \in R^{r_2}$. Tehát H *Ljapunov-függvény* az (5.9) rendszerhez, és

$$\dot{H}^{-1}_{(5.9)}(0) \cap H^{-1}(c) = \{(p, q) : \pi(q) = c,\ p = 0\} \quad (c \in R), \tag{5.11}$$

ezért az (5.9) rendszernek ezen halmazban lévő teljes trajektóriái pontosan azok a $p = 0$, $q = q_0$ egyensúlyi állapotai, amelyekre $\pi(q_0) = c$. Nyilvánvalóan,

$$H^{-1}_{r_1+r}[c, \infty]_0 \subset H^{-1}_{r_1+r}[c, \infty] \cap d^{-1}(0) =:$$
$$= \{(p, q_1) : p = 0,\ q_1 = \pi^{-1}_{r_1}[c, \infty]\} = E(c)$$

Mivel $\partial\pi/\partial q_1$ folytonos és egyenletesen konvergens, ha $|q_2| \to \infty$, a $\pi(\cdot, q_2) : \overline{B}_{r_1}(0, h') \to R$ függvények folytonosak q_2–re vonatkozóan egyenletesen ($q_2 \in R^{r_2}$), így

$$E(c) = \{(p, q_1) : p = 0,\ \pi_*(q_1) = c\}. \tag{5.12}$$

Mivel az (5.9) rendszer aszimptotikusan q_2–független, határegyenletrendszere a $|q_2| \to \infty$ határátmenet mellett a következő:

$$\begin{cases} \dot{p}^i = -\dfrac{1}{2}p^T\left(A_*^{-1}\dfrac{\partial A_*}{\partial q^i}A_*^{-1}\right)p - \dfrac{\partial \pi_*}{\partial q^i} + Q_*^i(q_1, A_*^{-1}p) \\ \dot{p}^j = Q_*^j(q_1, A_*^{-1}p) \\ \dot{q}^i = \displaystyle\sum_{k=1}^{r}[A_*^{-1}(q_1)]_{ik}p^k \end{cases} \tag{5.13}$$

($i = 1, 2, \ldots r_1$, $j = 1, 2, \ldots, r$). Ha az $E(c)$ halmaz tartalmazza a (4.6) rendszer egy teljes trajektóriáját, akkor ez a trajektória (5.12) szerint egy $p = 0$, $q_1 = (q_1)_0 =$konst. pont, amelyre

$$\pi_*((q_1)_0) = c, \qquad \left.\frac{d\pi_*}{dq_1}\right|_{q_1=(q_1)_0} = 0$$

is teljesülnek.

Ezek után könnyű ellenőrízni, hogy az 5.1.4. és 5.1.5. tételből — az 5.1.6. megjegyzést is figyelembe véve — az alábbi tételt kapjuk.

5.2.2. TÉTEL. *Tegyük fel, hogy az (5.8) rendszer aszimptotikusan q_2–független.*

I. Ha a π potenciális energia pozitív q_1–definit, az (5.8) rendszernek nincs egyensúlyi helyzete a $\{(q_1, q_2)^T : \pi(q_1, q_2) > 0,\ q_1 \neq 0\}$ tartományban, és a $d\pi_*(q_1)/dq_1 = 0$ egyenlőségből vagy $q_1 = 0$, vagy $\pi_*(q_1) = 0$ következik, akkor az (5.8) $q = \dot{q} = 0$ egyensúlyi állapota aszimptotikusan $(q_1, \dot{q})^T$–stabilis.

II. Ha a π potenciális energiának a $q = 0$ pontban nincs helyi minimuma, az (5.8) rendszernek nincs egyensúlyi helyzete a $\{q : \pi(q) < 0\}$ tartományban, és a $d\pi_*(q_1)/dq_1 = 0$ egyenlőségből $\pi_*(q_1) \geq 0$ következik, akkor az (5.8) rendszer $q = \dot{q} = 0$ egyensúlyi állapota q_1-instabilis.

Ebből a tételből azonnal következik, hogy az $u = (1/2)y^2[1 + 1/(1+z^2)]$ felület mentén történő mozgásnál viszkózus surlódás mellett az $y = z = \dot{y} = \dot{z} = 0$ egyensúlyi állapot aszimptotikusan $(y, \dot{y}, \dot{z})$–stabilis. Ugyanakkor, az $u = y^3[1 + 1/(1+z^2)]$ felület mentén való mozgásnál ugyanaz az egyensúlyi állapot y–instabilis.

Tekintsünk most egy *anholonóm* [7,12] mechanikai rendszert. Ez azt jelenti, hogy a rendszerre nemcsak geometriai, hanem kinematikai kényszerek is hatnak, azaz a konfigurációs tér egy-egy adott pontjában a sebességvektor nem lehet tetszőleges, hanem egy kijelölt lineáris altér egy eleme. Ez annyit jelent, hogy adott egy $B : R^r \to R^{r \times r}$ mátrixfüggvény, és a mozgásokra az (5.8) egyenleten kívül a

$$\dot{q}_2 = B(q_1, q_2)\dot{q}_1 \qquad (q_1 \in R^{r_1},\ q_2 \in R^{r_2},\ r_1 + r_2 = r) \tag{5.15}$$

feltételnek is teljesülni kell minden időpillanatban. Ilyen feltétel formáját ölti például az a kikötés, hogy egy golyó vagy korong adott felület mentén csúszás nélkül gördüljön (a felülettel érintkező pontnak a sebessége 0). Ez számunkra azért érdekes, mert itt eleve adott az általánosított koordinátáknak egy természetes partíciója.

Ha az (5.15) összefüggés segítségével az (5.8) egyenletből a $\dot{q}_2$ sebességet kiküszöböljük, akkor egy

$$\begin{cases} \dfrac{d}{dt}\dfrac{\partial\Theta}{\partial\dot{q}_1} = \dfrac{\partial(\Theta - \pi)}{\partial q_1} + \dfrac{\partial(\Theta - \pi)}{\partial q_2}B + \widehat{Q} + \dot{q}_1^T G \\ \dot{q}_2 = B(q)\dot{q}_1 \end{cases} \tag{5.16}$$

alakú rendszert kapunk, ahol $G = G(q, \dot{q}_1)$ ferdén szimmetrikus mátrixfüggvény: $G^T = -G$, vagyis giroszkópikus hatás lép fel; $\Theta = (1/2)\dot{q}_1^T\tilde{A}(q)\dot{q}_1$ – kvadratikus alak; $\widehat{Q} = \widehat{Q}(q, \dot{q}_1)$ egy R^{r_1}-vektorfüggvény (ez a *Voronec-egyenlet* ([71], 109. o.)).

Az ilyen anholonóm rendszer egyensúlyi helyzetének parciális stabilitási tulajdonságait V. V. RUMJANCEV [75] és C. RISITO [44] kezdték vizsgálni abban a speciális esetben, amikor $Q_1 = Q_1(q, \dot{q}_1)$ (*Rumjancevnél* $Q_1(q, \dot{q}_1) = D\dot{q}_1$, ahol D $r_1 \times r_1$-es konstans mátrix), a $Q_2 = 0$, és a disszipáció teljes, vagyis $Q_1^T\dot{q}_1 = 0$ akkor és csakis akkor, ha $\dot{q}_1 = 0$.

5.2.2* TÉTEL (V. V. RUMJANCEV [75]). *Tegyük fel, hogy minden $K \subset R^{r_1}$ kompakt halmazra*

(1) $B(q_1, q_2)$ korlátos a $K \times R^{r_2}$ halmazon;

(2) $\partial\pi/\partial q_1 + (\partial\pi/\partial q_2)B$ pozitív q_1–definit a $K \times R^{r_2}$ halmazon;

(3) π másodfokú Taylor–polinómja a $q = 0$ körül pozitív q_1–definit.

Akkor a $q = \dot{q} = 0$ egyensúlyi állapot aszimptotikusan $(q_1, \dot{q})^T$–stabilis.

RISITO–nak sikerült a (2) és (3) feltételt lényegesen gyengíteni, de csak az első feltétel szigorításával:

5.2.3* TÉTEL (C. RISITO) [44]). *Tegyük fel, hogy*

(1′) az (5.16) rendszernek a $q = 0$, $\dot{q} = 0$ kis környezetéből induló megoldásai egyenletesen q_2–korlátosak R_+-on;

(2′) az (5.16) rendszer egyensúlyi helyzetei pontosan a $q_1 = 0$, $q_2 = c =$ konst. pontok;

(3′) π pozitív q_1–definit minden kompakt $K \subset R^r$ halmazon.

Akkor a $q = 0$, $\dot{q} = 0$ egyensúlyi állapota aszimptotikusan $(q_1, \dot{q})^T$–stabilis.

Az első feltétel túlzottan szigorú, mivel a megoldások ismerete nélkül nehezen ellenőrízhető. Felmerül a kérdés: hogyan lehetne a két tétel előnyeit egyesítő eredményt kapni? Nevezetesen, milyen feltételekkel kell kiegészíteni az (1), (2′), (3′) feltételrendszert úgy, hogy a kapott feltételrendszer elegendő legyen az aszimptotikus $(q_1, \dot{q})^T$–stabilitáshoz? Erre a kérdésre választ lehet kapni az előző és a jelen fejezet eredményeinek a holonóm esethez lényegében hasonló módon való alkalmazásával (l. [27]).

5.3. Kiegészítő megjegyzések

Az eredmények [24,27]–ből valók, ahol nem-autonóm esetre vonatkozó tételek is találhatók. Ezekben az első lépés olyan feltételek kidolgozása, amelyek biztosítják az $\dot{x} = X(x,t)$ és az $\dot{y} = Y(y, \chi(t), t)$ egyenletek határegyenleteinek a létezését. Ezután az 5.1.1. tétel feltételein — durván szólva — az alábbi módosításokat kell végrehajtani:

— a (2) feltétel helyett azt követeljük meg, hogy tetszőleges pozitív c–re a $V_k^{-1}[c,\infty]_0$ halmaz ne tartalmazza az $\dot{x} = X(x,t)$ egyenlet egyetlen határegyenletének egyetlen pozitív fél-trajektóriáját sem;

— tetszőleges pozitív c–re és olyan $\chi : R_+ \to R^n$ függvényre, amelyre $|\chi(t)| \to \infty$ $(t \to \infty)$, a $V_m^{-1}[c,\infty]_0$ halmaz ne tartalmazza az $\dot{y} = Y(y, \chi(t), t)$ egyenlet egyetlen határegyenletének egyetlen pozitív féltrajektóriáját sem a $\{0\}$ kivételével.

Az így kapott állításoknak az 5.1.3.–5.1.4. tételekkel analóg finomítását a [27] dolgozat tartalmazza.

6. fejezet

Energia-típusú Ljapunov-függvények

A. M. LJAPUNOV–nak az aszimptotikus stabilitásról szóló klasszikus tételében ([65], l. Bevezetés) a *Ljapunov-függvényről* az alábbi két feltételt találjuk: alkalmas $a, b, c \in \mathcal{K}$ függvényekkel teljesülnek az

(1) $a(|x|) \leq V(x,t) \leq b(|x|)$;

(2) $\dot{V}(x,t) \leq -c(|x|)$

egyenlőtlenségek az $U \times R_+$ halmazon, ahol $U \subset R^k$ az origó egy környezete. Az a, b, c függvények szigorú monotonitása miatt a (2) feltétel ekvivalens egy $\dot{V}(x,t) \leq -d(V(x,t))$ egyenlőtlenség teljesülésével, ahol $d \in \mathcal{K}$. A mechanikában — mint az eddigi tapasztalatokból kiderült — a legtipikusabb *Ljapunov-függvénynek*, a $V(q,\dot{q}) = T(q,\dot{q}) + \pi(q) = (1/2)\dot{q}^T A(q)\dot{q} + \pi(q)$ teljes mechanikai energiának a mozgások mentén vett deriváltja a q pozíciós koordinátákra vonatkozóan nem negatív definit, tehát csak egy $\dot{V}(q,\dot{q}) \leq -c(|\dot{q}|)$, vagyis egy $\dot{V}(q,\dot{q}) \leq -d(T(q,\dot{q}))$ egyenlőtlenség teljesülése várható. Ez adja az ötletet, hogy vizsgáljuk a következő kérdést: Tegyük fel, hogy a *Ljapunov-függvény* két segédfüggvény összege, $V = V_1 + V_2$, és a rendszerre vonatkozó derivált egy $\dot{V}(x,t) \leq -\phi(t)c(V_1(x,t))$ egyenlőtlenségnek tesz eleget. Milyen feltételek mellett tudjuk ezt a *Ljapunov-függvényt* a megoldások aszimptotikus viselkedésének tanulmányozására használni? A kérdést megválaszoló eredményünk alkalmas mechanikai rendszerekben a sebességekre vonatkozó aszimptotikus stabilitás feltételeinek tanulmányozására.

6.1. Általános rendszerekre vonatkozó tételek

6.1.1. TÉTEL. *Tegyük fel, hogy a*

$$\dot{x} = X(x,t) \tag{6.1}$$

differenciálegyenletrendszerhez léteznek olyan $V_1, V_2 \in \mathrm{Lip}(\Gamma_m; R)$ *függvények, amelyek kielégítik a következő feltételeket a* Γ_m *halmazon:*

(1) $V(x,t) := V_1(x,t) + V_2(x,t) \geq 0$;

(2) $V_1(x,t) \geq 0$;

(3) létezik olyan $c \in \mathcal{K}$ *és* $\alpha : R_+ \to R_+$ *integrálisan pozitív függvény, hogy*

$$\dot{V}(x,t) \leq -\alpha(t)c(V_1(x,t));$$

(4) tetszőleges $\gamma, \gamma_1 > 0$ *számokra és minden olyan* $\xi : R_+ \to R^k$ *folytonos függvényre, amely eleget tesz a* $V(\xi(t),t) \leq \gamma$ $(t \in R_+)$ *feltételnek, az*

$$\int_{H(t;\gamma_1)} [\dot{V}_2(\xi(s),s]_+ ds$$

függvény egyenletesen folytonos az R_+ *félegyenesen, ahol*

$$H(t;\gamma_1) := \{s : 0 \leq s < t,\ V_1(s,\xi(s)) \geq \gamma_1\}.$$

Ekkor a (6.1) egyenlet minden akármeddig folytatható $x(t)$ *megoldására* $V_1(x(t),t) \to 0$, *és* $V_2(x(t),t)$*-nek létezik határértéke, ha* $t \to \infty$.

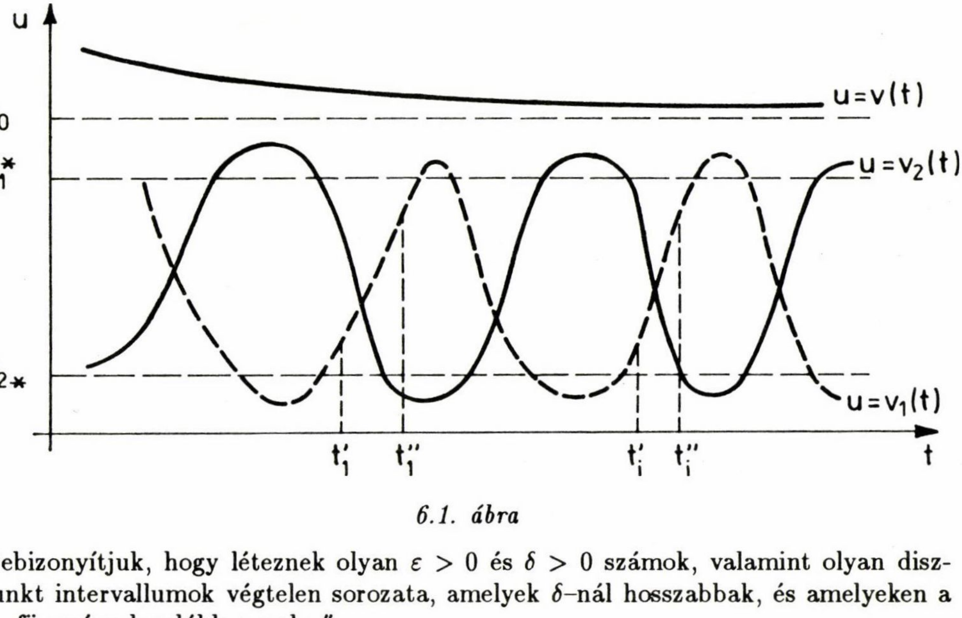
Bizonyítás. Legyen $x(t)$ megoldása a (6.1) egyenletnek a $[t_0,\infty)$ intervallumon, és tekintsük a $v_1(t) := V_1(x(t),t)$, $v_2(t) := V_2(x(t),t)$, $v(t) := v_1(t) + v_2(t)$

függvényeket. Jelölje $v_0 \geq 0$ a v függvény határértékét, ha $t \to \infty$. Azt kell megmutatnunk, hogy $v_1(t) \to 0$, ha $t \to \infty$.

Tegyük fel, hogy ez nem igaz. Akkor az (1)–(3) feltételek teljesülése miatt

$$\liminf_{t\to\infty} v_1(t) < \limsup_{t\to\infty} v_1(t) =: v_1^* \leq \infty,$$

$$v_{2*} := \liminf_{t\to\infty} v_2(t) = v_0 - v_1^* < v_0 = \limsup_{t\to\infty} v_2(t).$$



6.1. ábra

Bebizonyítjuk, hogy léteznek olyan $\varepsilon > 0$ és $\delta > 0$ számok, valamint olyan diszjunkt intervallumok végtelen sorozata, amelyek δ–nál hosszabbak, és amelyeken a v_2 függvény legalább ε–nal nő.

Valóban, legyen $\varepsilon := v_1^*/4 > 0$, ha $v_1^* < \infty$, és legyen $\varepsilon > 0$ tetszőleges, ha $v_1^* = \infty$. Létezik olyan T, hogy $v_0 \leq v(t) < v_0 + \varepsilon$, ha $t \geq T$, továbbá létezik olyan $T < t_1' < t_1'' < \cdots < t_i' < t_i'' < \ldots$, hogy

$$v_1(t_i') = 3\varepsilon,\ v_1(t_i'') = \varepsilon,\ \varepsilon \leq v_1(t) \leq 3(\varepsilon) \quad (t \in [t_i', t_i''],\ i = 1, 2, \ldots).$$

Felhasználva a $v_2(t) \equiv v(t) - v_1(t)$ azonosságot, a

$$v_2(t_i') \leq v_0 - 2\varepsilon,\ v_2(t_i'') \geq v_0 - \varepsilon \quad (i = 1, 2, \ldots)$$

egyenlőtlenségeket kapjuk, amelyekből a

$$0 < \varepsilon \leq v_2(t_i'') - v_2(t_i') \leq \int_{t_i'}^{t_i''} [\dot{V}_2(x(t), t)]_+ dt \quad (i = 1, 2, \ldots)$$

becslés adódik. A (4) feltétel miatt ebből következik, hogy $t_i'' - t_i' \geq \delta > 0$ $(i = 1, 2, \dots)$ teljesül valamely δ–val $(\gamma := v_0 + \varepsilon,\ \gamma_1 := \varepsilon)$.

Másrészt, a (3) feltétel szerint ekkor $v(t) \to -\infty$, ha $t \to \infty$, ami ellentmond annak, hogy $v(t)$ nem–negatív. □

Ha a V_1 függvény pozitív y–definit, akkor a tétel feltételeiből következik, hogy minden olyan $x(t) = (y(t), z(t))^T$ megoldásra, amely bármeddig folytatható, $|y(t)| \to 0$ $(t \to \infty)$ teljesül. Ha még a $V = V_1 + V_2$ pozitív y–definitását is biztosítjuk, akkor egy olyan tételt kapunk, amely aszimptotikus y–stabilitást biztosít, és ugyanakkor — ez rendkívül fontos! — a $V = V_1 + V_2$ *Ljapunov–függvényre* semmilyen felső becslésnek („végtelen kicsiny felső korlát") nem kell teljesülni:

6.1.2. KÖVETKEZMÉNY. *Tegyük fel, hogy a (6.1) egyenlethez létezik két olyan $V_1, V_2 \in \mathrm{Lip}(\Gamma_m'; R)$ függvény, amely kielégíti a következő feltételeket a Γ_m' halmazon:*

(1) $V_1(0, t) \equiv V_2(0, t) \equiv 0,\ V_2(x, t) \geq 0$;

(2) létezik olyan $a_1 \in \mathcal{K}$ függvény, hogy $a_1(|y|) \leq V_1(y, z, t)$;

(3)—(4) teljesül a 6.1.1. tétel (3)—(4) feltétele.

Ekkor a (6.1) egyenlet 0–megoldása aszimptotikusan y–stabilis, és minden elég kicsiny kezdeti értékekkel induló $x(t)$ megoldásra a $V_2(x(t), t)$ függvénynek létezik véges határértéke, ha $t \to \infty$.

Legyen $x = (y', z')^T$ az $x \in R^k$ vektornak egy újabb partíciója: $y' \in R^{m'}$, $z' \in R^{n'}$, $m \leq m' \leq k$, $n' = k - m'$. Ha $V_1(y', z', t) \to 0$ $(y' \to 0)$ a $z' \in R^{n'}$, $t \in R_+$ változókra vonatkozóan egyenletesen, akkor a (3) feltétel igen egyszerű alakot ölt.

6.1.3. KÖVETKEZMÉNY. *Tegyük fel, hogy a (6.1) egyenlethez létezik két olyan $V_1, V_2 \in \mathrm{Lip}(\Gamma_m'; R)$ függvény, amely kielégíti a következő feltételeket a Γ_m' halmazon:*

(1) $V_1(0, t) \equiv V_2(0, t) \equiv 0,\ V_2(x, t) \geq 0$;

(2) léteznek olyan $a_1, b_1 \in \mathcal{K}$ függvények, hogy $a_1(|y|) \leq V_1(x, t) \leq b_1(|y'|)$;

(3) létezik olyan $c \in \mathcal{K}$ és olyan integrálisan pozitív $\alpha : R_+ \to R_+$ függvény, hogy

$$\dot{V}(x, t) \leq -\alpha(t) c(|y'|);$$

(4) bármely $\gamma, \mu > 0$ számokra az

$$\int_0^t \left[\sup\left\{\dot{V}_2(y', z', s) : (y', z') \in M_{\gamma,\mu}(s)\right\}\right]_+ ds$$

függvény egyenletesen folytonos az R_+ félegyenesen, ahol

$$M_{\gamma,\mu}(s) := \left\{(y', z') \in R^{m'} \times R^{n'} : V(y', z', s) \leq \gamma,\ |y'| \geq \mu\right\}.$$

Ekkor a 6.1.2. következmény állítása igaz.

6.2 Az egyensúlyi helyzet sebességekre vonatkozó aszimptotikus stabilitása

Tekintsük ismét a 3. fejezetben már tárgyalt disszipatív giroszkópikus mechanikai rendszert:

$$\frac{d}{dt}\frac{\partial T}{\partial \dot{q}} - \frac{\partial T}{\partial q} = -\frac{\partial \pi}{\partial q} - B\dot{q} + G\dot{q}, \tag{6.2}$$

(a jelölésekre vonatkozóan l. a (3.5) rendszert). Feltéve, hogy $\partial\pi(q)/\partial q_{|q=0} = 0$, vizsgáljuk a $q = \dot{q} = 0$ egyensúlyi állapot $\dot{q}$ általánosított sebességvektorra vonatkozó aszimptotikus stabilitásának a feltételeit.

A $B(t,q) = B(q)$, $G(t,q) = G(q)$ stacionárius esetre V. V. RUMJANCEV [76] bebizonyította, hogy ha a disszipáció teljes, és potenciális erő nem hat ($\pi(q) \equiv 0$), akkor az egyensúlyi állapot aszimptotikusan stabilis a sebességekre vonatkozóan. Ebben az állításban az első feltétel természetes, a második viszont nem. Hiszen, például, ha csak a gravitációs térben adott felület mentén mozgó anyagi pontot vizsgáljuk, tapasztalataink szerint teljes disszipációnál nem csak akkor áll be a sebességekre vonatkozó aszimptotikus stabilitás, amikor az adott felület egy vízszintes sík, hanem, például akkor is, ha a felület csésze alakú (hat potenciális erő, de az egyensúlyi állapot stabilis). RUMJANCEV-nek a $\pi(q) \equiv 0$ feltételt azért kellett megkövetelnie, mert LJAPUNOV klasszikus tételének a parciális y–stabilitásra vonatkozó általánosításában a $V(y,z)$ *Ljapunov-függvényre* egy $V(y,z) \leq b(|y|)$ ($b \in \mathcal{K}$) felső becslésnek is teljesülni kell, ami esetünkben a $H(q,\dot{q})$ teljes mechanikai energiára vonatkozóan egy $H(q,\dot{q}) = T(q,\dot{q}) + \pi(q) \leq b(|\dot{q}|)$ egyenlőtlenség teljesülését jelenti, ami maga után vonja a $\pi(q) \equiv 0$–t. Ez is mutatja, mennyire fontos olyan feltételeket keresni az aszimptotikus stabilitásra, amelyek a V *Ljapunov-függvényt* felülről nem korlátozzák. Ha a rendszer autonóm (vagyis a mechanikai rendszer stacionárius), és a megoldások korlátosak, akkor az invariancia–elv segítségével nyerhetők ilyen feltételek (l. 4. fejezet). Nem–autonóm rendszerekre azonban az invariancia–elv nem igaz, márpedig igen fontos lenne, például, a sebességekre vonatkozóan aszimptotikus stabilitási feltételeket kapni arra az esetre is, amikor a fékezés intenzitása, vagyis a $B(q,t)$ együttható–mátrix az időtől is függ.

Az ilyen irányú korábbi tételeket közösen jellemzi, hogy megkövetelik a $B(q,t)$ mátrixban szereplő surlódási együtthatók időbeni korlátosságát ([21,69,73]). Ez szintén egy természetellenes technikai feltétel, hiszen tapasztalataink szerint minél nagyobb a surlódás, annál inkább számíthatunk az egyensúlyi helyzet aszimptotikus stabilitására a sebességekre vonatkozóan. A 6.1.1. tétel felhasználásával olyan eredményt kapunk, amely a $\| B(q,t) \|$, $\| G(q,t) \|$ függvényeket felülről egyáltalán nem korlátozza.

Jelölje $\lambda(q)$, illetve $\Lambda(q)$ a kinetikai energia $A(q)$ szimmetrikus mátrixának legkisebb, illetve legnagyobb sajátértékét, $\beta(q,t)$ pedig a surlódási együtthatók $B(q,t)$ mátrixának legkisebb sajátértékét. Tetszőleges $M > 0$ számra E_M jelöli a potenciális energia $\{q \in R^r : \pi(q) < M\}$ nívóhalmazának az origót tartalmazó komponensét.

6.2.1. TÉTEL. *Tegyük fel, hogy létezik olyan $M > 0$ szám, amellyel teljesülnek a következő feltételek:*

(i) a disszipáció „integrálisan teljes", vagyis az

$$\alpha_M(t) := \inf\{\beta(q,t)/\Lambda(q) : q \in E_M\}$$

függvény integrálisan pozitív;

(ii) $\lambda_M := \inf\{\lambda(q) : q \in E_M\} > 0$;

(iii) $\pi(q) \geq 0$, *és* $\mathrm{grad}\pi(q)$ *korlátos az* E_M *halmazon.*

Ekkor a (6.2) rendszer $q = \dot{q} = 0$ egyensúlyi állapota aszimptotikusan stabilis a sebességekre vonatkozóan, és minden elég kicsiny kezdeti értékekkel induló $q(t)$ mozgásra $\lim_{t\to\infty}\pi(q(t))$ *létezik.*

Bizonyítás. Alkalmazzuk a 6.2. következményt a $V_1 = T(q,\dot{q})$, $V_2 = \pi(q)$ választással.

Legyen $\Delta > 0$ olyan kicsi, hogy ha $|q_0| < \Delta$ és $|\dot{q}_0| < \Delta$, akkor $T(q_0,\dot{q}) + \pi(q_0) < M$. Mivel $(T+\pi)^{\cdot} \leq 0$ és $T(q,\dot{q}) \geq 0$, az R^{2r} állapottér origójának Δ sugarú környezetéből induló $q(t)$ mozgásokra teljesül a $\pi(q(t)) \leq M$ $(t \in R_+)$ egyenlőtlenség, vagyis $q(t) \in E_M$ minden $t \in R_+$ időpillanatban. Ezért a 6.1.2. következmény feltételeinek ellenőrzésénél eleve feltehetjük, hogy $q \in E_M$. Ekkor viszont $V_2(q,\dot{q}) = (1/2)\dot{q}^T A(q)\dot{q} \geq \lambda_M|\dot{q}|^2$, és

$$(V_1 + V_2)^{\cdot} = -\dot{q}^T B(q,t)\dot{q} \leq -\beta(q,t)|\dot{q}|^2 \leq$$
$$\leq -(\beta(q,t)/\Lambda(q))V_1 = -\alpha_M(t)V_1,$$

vagyis teljesül a 6.1.2. következmény (1)—(3) feltétele. Mivel

$$|\dot{V}_2| = |\dot{q}^T \mathrm{grad}\pi(q)| \leq |\mathrm{grad}\pi(q)|/\sqrt{\lambda_M}\ \sqrt{V_1},$$

a (iii) feltételből következik, hogy a (4) feltétel is teljesül. □

Ha eleve ismeretes, hogy az R^{2r} konfigurációs tér origójának egy környezetéből induló mozgásokra $q(t)$ korlátos a $t \geq t_0$ félegyenesen (individuálisan), akkor a (ii) feltételre nincs szükség, hiszen $\lambda(q) > 0$ $(q \in R^r)$, tehát minden mozgáshoz létezik olyan $\lambda_0 > 0$, hogy $\lambda(q(t)) \geq \lambda_0$ $(t \geq t_0)$ teljesül. Ugyanígy felesleges ebben az esetben a $\mathrm{grad}\pi(q)$ korlátosságára vonatkozó feltétel is (iii)–ban. Ha $\pi(q)$ pozitív definit, akkor az egyensúlyi állapot stabilis (q–ra vonatkozóan is), így feltehetjük, hogy a mozgások az origó egy kompakt környezetében vannak, ami a feltételeket lényegesen egyszerűsíti.

6.2.2. KÖVETKEZMÉNY. *Tegyük fel, hogy*

(i) létezik olyan $K > 0$ szám és $\beta_0(t)$ integrálisan pozitív függvény, hogy

$$\beta(q,t) \geq \beta_0(t) \quad (|q| \leq K);$$

(ii) $\pi(q)$ *pozitív definit.*

Ekkor a (6.2) rendszer $q = \dot{q} = 0$ egyensúlyi állapota stabilis, a sebességekre vonatkozóan aszimptotikusan stabilis, és minden elég kicsiny kezdeti értékekkel induló mozgásra létezik a $\lim_{t\to\infty}\pi(q(t))$ *határérték.*

6.3. Kiegészítő megjegyzések.

Az eredmények [28]-ból valók. TERJÉKI JÓZSEFfel közös [79] dolgozatunkban jelent meg a 6.1.1. tétel egy általánosítása, amelyben a (3) feltételben

$$\dot{V}(x,t) \leq -\omega(t, V_1(x,t)) + r(V(x,t),t)$$

típusú egyenlőtlenséget is megengedünk, és a (4) feltételt is csak egy gyengítettebb formában követeljük meg. Ennek köszönhetően a tétel olyan mechanikai rendszerek vizsgálatára is használhatóvá válik, amelyben a potenciális erők az időtől is függnek.

A 6.1.1. tételnek fontos momentuma, hogy a V *Ljapunov-függvényt* felülről nem korlátozza. További ilyen típusú eredmények találhatók [25] dolgozatomban, ahol alkalmazásként a változó fonalhosszúságú matematikai inga

$$\ddot{\phi} + \left[2\frac{\dot{l}(t)}{l(t)} + \frac{h(t)}{m}\right]\dot{\phi} + \frac{g}{l(t)}\sin\phi = 0$$

mozgásegyenletét tárgyalom. Ebben ϕ a függőlegesen lefelé mutató iránytól való szögkitérést jelöli, $l(t)$ az inga hossza a t időpillanatban (az l függvény „kívülről" tetszőlegesen előírható), m az anyagi pont tömege, g a gravitációs állandó, $h(t)$ a surlódási együttható a t időpillanatban. Bebizonyítom, hogy ha

$$\frac{3\dot{l}(t)}{l(t)} + \frac{2h(t)}{m}$$

integrálisan pozitív, akkor $\phi(t) = o\left(1/\sqrt{l(t)}\right) \quad (t \to \infty)$.

A változó fonalhosszúságú matematikai inga modelljével kapcsolatban sok érdekes probléma merül fel, amelyek megoldása a jövőben is elő fogja mozdítani a nem–autonóm rendszerek stabilitáselméletének fejlődését.

7. fejezet

Utószó

1979–ben közöltem egy hosszabb lélegzetű összefoglaló dolgozatot [H4] ugyanebben a folyóiratban a nem-autonóm differenciálegyenlet-rendszerek stabilitáselméletéről, amelyben számot adtam ezen a területen elért saját eredményeimről is. Jelen dolgozatom ezen áttekintés folytatása; azonos akadémiai doktori értekezésemmel, amelyben az utóbbi években elért, magyarul nem publikált eredményeimet dolgoztam fel. Az értekezés megírása óta is eltelt már egy kis idő, ezért szükségesnek látom az irodalomjegyzék kiegészítését néhány, a témához tartozó azóta megjelent dolgozattal (még akkor is, ha az eredeti irodalomjegyzék sem törekedett teljességre), továbbá az eredmények, módszerek továbbfejlődésének egy rövid, csak az új irányokat megjelölő áttekintését .

A megoldások határhalmazainak *Ljapunov direkt módszerével* való lokalizálása egyre általánosabb rendszerekben nyer alkalmazást.

A nem-autonóm differenciálegyenlet-rendszerek megoldásainak általánosítása a következő fogalom:

7.1. *Definíció.* Legyen M metrikus tér, és legyen egy $u : R_+ \times M \times R \to M$ leképezés adott. Rögzített $t \in R_+$, $t_0 \in R$ számok mellett definiáljuk az $U(t, t_0) : M \to M$ leképezést az $U(t, t_0)x = u(t, x, t_0)$ egyenlőséggel ($x \in M$).

Az $u : R_+ \times M \times R \to M$ leképezést *folyamatnak* nevezzük, ha

(1) u folytonos;

(2) $U(0, t_0) = I$ – az identitás-operátor ($t_0 \in R$);

(3) $U(t, t_0 + s)\ U(s, t_0) = U(t + s, t_0)$ ($t, s \in R_+$, $t_0 \in R$).

Ha $U(t, t_0)$ nem függ t_0-tól, akkor U egy un. *autonóm folyamat.* Ha még azt is feltesszük, hogy u az $R \times M \times R$ halmazon van definiálva, és (3) teljesül minden $t, s \in R$ esetén, akkor u egy dinamikus rendszer a *klasszikus Birkhoff-féle értelemben.*

Az (1.2) kezdetiérték-probléma $x = x(t; x_0, t_0)$ megoldása ($x_0 \in R^n$, $t_0 \in R_+$, $t \in R$) nyilvánvalóan egy folyamat, feltéve, hogy a megoldások folytonosan függnek a kezdeti feltételektől és akármeddig folytathatók *előre* (visszafelé nem feltétlenül!).

Tekintsük az

$$\dot{x}(t) = f(x_t, t) \tag{7.1}$$

funkcionál-differenciálegyenletet a következő standard jelölésekkel: $r > 0$ rögzített valós szám; C a $[-r, 0]$ intervallumot R^k-ba képező folytonos függvények halmaza a maximum-normával; adott $x : [-r, T) \to R^n$ ($T > 0$) függvény esetén $x_t \in C$ az $x_t(s) := x(t + s)$ ($-r \le s \le 0$) egyenlőséggel van definiálva. Be lehet bizonyítani [H1], hogy az

$$\dot{x}(t) = f(x_t, t), \quad x_\sigma = \varphi, \quad (\sigma \in R,\ \varphi \in C)$$

kezdetiérték-probléma $x = x(t; \varphi, \sigma)$ megoldása folyamat a 7.1. definíció értelmében.

További példák hozhatók a folyamat fogalmára a parciális diferenciálegyenletek köréből: a parabolikus és hiperbolikus egyenletekre megfogalmazott kezdeti- és peremértékproblémák megoldásai kielégítik a folyamat defininicióját alkalmasan definiált függvénytereken. Ez a felfogás lehetővé teszi a megoldások kvalitatív jellemzőinek, aszimptotikus viselkedésének a közönséges differenciálegyenletekre kidolgozott módszerekkel való tanulmányozását (l. pl. [H2]). Viszont, mivel az állapottér függvénytér, végtelen dimenziós téren definiált folyamattal van dolgunk, amely tény a közönséges differenciálegyenlet R^k állapotterével összehasonlítva komoly nehézségeket okoz. Ezeket már a (7.1) egyenlettel is jól illusztrálhatjuk, hiszen az ott fellépő C tér végtelen dimenziós.

Tegyük fel, hogy $f(0, t) \equiv 0$ és a (7.1) egyenlet $x = 0$ megoldásának attraktivitására keresünk feltételeket *Ljapunov direkt módszerével.* Ennek legkézenfekvőbb módja az az N. N. KRASZOVSZKIJ [64] által javasolt megközelítés, hogy a *Ljapunov-függvény* szerepét adjuk át egy $V : C \times R \to R$ funkcionálnak, és a funkcionál (7.1) szerinti deriváltjának negatív definitását követeljük meg a

$$\dot{V}(x_t, t) \le -W(\| x_t \|) \tag{7.2}$$

formában, ahol $\|\cdot\|$ a maximum-normát jalöli C–ben. Ezen a módon egy igen jó (megfordítható!) tételhez jutunk, aminek viszont csak elméleti jelentősége van, mivel *közvetlenül* nem alkalmazható: a (7.2) feltételt kielégítő *Ljapunov–Kraszovszkij funkcionált* konstruálni a megoldások ismerete nélkül nagyon nehéz. Könnyebb viszont sok, a gyakorlati alkalmazások szempontjából fontos esetben olyan funkcionált találni, amely eleget tesz egy (7.2) típusú egyenlőtlenségnek, de az $|x(t)|$, illetve az $|||x_t|||$ „normával" az $\| x_t \|$ helyett, ahol $|||\cdot|||$ az L_2–normát jelöli C–ben. Ez is mutatja, hogy a végtelen dimenziós esetben a *Ljapunov–típusú tételekben* a normákat különös gonddal kell kiválasztani. Ehhez a témakörhöz kapcsolódnak T. A. BURTON-nel közös [B1,B2] dolgozataink, amelyekben azt vizsgáljuk, hogyan használható a

$$\dot V(x_t,t) \le -\eta_1(t)W_1(|\,x(t)\,|) - \eta_2(t)W_2(|||x_t|||)$$

$$(\eta_1,\eta_2 : R \to R_+)$$

egyenlőtlenségnek eleget tevő *Ljapunov–Kraszovszkij-funkcionál* a (7.1) egyenlet megoldásai aszimptotikus viselkedésének tanulmányozására. Az így nyert eredmények alkalmazásaként feltételeket adunk a (3.3) típusú rezgés általánosítását szolgáltató

$$\ddot x(t) + a(\dot x(t),t) + f(x(t-h(t))) = 0 \tag{7.3}$$

egyenlet $x = 0$ egyensúlyi helyzetének aszimptotikus stabilitására. A (7.3) egyenlet olyan mozgást ír le, ahol a visszatérítő erő nem a pillanatnyi kitérés, hanem egy korábbi időpontbeli kitérés függvénye. (Ez az egyenlet írja le pl. a napraforgó mozgását.)

KLINCSIK MIHÁLY [K5] dolgozatában azt vizsgálja, hogy az

$$u_{tt} = u_{xx} - p(t)q(u_t) \quad (p(t) > 0) \tag{7.4}$$

hullámegyenletre felírt kezdeti- és peremértékprobléma megoldása hogyan viselkedik a t idő nagy értékeire „kis" és „nagy" $p(t)$ surlódási együttható mellett. Ehhez *Ljapunov direkt módszerét* használja egy alkalmas függvénytéren (az energiatéren) felírt folyamatra, amelyet (7.4) megoldásai határoznak meg. Bizonyításának fő eszköze a 6.1.1. tételnek végtelen dimenziós *Banach–terekre* való általánosítása. Sikerül belátnia, hogy ha a súrlódás elég nagy, akkor az $u_t(t,\cdot) \to 0$ $(t \to \infty)$ az energiatér normájában, és ha a surlódás nem nő „túl gyorsan", akkor $u(t,\cdot) \to 0$ is teljesül.

A *Ljapunov–féle direkt módszernek* a folyamatokra való kiterjesztéséről, annak gyakorlati alkalmazásairól szólnak az összefoglaló [S1, S2] dolgozatok.

A dolgozat 4. és 5. fejezetének témáját adó parciális stabilitás elméletének nagy eseménye az [R1] monográfia megjelenése, amelyet az elmélet megalapítója, V. V. RUMJANCEV és munkatársa, A. SZ. OZIRANER írt. Ez a könnyen olvasható, mind az elméleti megalapozás, mind a gyakorlati alkalmazások szempontjából fontos könyv remélhetőleg további stabilitáselméleti kutatások kiindulópontja lesz.

Végezetül, hálás köszönetemet fejezem ki ARATÓ MÁTYÁS főszerkesztő-helyettesnek és FARKAS MIKLÓS szerkesztőnek hasznos megjegyzéseikért és bátorító tanácsaikért.

Jelölések

R^k: valós számokból álló, k elemű
  $x = \mathrm{col}\,(x^1, x^2, \ldots, x^k)$ oszlopvektorok tere ($k \geq 1$)–állapottér;

$R_+ := [0, \infty)$;

$R_- := (-\infty, 0]$;

$|x|$ ($x \in R^k$): az x vektor normája;

$d(x_1, x_2) := |x_1 - x_2|$ ($x_1, x_2 \in R^k$) – távolság;

$R^k_\infty := R^k \cup \{\infty\}$ – kompaktifikáció;

$d(H, K) := \inf\{d(x_1, x_2) : x_1 \in H,\ x_2 \in K\}$ $(H, K \subset R^k)$;

H^c ($H \subset R^k$): H komplementere;

$\overline{H}$ ($H \subset R^k$): H lezártja;

$H_\infty := H \cup \{\infty\}$: ($H \subset R^k$);

$[a]_+ := \max\{a, 0\}$ ($a \in R$) – szám pozitív része;

$[a]_- := \max\{-a, 0\}$ ($a \in R$) – szám negatív része;

$B_k(p, \rho) := \{x \in R^k : d(x, p) < \rho\}$ ($p \in R^k$, $\rho > 0$): p körüli ρ–sugarú gömb;

$B(H, \rho) := \{x \in R^k : d(x, H) < \rho\}$ ($H \subset R^k$, $\rho > 0$): H ρ–sugarú környezete;

A^T($A \in R^{p\times q}$ – $p \times q$ típusú mátrix): A transzponáltja;

$C^i(H; K)$ ($H \subset R^p$, $K \subset R^q$): i–szer folytonosan differenciálható $H \to K$ függvények osztálya;

$\mathrm{Lip}\,(H; K)$: *Lipschitz-tulajdoságú* függvények osztálya;

$\mathrm{Lip}_x(\Gamma; R)$ ($\Gamma \subset R^k \times R$, $(x, t) \in \Gamma$): x–ben *Lipschitz-tulajdonságú* függvények osztálya;

$h^{-1}(c) := \{x \in G : h(x) = c\}$ ($h : G \subset R^k \to R^r$, $c \in R^r$);

$\dot{f}(t) := df(t)/dt$ ($f : R \to R$ – differenciálható);

$D^+ f(t) := \limsup\limits_{t \to t_0 + 0} \frac{f(t) - f(t_0)}{t - t_0}$ $(f : R \to R)$,

$D_+ f(t) := \liminf\limits_{t \to t_0 + 0} \frac{f(t) - f(t_0)}{t - t_0}$,

$D^- f(t) := \limsup\limits_{t \to t_0 - 0} \frac{f(t) - f(t_0)}{t - t_0}$,

$D_- f(t) := \liminf\limits_{t \to t_0 - 0} \frac{f(t) - f(t_0)}{t - t_0}$, (*Dini-deriváltak*);

$G \subset R^k$: tartomány (összefüggő nyitott halmaz);

$\Gamma := G \times R_+$

$\Lambda \subset \Gamma$

$L(t) := \{x : (x,t) \in \Lambda\}$,

$L := \bigcup_{t \geq 0} L(t)$,

$L^* := \bigcap_{T \geq 0} \overline{\bigcup_{t \geq T} L(t)}$;

$D_t W(x,t) := \partial W(x,t)/\partial t \quad (W : \Gamma \to R^r)$;

$\dot{V}(x,t) = \limsup_{h \to 0+0} (1/h)[V(x + hX(x,t), t+h) - V(x,t)]$: a $V : \Gamma \to R$ függvény $\dot{x} = X(x,t)$ egyenlet szerinti deriváltja:

$x(t; x_0, t_0)$: a differenciálegyenlet $x(t_0; x_0, t_0) = x_0$ kezdeti feltételnek eleget tevő megoldása;

$x = \mathrm{col}\,(y,z) = (y,z)^T$: az $x \in R^k$ vektor partíciója $(y \in R^m,\ z \in R^n;\ 1 \leq m \leq k,\ m + n = k)$;

$y \in R^m$: ellenőrzött koordináták vektora;

$z \in R^n$: nem-ellenőrzött koordináták vektora;

$G_m := B_m(0,h) \times R^n \quad (0 < h \leq \infty)$;

$\Gamma_m := G_m \times R_+$;

$G'_m := B_m(0,h') \times R^n \quad (0 < h' < h)$;

$\Gamma'_m := G'_m \times R_+$;

$\gamma^+(\phi)$, $\gamma(\phi)$: pozitív féltrajektória, ill. teljes trakektória;

$\Omega(\phi)$: pozitív határhalmaz

$\mathcal{K}$: azon $a : R_+ \to R_+$ folytonos, szigorúan monoton növő függvények osztálya, amelyekre $a(0) = 0$;

$V(x,t) \rightrightarrows 0 \quad (x \to 0,\ t \in R_+)$ – egyenletes konvergencia, azaz $\exists b \in K$: $V(x,t) \leq b(|\, x \,|)$ teljesül valamely $B(0,h') \times R_+$ halmazon $(0 < h' < \infty)$;

$(x,t) \mapsto V(x,t) \in R \quad (x \in R^k;\ t \in R_+;\ 0 < \omega \leq \infty,\ c \in R)$

$V^{-1}[c,\omega] := \{p \in R^k \mid \exists \{(x_i,t_i)\}_{i=1}^{\infty} : x_i \to p,\ t_i \to \omega - 0,\ V(x_i,t_i) \to c,\ (i \to \infty)\}$,

$V^{-1}[c,\omega]_0 := \{p \in R^k \mid \exists \{(x_i,t_i)\}_{i=1}^{\infty} : x_i \to p,\ t_i \to \omega - 0,\ V(x_i,t_i) \to c,\ \dot{V}(x_i,t_i) \to 0,\ (i \to \infty)\}$,

$(y,z,t) \mapsto V(y,z,t) \in R \ (y \in R^m,\ z \in R^n;\ t \in R_+,\ c \in R)$

$V_m^{-1}[c,\infty] := \{q \in R^m \mid \exists \{(y_i,z_i,t_i)\}_{i=1}^{\infty} : y_i \to q,\ |z_i| \to \infty,\ t_i \to \infty,\ V(y_i,z_i,t_i) \to c,\ (i \to \infty)\}$,

$V_m^{-1}[c,\infty]_0 := \{q \in R^m \mid \exists \{(y_i,z_i,t_i)\}_{i=1}^{\infty} : y_i \to q,\ |z_i| \to \infty,\ t_i \to \infty,\ V(y_i,z_i,t_i) \to c,\ \dot{V}(y_i,z_i,t_i) \to 0 \ (i \to \infty)$;

$\square$: állítás, bizonyítás vége.

IRODALOM

[1] ARTSTEIN, Z., INFANTE, E. F., "On the asymptotic stability of oscillators with unbounded damping ", *Quarterly Appl. Math.* **34** (1976), 195–199.

[2] ARTSTEIN, Z., "Topological dynamics of an ordinary differential equation ", *J. Differential Equations* **23** (1977), 216–223.

[3] ARTSTEIN, Z., "Topological dynamics of ordinary differential equations and Kurzweil equations ", *J. Differential Equations* **23** (1977), 224–243.

[4] ARTSTEIN, Z., "The limiting equations of nonautonomous ordinary differential equations ", *J. Differential Equations* **25** (1977), 184–202.

[5] BALLIEU, R. J., PEIFFER, K., "Attractivity of the origin for the equation $\ddot{x} + f(t, x, \dot{x})|\dot{x}|^{\alpha}\dot{x} + g(x) = 0$ ", *J. Math. Anal. Appl.* **65** (1978), 321–332.

[6] BIHARI, I., "Extension of a theorem of Armellini–Tonelli–Sansone to the nonlinear equation $u'' + a(t)f(u) = 0$ ", *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **5** (1960), 179–202.

[7] BUDÓ, Á., *Mechanika* (Tankönyvkiadó, Budapest, 1957).

[8] BURTON, T. A., WONG, J. S. W., "Some properties of solutions of $u'' + a(t)f(u)g(u') = 0$, II ", *Monatsh. Math.* **69** (1965), 368–374.

[9] BURTON, T. A., "An extension of Liapunov's direct method ", *J. Math. Anal. Appl.* **28** (1969), 545–522; **32** (1970) 681–691.

[10] BURTON, T. A., "A continuation result for differential equations ", *Proc. Amer. Math. Soc.* **67** (1977), 272–276.

[11] CESARI, L., *Asymptotic Behaviour and Stability Problems in Ordinary Differential Equations* (Springer–Verlag, Berlin, 1959).

[12] GANTMACHER, F., *Lectures in Analytical Mechanics* (Mir, Moscow, 1975).

[13] GRAEF, J. R., SPIKES, P. W., "Asymptotic properties of solutions of a second order nonlinear differential equation ", *Publicationes Math. Debrecen* **24** (1977), 39–51.

[14] HADDOCK, J., "On Liapunov functions for nonautonomous systems ", *J. Math. Anal. Appl.* **47** (1974), 599–603.

[15] HALE, J. K., *Ordinary Differential Equations* (Wiley–Interscience, New York – London – Sydney – Toronto, 1969).

[16] HARTMAN, P., *Ordinary Differential Equations* (Wiley, New York, 1964).

[17] HATVANI, L., "On the asymptotic behaviour of the solutions of $y'' + \Sigma_{i=0}^{n} a_i(x)y^{2i+1} = 0$ ", *Studia Sci. Math. Hungar.* **4** (1969), 355–365.

[18] HATVANI, L., "On the stability of the zero solution of certain second order nonlinear differential equations ", *Acta Sci. Math.* **32** (1971), 1–9.

[19] HATVANI, L., "On the asymptotic behaviour of the solutions of $(p(t)x')' + q(t)f(x) = 0$ ", *Publicationes Math. Debrecen* **19** (1972), 225–237.

[20] HATVANI, L., "Attractivity theorems for nonautonomous systems of differential equations ", *Acta Sci. Math.* **40** (1978), 271–283.

[21] HATVANI, L., "A generalization of Barbashin–Krasovskij theorems to partial stability in nonautonomous systems ", Colloquia Math. Soc. János Bolyai, **30**, *Qualitative Theory of Differential Equations*, Szeged 1979, 381–409.

[22] HATVANI, L., "On the continuation of solutions of differential equations by vector Ljapunov functions ", *Proc. Amer. Math. Soc.* **79** (1980), 59–62.

[23] HATVANI, L., "On partial asymptotic stability and instability. I (Autonomous systems) ", *Acta Sci. Math.* **45** (1983), 219–231.

[24] HATVANI, L., "On partial asymptotic stability and instability. II (The method of limiting equation) ", *Acta Sci. Math.* **46** (1983), 143–156.

[25] HATVANI, L., "On the asymptotic stability by non–decrescent Lyapunov function ", *Nonlinear Anal.* **8** (1984), 67–77.

[26] HATVANI, L., "On location of positive limit sets of solutions of nonautonomous systems ", Colloquia Math. Soc. János Bolyai, **47**, *Differential Equations: Qualitative Theory,* Szeged ,

1984, 413–428.

[27] HATVANI, L., "On partial asymptotic stability by the method of limiting equation ", *Annali di Mat. Pura Appl.* (4) **139** (1985), 65–82.

[28] HATVANI, L., "On partial asymptotic stability and instability. III (Energy–like Ljapunov functions) ", *Acta Sci. Math.* **49** (1985), 157–167.

[29] HATVANI, L., KRÁMLI, A., "A condition for non–asymptotical stability ", Colloquia Math. Soc. János Bolyai, **15**, *Differential Equations,* Keszthely, 1975, 269–276.

[30] KAMKE, E., *Differentialgleichungen (Lösungsmethoden und Lösungen)* (Leipzig, 1959).

[31] KARSAI, J., "On the asymptotic stability of the zero solution of certain nonlinear differential equations ", Colloquia Math. Soc. János Bolyai, **47**, *Differential Equations: Qualitative Theory,* Szeged, 1984, 495–504.

[32] KARSAI, J., "On the global asymptotic stability of the zero solution of the equation $\ddot{x}+$ $+g(t,x,\dot{x})\dot{x}+f(x)=0$ ", *Studia Sci. Math. Hungar.* (megjelenés alatt).

[33] KATO, J., "Asymptotic estimation in functional differential equations via Liapunov function ", Colloquia Math. Soc. János Bolyai, **47**, *Differential Equations: Qualitative Theory,* Szeged, 1984, 505–526.

[34] KOITER, W. T., "On the instability of equilibrium in the absence of a minimum of potential energy ", *Nederl. Akad. Wetensch. Proc. Ser. B* **68** (1965), 107–113.

[35] LASALLE, J. P, "Stability theory for ordinary differential equations ", *J. Differential Eqautions* **4** (1968), 57–65.

[36] LASALLE, J. P., "Stability of nonautonomous systems ", *Nonlinear Anal.* **1** (1976), 83–91.

[37] MARTYNYUK, A. A., "Extension of space states of dynamic systems and the problem of stability ", Colloquia Math. Soc. János Bolyai, **47**, *Differential Equations: Qualitative Theory,* Szeged, 1984, 711–750.

[38] MEIR, A., WILLETT, D., WONG, J. S. W., "On the asymptotic behavior of the solutions of $x''+a(t)x=0$ ", *Michigan Math. J.* **14** (1967), 47–51.

[39] MURAKAMI, S., "Stability of a mechanical system with unbounded dissipative forces ", *Tôhoku Math. Journ.* **36** (1984), 401–406.

[40] MURAKAMI, S., "Asymptotic behavior of solutions of some differential equations ", *J. Math. Anal. Appl.* **109** (1985), 534–545.

[41] ONUCHIC, N., ONUCHIC, L. R., TABOAS, P., "Invariance properties in the theory of stability for ordinary differential systems and applications ", *Applicable Anal.* **5** (1975), 101–107.

[42] PEIFFER, K., "An example of non–isolated equilibrium with maximum potential, stabilized by dissipative forces ", *Z. Angew. Math. Phys.* **30** (1979), 835–837.

[43] PEIFFER, K., ROUCHE, N., "Liapunov's second method applied to partial stability ", *J. Mécanique* **8** (1969), 323–334.

[44] RISITO, C., "Sulla stabilità asintotica parziale ", *Ann. Mat. Pura Appl.* (4) **84** (1970), 279–292.

[45] ROUCHE, N., "On the stability of motion ", *Int. J. Non–Linear Mech.* **3** (1968), 295–306.

[46] ROUCHE, N., HABETS, P., LALOY, M., *Stabilitáselmélet. A Ljapunov–féle direkt módszer* (Műszaki Könyvkiadó, Budapest, 1984).

[47] SALVADORI, L., "Sull' estensione ai sistemi dissipativi del criterio di stabilità del Routh ", *Ricerche Mat.* **15** (1966), 162–167.

[48] SALVADORI, L., "Famiglie ad un parametro di funzioni di Liapunov nello studio della stabilità ", *Symposia Mathematica* **6** (1971), 309–330.

[49] SANSONE, G., *Equazioni differenziali nel campo reale. II,* second ed. (Bologna, 1984).

[50] SELL, G. R., *Topological Dynamics and Ordinary Differential Equations* (Van Nostrand Reinhold Company, London, 1971).

[51] SHESTAKOV, A. A., "A survey on the modern theory of localization of limit sets of dynamical processes via Lyapunov functions ", Colloquia Math. Soc. János Bolyai, **47**, *Differential Equations: Qualitative Theory,* Szeged, 1984, 997–1028.

[52] TERJÉKI, J., "Stability equivalence between damped mechanical and gradient systems", Colloquia Math. Soc. János Bolyai, **47**, *Differential Equations: Qualitative Theory*, Szeged, 1984, 1059–1072.

[53] THOMSON, W., TAIT, P. G., *Principles of Mechanics and Dynamics* (Cambridge Univ. Press, 1912).

[54] THURSTON, L. H., WONG, J. S. W., "On global asymptotic stability of certain second order differential equations with integrable forcing terms", *SIAM J. Appl. Math.* **24** (1979), 50–61.

[55] YOSHIZAWA, T., *Stability Theory by Liapunov's Second Method* (The Mathematical Society of Japan, Tokyo, 1966).

[56] YOSHIZAWA, T., "Asymptotic properties in nonautonomous systems", in *Qualitative Methods of the Theory of Nonlinear Oscillations* (IX. ICNO, Kiev, 1981, Ed.: Yu. Mitropolsky), 159–162.

[57] YOSHIZAWA, T., "Asymptotic behavior of solutions in nonautonomous systems", in *Trends in theory and practice of nonlinear differential equations* (Arlington, Texas, 1982), pp. 553–562, Lecture Notes in Pure and Appl. Math., **90**, Dekker, New York, 1984.

[58] YOSHIZAWA, T., "Asymptotic behaviors of solutions of differential equations", Colloquia Math. Soc. János Bolyai, **47**, *Differential Equations: Qualitative Theory*, Szeged, 1984, 1141–1165.

[59] WALTER, W., *Differential and Integral Inequalities* (Springer–Verlag, Berlin–Heidelberg–New York, 1970).

[60] ANAPOLSKIĬ, L. Ju., IRTEGOV, V. D., MATROSOV, V. M., "Sposoby postroenija funkciĭ Ljapunova", *Itogi nauki i tehniki, Obščaja Mehanika* t. **2**, 53–112.

[61] ANDREEV, A. C., "Ob asimptotičeskoĭ ustoĭčivosti i neustoĭčivosti neavtonomnyh sistem", *Prikl. mat. meh.* **43** (1979), 796–805.

[62] BARBAŠIN, E. A., *Funkcii Ljapunova* (Nauka, Moskva, 1970).

[63] DEMIDOVIČ, B. P., *Lekcii po matematičeskoĭ teorii ustoĭčivosti* (Nauka, Moskva, 1967).

[64] KRASOVSKIĬ, N. N., *Nekotorye zadači teorii ustoĭčivosti dviženija* (Fizmatgiz, Moskva, 1959).

[65] LJAPUNOV, A. M., *Obščaja zadača ob ustoĭčivosti dviženija* (Harkovskogo mat. obšč., Harkov, 1892).

[66] MALKIN, I. G., *Teorija ustoĭčivosti dviženija* (2-oe izd., Nauka, Moskva, 1966).

[67] MARAČKOV, V. P., "Ob odnoĭ teoreme ustoĭčivosti", *Izv. fiz.-mat. ob.-va i NII matematiki i mehaniki pri Kazanskom un.-te* ser. 3 **12** (1940), 171–174.

[68] MARTYNJUK, A. A., GUTOVSKIĬ, P., *Integralnoe neravenstva i ustoĭčivost dviženija* (Naukova dumka, Kiev, 1979).

[69] MATROSOV, V. M., "Ob ustoĭčivosti dviženija", *Prikl. mat. meh.* **26** (1962), 885–895.

[70] MERKIN, D. P., *Giroskopičeskie sistemy* (Nauka, Moskva, 1974).

[71] NEĬMARK, Ju. I., FUFAEV, N. A., *Dinamika negolonomnyh sistem* (Nauka, Moskva, 1967).

[72] OZIRANER, A. S., "Ob asimptotičeskoĭ i neustoĭčivosti otnositelno časti peremennyh", *Prikl. mat. meh.* **37** (1973), 659–665.

[73] OZIRANER, A. C., RUMJANCEV, V. V., "Metod funkciĭ Ljapunova v zadače ob ustoĭčivosti dviženija otnositelno časti peremennyh", *Prikl. mat. meh.* **39** (1972), 364–384.

[74] RUMJANCEV, V. V., "Ob ustoĭčivosti dviženija po otnošeniju k časti peremennyh", *Vestnik MGU* vyp **4** (1957), 9–16.

[75] RUMJANCEV, V. V., "Ob ustoĭčivosti dviženija negolonomnyh sistem", *Prikl. mat. meh.* **31** (1967), 260–271.

[76] RUMJANCEV, V. V., "Ob asimptotičeskoĭ ustoĭčivosti i neustoĭčivosti dviženija po otnošeniju k časti peremennyh", *Prikl. mat. meh.* **35** (1971), 138–143.

[77] TEREKI, J., HATVANI, L., "O častičnoĭ ustoĭčivosti i shodimosti dviženiĭ", *Prikl. mat. meh.* **45** (1981), 428–435.

[78] TEREKI, J., HATVANI, L., "Ob asimptotičeskom ostanavlivanii pri naličii vjazkogo trenija", *Prikl. mat. meh.* **46** (1982), 20–26.

[79] TEREKI, J., HATVANI, L., "Funkcii Ljapunova tipa mehaničeskoĭ energii", *Prikl. mat. meh.*

49 (1985), 894–899.

[80] HATVANI, L., "O primenenii differencialnyh neravenstv k teorii ustoǐčivosti ", *Vetnik MGU. mat.-meh.* vyp. 3 (1975), 83–89.

[81] ČETAEV, N. G., *Ustoǐčivost dviženija* (Gostehizdat, Moskva, 1956).

[82] ŠESTAKOV, A. A., MERENKOV, Ju. N., "O lokalizacii predelnogo množestva v neavtonomnoǐ differencialnoǐ sisteme s pomoščju funkcii Ljapunova ", *Differencialnye uravnenija* 17 (1981), 2017–2028.

KIEGÉSZÍTŐ IRODALOM

[A1] AMINOV, A. B., SIRAZETDINOV , T. K., "Metod funkciǐ Ljapunova v zadačah o poliustoǐčivosti dviženija ", *Prikladnaja matematika i mechanika* 51 (1987), 709–716.

[A2] ANDREEV, A., "Sulla stabilità asintotica ed instabilità ", *Rendiconti del Seminario Mathematico della Universita di Padova* 75 (1986), 235–245.

[A3] ANDREEV, A. S., "Ob issledovanii častičnoǐ asimptotičeskoǐ ustoǐčivosti i neustoǐčivosti na osnove predelnyh uravneniǐ ", *Prikladnaja matematika i mechanika* 51 (1987), 253–259.

[A4] ATHANASSOV, ZHIVKO, S., "Total stability of sets for nonautonomous differential systems ", *Transactions of the Amer. Math. Soc.* 295 (1986), 649–663.

[B1] BURTON, T., HATVANI, L., "Stability theorems for nonautonomous functional differential equations by Liapunov functionals ", *Tohôku Math. Journal* 41 (1989), 65–104.

[B2] BURTON, T., HATVANI, L., "On nonuniform asymptotic stability for nonautonomous functional differential equations ", *Differential and Integral Equations* 3 (1990), 285–293.

[F1] FARKAS, M., GÁSPÁR, Z., KOLLÁR, L., PATKÓ, G., STÉPÁN, G., "Stability investigations of mechanical systems—State of Art ", *Acta Technica Acad. Sci. Hung.* 100(1–2) (1987), 67–99.

[H1] HALE, JACK, *Theory of Functional Differential Equations* (Springer–Verlag, New York–Heidelberg–Berlin, 1977).

[H2] HALE, JACK, *Asymptotic Behavior of Dissipative Systems* (American Mathematical Society, Providence, Rhode Island, 1988).

[H3] HARA, T., YONEYAMA, T., SAITOH, S., HIRANO, M., "Relation between ultimate boundedness and radially unbouded Liapunov function ", *Nonlinear Analysis* 10 (1986), 471–482.

[H4] HATVANI, L., "Nem–autonóm differenciálegyenlet–rendszerek megoldásainak stabilitása és parciális stabilitása ", *Alkalmazott Matematikai Lapok* 5 (1979), 1–48.

[H5] HATVANI, L., "Asymptotic stability of equilibria of damped mechanical systems ", *Differential Equations and Applications,* Ohio University, (1988), 423–427.

[K1] KALOCSAI, V., "Elgséges kritérium másodrendű differenciálegyenlet zérus egyensúlyi helyzetének globális aszimptotikus stabilitsára ", *Alkalmazott Matematikai Lapok* 13 (1987–88), 77–81.

[K2] KARSAI, J., "Egy csillapított rezgőmozgás nemattraktív egyensúlyi helyzettel ", *Alkalmazott Matematikai Lapok* 11 (1985), 167–170.

[K3] KERTÉSZ, V., "Másodrendű differenciálegyenletek parciális aszimptotikus becslései ", *Alkalmazott Matematikai Lapok* 12 (1986), 91–106.

[K4] KERTÉSZ, V., "Zérushoz tartó megoldások létezésének feltételei homogén lineáris differenciálegyenletek esetében ", *Alkalmazott Matematikai Lapok* 12 (1986), 191–205.

[K5] KLINCSIK, M., "Asymptotic behaviour of solutions of nonautonomous dissipative hyperbolic equations ", in *Qualitative Theory of Differential Equations,* Ed. B. Sz.–Nagy and L. Hatvani (János Bolyai Mathematical Society and North–Holland Publishing Company, Amsterdam–Oxford–New York, 1988), pp. 339–346.

[K6] KNJAŽIŠČE, L. B., RAPTUNOVIĆ, "Lokalizacija časti predelnogo množestva i uslovija asimptotičeskoǐ ustoǐčivosti neavtonomnyh uravneniǐ I. ", *Differencialnye uravnenija* 25 (1989), 388–400.

[L1] LAKSHMIKANTHAM, V., XINZHI, LIU, "Perturbing families of Liapunov functions and stability in terms of two measures ", *Journal of Math. Anal. Appl.* **140** (1989), 107–114.

[L2] LAKSHMIKANTHAM, V., "Recent advances in stability theory of nonlinear systems ", in *Qualitative Theory of Differential Equations*, Ed. B. Sz.-Nagy and L. Hatvani (János Bolyai Mathematical Society and North-Holland Publishing Company, Amsterdam-Oxford-New York, 1988), pp. 383–393.

[M1] MARTYNJUK, A. A., "Rasširenie prostranstva sostojaniĭ dinamičeskih sistem i problem ustoĭčivosti ", *Prikladnaja Mehanika* **22** No. 12 (1986), 10–25.

[R1] RUMJANCEV, V. V., OZIRANER, A. C., *Uctoĭčivost i stabilizacija po otnošeniju k časti peremennyh* (Nauka, Moskva, 1987).

[S1] ŠESTAKOV, A. A., "Obobščennyĭ prjamyĭ metod Ljapunova dlja abstraktnyh dinamičeskih processov. I. Poludinamičeskie processy kak poludinamičeskih sistemy. Lokalizacija predelnogo množestva avtonomnyh i asimptotičeski avtonomnyh polydinamičeskix processov ", *Differencialnye uravnenija* **22** (1986), 1475–1489.

[S2] ŠESTAKOV, A. A., "Obobščennyĭ prjamoĭ metod Ljapunova dlja abstraktnyh poludinamičeskih processov. II. Lokalizacija predelnogo množestva kompaktnyh poludinamičeskih processov. Priloženija k uravnenijam s častnymi proizvodnymi ", *Differencialnye uravnenija* **23** (1987), 371–387.

[V1] VORONOV, A. A., MATROSOV, V. M., *Metod vektornyh funkciĭ Ljapunova v teorii ustoĭčivosti* (Nauka, Moskva, 1987).

[V2] VOROTNIKOV, V. I., "Ob ustoĭčivosti po zadannomu čislu peremennyh ", *Prikladnaja matematika i mehanika* **50** (1986), 353–359.

[W1] WADA, T., YAMAMOTO, M., "On the global asymptotic stability of solutions of ordinary differential equations ", *Proc. Japan Acad.* **63** Ser. A (1987), 102–105.

(Beérkezett: 1989. május 8.)

HATVANI LÁSZLÓ
BOLYAI INTÉZET
H-6720 SZEGED, ARADI VÉRTANÚK TERE 1.

ON THE STABILITY OF THE SOLUTIONS OF ORDINARY DIFFERENTIAL EQUATIONS WITH MECHANICAL APPLICATIONS

L. HATVANI

This paper is devoted to the further development and extension of Lyapunov's direct method for nonautonomous systems, for partial stability investigations. The positive limit sets, the positive partial limit sets of solutions are located by a Lyapunov function with negative semidefinite (but not negative definite) derivative. These results are used for establishing theorems on the asymptotic stability and instability of mechanical equilibria. The method of limiting equation is combined with the partial stability investigation.

Alkalmazott Matematikai Lapok 15 (1990–91) 91–113

A CLARKE–FÉLE DERIVÁLT

SZABÓ IMRE

Budapest

Az operációkutatási módszerek hatásosságának a kiterjesztése során egyre nyilvánvalóbbá vált, hogy a hagyományos simasági, általában folytonosan differenciálhatósági feltételek a finomabb modellekben gyakran nem teljesülnek, így az alkalmazhatóság gátjaivá váltak. Ezért szükségessé vált egy olyan elmélet kifejlesztése, amely alkalmas általánosabb szélsőérték-feladatok kezelésére. A *Clarke–féle derivált* segítségével azokra a szélsőérték-feladatokra, amelyekben szereplő függvények nem differenciálhatók, sőt nem is konvexek, de *lokálisan Lipschitz tulajdonságúak*, az ismert szükséges feltételekkel analóg tételek fogalmazhatók meg, például a matematikai programozási feladatra vonatkozó *Lagrange multiplikátor szabály*.

0. Előkészületek

A hatvanas évek elején elsősorban J. J. MOREAU és R. T. ROCKAFELLAR munkásságának köszönhetően a kiterjesztett valós értékű konvex függvényekre bevezették a szubderivált fogalmát. FRANK H. CLARKE a lokálisan Lipschitz tulajdonságú függvényekre általánosította ezt a fogalmat doktori dolgozatában ([1]). A *támaszfüggvény* és a *támaszhalmaz* fogalom pár segítségével a *Clarke–féle derivált* a szubderiválttal egységes keretbe ágyazható. Ebből a felépítésből jól látható a szubderivált és az iránymenti derivált, illetve a *Clarke–féle általánosított derivált* és az iránymenti derivált általánosítása közötti szoros kapcsolat.

Egy adott *Banach téren* értelmezett folytonos és „szublineáris" függvények, valamint a duális tér nem üres, konvex, gyenge*–kompakt részhalmazai kölcsönösen egyértelműen megfeleltethetők egymásnak. A szubderivált és a *Clarke–féle derivált* felépítésének a keretéül ez a kapcsolat szolgál.

Jelöljön a továbbiakban az $(X, \|\cdot\|)$ egy *Banach teret.* Legyen az f ebből a *Banach térből* a valós számokba képező függvény, amely az X egy összefüggő nyílt részhalmazán van értelmezve. Ehhez képest nem jelent megszorítást az, ha azt tesszük fel, hogy az f értelmezési tartománya az egész tér, azaz $D(f) = X$.

0.1. *Definíció.* Az $f : X \to R$ függvényt szublineárisnak nevezzük, ha pozitívan homogén és szubadditív, azaz

(1) $\quad \forall x \in X, \forall \lambda > 0 : f(\lambda x) = \lambda f(x)$

(2) $\quad \forall x, y \in X : f(x + y) \leq f(x) + f(y).$

Ha az $f : X \to R$ függvény szublineáris, akkor konvex, ugyanis $\forall \lambda \in (0,1)$, $\forall x, y \in X$ esetén

$$f(\lambda x + (1-\lambda)y) \leq f(\lambda x) + f((1-\lambda)y) = \lambda f(x) + (1-\lambda)f(y).$$

A következőkben a folytonos és szublineáris függvények fontos szerepet fognak játszani. Jelölje a halmazukat $\mathcal{P}$, azaz

$$\mathcal{P} := \{f : X \to R,\ f \text{ folytonos és szublineáris}\}.$$

Ha az X véges dimenziós tér, akkor a folytonossági megkötés redundáns, ugyanis ismert, hogy véges dimenziós téren értelmezett konvex függvény folytonos az értelmezési tartomány belsejében (lásd például [9]).

Hasonlóképpen fontos szerepet fognak játszani az $X^\star$ duális tér nem üres, konvex, gyenge*–kompakt részhalmazai. Jelölje ezek összességét $\mathcal{M}$, azaz

$$\mathcal{M} := \{U \subseteq X^\star,\ U \neq 0,\ \text{konvex, gyenge}^\star\text{–kompakt}\}.$$

0.2. *Definíció.* Egy $f : X \to R$ függvény támaszhalmazának nevezzük az

$$M_f := \{x^\star \in X^\star : \langle x^\star, x\rangle \leq f(x),\ \forall x \in X\} \subseteq X^\star \text{ halmazt.}$$

0.3. ÁLLÍTÁS. *Ha $f \in \mathcal{P}$, azaz $f : X \to R$ folytonos és szublineáris, akkor $M_f \in \mathcal{M}$, azaz*
(1) $M_f \neq \emptyset$
(2) M_f konvex és gyenge*–zárt
(3) M_f gyenge*–kompakt,
ami azt jelenti, hogy az $M : \mathcal{P} \to \mathcal{M}$ leképezés.

Bizonyítás. (1) A *Banach–Hahn tételből* következik.
(2) A definíció alapján könnyen látható.
(3) Az $M_f(x) := \{\langle x^\star, x\rangle : x^\star \in M_f\} = [-f(-x), f(x)]$ egyenlőség szerint M_f gyenge*–korlátos halmaz, így a Banach-féle egyenletes korlátosság tétele miatt korlátos az $X^\star$ normájában. Mivel (2) szerint gyenge*–zárt is, azért a *Banach–Alaoglu tétel* szerint gyenge*–kompakt. □

0.4. *Definíció.* A $C \subseteq X^\star$ nem üres halmaz támaszfüggvényének nevezzük a

$$\sigma_C(x) := \sup\{\langle x^\star, x\rangle : x^\star \in C\}, \quad \sigma_C : X \to R \cup \{+\infty\} \text{ függvényt.}$$

0.5. ÁLLÍTÁS. *A σ_C támaszfüggvény szublineáris, továbbá, ha a C halmaz korlátos, akkor folytonos is, azaz a $\sigma_C : X \to R$ függvény $\mathcal{P}$–beli. Ami azt jelenti, hogy $\sigma : \mathcal{B} \to \mathcal{P}$, ahol $\mathcal{B}$ az $X^\star$ nem üres, korlátos halmazait jelöli.*

Bizonyítás. $\sigma_C(\lambda x) = \sup\{\langle x^\star, \lambda x\rangle : x^\star \in C\} = \sup\{\lambda\langle x^\star, x\rangle : x^\star \in C\} =$
$= \lambda \sup\{\langle x^\star, x\rangle : x^\star \in C\} = \lambda\sigma_C(x)$, $\quad \sigma_C(x+y) = \sup\{\langle x^\star, x+y\rangle : x^\star \in C\} =$
$= \sup\{\langle x^\star, x\rangle + \langle x^\star, y\rangle : x^\star \in C\} \leq \sup\{\langle x^\star, x\rangle : x^\star \in C\} + \sup\{\langle x^\star, y\rangle : x^\star \in C\} = \sigma_C(x) + \sigma_C(y)$.

Ha C korlátos, akkor $\sigma_C(x) = \sup\{\langle x^\star, x\rangle : x^\star \in C\} \leq \sup\{\|x^\star\| \cdot \|x\| : x^\star \in C\} = \sup\{\|x^\star\| : x^\star \in C\} \cdot \|x\|$. □

0.6. ÁLLÍTÁS. *$\sigma \circ M = \mathrm{id}_{\mathcal{P}}$, azaz bármely $f \in \mathcal{P}$ esetén $f = \sigma(M(f)) = \sigma_{M_f} = \max\{\langle x^\star, \cdot\rangle : x^\star \in M_f\}$.*

Bizonyítás. Nyilván $f \geq \sigma_{M_f}$. Másrészt az f konvex és folytonos volta miatt a *Banach–Hahn tétel* szerint $\forall \varepsilon > 0$ esetén $\forall (x, f(x) - \varepsilon)$ pont szeparálható az $\mathrm{epi} f = \{(x, \alpha) : f(x) \leq \alpha\}$ halmaztól valamilyen nem nulla funkcionállal. □

A $\sigma : \mathcal{B} \to \mathcal{P}$ és az $M : \mathcal{P} \to \mathcal{M}$ leképezésekre tehát teljesül, hogy $\sigma \circ M = \mathrm{id}_{\mathcal{P}}$. Nyilván ekkor a σ szurjektív, és az is könnyen látható, hogy az M injektív. Ha megmutatjuk, hogy a $\sigma|_{\mathcal{M}} : \mathcal{M} \to \mathcal{P}$ leképezés injektív, vagy azt, hogy az $M : \mathcal{P} \to \mathcal{M}$ szurjektív, azaz, hogy $\forall U \in \mathcal{M}$ esetén $\mathcal{M}(\sigma(U)) = \mathcal{M}_{\sigma_u} = U$, akkor $\mathcal{M}$ és $\mathcal{P}$ kölcsönösen egyértelműen megfeleltethetők.

0.7. MEGFELELTETÉSI TÉTEL (*Hörmander*).
Az $X^\star$ nem üres, konvex, gyenge*-kompakt részhalmazai, valamint az X-en értelmezett, valós értékű, folytonos, szublineáris függvények egymásnak kölcsönösen egyértelműen megfeleltethetők, azaz létezik $\Phi : \mathcal{M} \to \mathcal{P}$ bijekció.

Φ nem más mint a támaszfüggvény képzés, azaz $\Phi = \sigma/_{\mathcal{M}}$, az inverze Φ^{-1} pedig a támaszhalmaz képzés, azaz $\Phi^{-1} = M$.

Φ és Φ^{-1} rendezéstartó, azaz, ha $U, V \in \mathcal{M}$, $U \subseteq V$, akkor $\Phi(U) \leq \Phi(V)$, ha $f, g \in \mathcal{P}$, $f \leq g$, akkor $\Phi^{-1}(f) \subseteq \Phi^{-1}(g)$.

Bizonyítás. Φ nyílván szurjektív, azaz $\mathcal{R}(\Phi) = \mathcal{P}$, ugyanis a 0.6. állítás szerint $\forall f \in \mathcal{P}$ esetén az $M_f \in \mathcal{M}$ halmazra $\Phi(M_f) = \sigma_{\mathcal{M}_f} = f$.

Φ injektív is, ugyanis $\forall U, V \in \mathcal{M}$, $U \neq V$ esetén tetszőleges adott $y^\star \in U \setminus V$ és a V gyenge*-kompakt, konvex halmaz szigorúan szeparálhatók egy hipersíkkal, azaz létezik olyan $x_0 \in X$ nem nulla elem amelyre: $\Phi(U)(x_0) = \sigma_U(x_0) = \max\{\langle x^\star, x_0\rangle : x^\star \in U\} \geq \langle y^\star, x_0\rangle > \max\{\langle z^\star, x\rangle : z^\star \in V\} = \sigma(x_0, V) = \Phi(V)(x_0)$, azaz $\Phi(U)(x_0) > \Phi(V)(x_0)$, azaz $\Phi(U) \neq \Phi(V)$.

Φ rendezéstartó : ha $U, V \in \mathcal{M}$, $U \subseteq V$, akkor $\Phi(U) = \sigma_U = \max\{\langle x^\star, \cdot\rangle : x^\star \in U\} \leq \max\{\langle x^\star, \cdot\rangle : x^\star \in V\} = \sigma_V = \Phi(V)$.

Φ^{-1} rendezéstartó : ha $f, g \in \mathcal{P}$, $f \leq g$, akkor $\Phi^{-1}(f) = M_f = \{x^\star \in X^\star : \langle x^\star, \cdot\rangle \leq f\} \leq \{x^\star \in X : \langle x^\star, \cdot\rangle \leq g\} = M_f = \Phi^{-1}(g)$. □

1. A szubderivált

1.1. *Definíció.* Azt mondjuk, hogy az $f : X \to R$ függvény az $x \in X$ pontban a *v irány mentén differenciálható*, ha létezik a következő határérték:

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

Látható, hogy az $f(x, \cdot) : X \to R$ függvény pozitívan homogén, ugyanis $\forall \mu > 0$ esetén $f'(x, \mu v) = \lim\limits_{\lambda \to +0} \dfrac{f(x + \mu\lambda v) - f(x)}{\lambda} = \mu \lim\limits_{\mu \to +0} \dfrac{f(x + \mu\lambda v) - f(x)}{\mu\lambda} = \mu f'(x, v)$. Konvex függvények esetén több is igaz:

1.2. ÁLLÍTÁS. *Ha $f : X \to R$ konvex függvény, akkor*
1. f minden irány mentén differenciálható az egész X–en, azaz bármely $x \in X$ és bármely $v \in X$ esetén létezik $f'(x,v)$.
2. $f'(x,v)$ előáll $f'(x,v) = \inf\limits_{\lambda>0} \dfrac{f(x+\lambda v) - f(x)}{\lambda}$ alakban.
3. Az $f'(x,\cdot) : X \to R$ függvény szubadditív.
Ez a pozitívan homogén tulajdonsággal együtt azt jelenti, hogy az $f'(x,\cdot)$ függvény szublineáris.
4. Ha az f függvény folytonos az $x \in X$ pontban, akkor az $f'(x,\cdot)$ függvény folytonos X–en.

Bizonyítás. Lásd például [9] 4.1. §3. és 4. állítás. □

1.3. *Definíció.* Az $f : X \to R$ konvex függvény $x \in X$ pontjához tartozó *szubderiváltnak* nevezzük a következő $x^\star$-beli halmazt:

$$\partial f(x) := M_{f'(x,\cdot)} = \{x^\star \in X^\star : \langle x^\star, v\rangle \le f'(x,v),\ \forall v \in X\}.$$

A $\partial f(x)$ halmaz elemeit szubgradienseknek nevezzük.

Az elnevezést az indokolja, hogy a szubderivált halmaz elemei — a szubgradiensek — lineáris funkcionálok, amelyek az eredeti függvényt alulról approximálják.

1.4. ÁLLÍTÁS. *Legyen az $f : X \to R$ konvex függvény.*
1. Ha az f folytonos az $x \in X$ pontban, akkor a $\partial f(x) \subseteq X^\star$ nem üres, konvex, gyenge*-kompakt halmaz.
2. $f'(x,\cdot) = \sigma_{\partial f(x)} = \max\{\langle x^\star, \cdot\rangle : x^\star \in \partial f(x)\}$

Bizonyítás.
1. Az 1.2. állítás 4. szerint ekkor az $f'(x,\cdot)$ függvény folytonos az X–en, így a 0.7. megfeleltetési tétel alapján ez a definícióból következik.
2. A 0.6. állítás alapján a definícióból adódik. □

Már említettük, hogy ha az X véges dimenziós, úgy ha az f konvex, akkor folytonos is az értelmezési tartományának a belsejében, azaz most az egész X–en, tehát ebben az esetben a $\partial f(x)$ sohasem üres, konvex, gyenge*-kompakt halmaz. Ha az X nem véges dimenziós, akkor is ismert az a tény, hogy egy konvex függvény pontosan akkor folytonos az értelmezési tartományának a belsejében, azaz most az egész X–en, ha felülről korlátos egy x pont környezetében, ami pedig nem túl erős megkötés.

1.5 *Példa.* (Az abszolútérték és a norma szubderiváltja)
1. Legyen $f(x) := |x|$ $(x \in R)$, ekkor $f'(0,v) = \inf\limits_{\lambda>0} \dfrac{|0+\lambda v| - |0|}{\lambda} = |v|$, így $\partial f(0) = \{x : x \cdot v \le |v|,\ v \in R\} = [-1,+1]$.
2. Legyen $f(x) := \|x\|$ $(x \in X)$, ekkor $f'(0,v) = \inf\limits_{\lambda>0} \dfrac{\|0+\lambda v\| - \|0\|}{\lambda} = \|v\|$, így

$\partial f(0) = \{x^\star \in X^\star : \langle x^\star, v\rangle \leq \|v\|,\ \forall v \in X\} = \mathrm{cl}B^\star$, ahol $B^\star$ az $X^\star$ duális térbeli nulla közepű egységgömböt jelöli.

A fenti fogalmak felépítéséből jól látható, hogy konvex függvény esetén egy adott ponthoz tartozó iránymenti derivált függvény és a szubderivált ugyanannak a dolognak a kétféle reprezentációja. Ez a tény a szubderivált hagyományos definíciójából már nem látszik :

1.6. *Definíció.* Az $f : X \to R$ konvex függvény $x \in X$ pontjához tartozó szubderiváltnak nevezzük a következő halmazt :

$$\partial f(x) := \{x^\star : \langle x^\star, z - x\rangle \leq f(z) - f(x),\ \forall z \in X\}.$$

A definíció geometriai jelentése igen szemléletes. Azt mondja, hogy az f függvény gráfja $(x, f(x))$ pontján keresztül húzott érintők a gráf alatt vannak. Ez is magyarázza a szubderivált elnevezést.

1.7. ÁLLÍTÁS. *A szubderivált 1.3. és 1.6. definíciója ekvivalens.*

Bizonyítás. Jelölje a bizonyításban az 1.6. definícióbeli halmazt $\bar{\partial} f(x)$.

Legyen $x^\star \in \partial f(x)$, azaz $x^\star \in M_{f'(x,\cdot)}$, legyen $z \in X$ tetszőleges, ekkor $\langle x^\star, z - x\rangle \leq f'(x, z - x) = \inf\limits_{\lambda > 0} \dfrac{f(x + \lambda(z - x)) - f(x)}{\lambda} \leq f(z) - f(x)$, azaz $z^\star \in \bar{\partial} f(x)$.

Fordítva : Legyen $x \in \bar{\partial} f(x)$, azaz $\forall z \in X$ esetén $\langle x^\star, z - x\rangle \leq f(z) - f(x)$. Ekkor $\forall v \in X$ és $\forall \lambda > 0$ esetén $\lambda\langle x^\star, v\rangle = \langle x^\star, \lambda v\rangle \leq f(x + \lambda v) - f(x)$, azaz

$$\langle x^\star, v\rangle \leq \frac{f(x + \lambda v) - f(x)}{\lambda} \quad (\lambda > 0), \text{ ebből}$$

$$\langle x^\star, v\rangle \leq \inf_{\lambda > 0} \frac{f(x + \lambda v) - f(x)}{\lambda} = f'(x, v), \text{ azaz } x^\star \in \partial f(x). \qquad \square$$

1.8. *Megjegyzés.* Konvex függvények esetén eltekinthettünk volna attól a feltevéstől, hogy a függvény értelmezési tartománya az egész X. Ugyanis, ha egy f konvex függvényre $D(f) \neq X$, akkor legyen $f_1(x) := \begin{cases} f(x), & \text{ha } x \in D(f) \\ +\infty, & \text{ha } x \notin D(f) \end{cases}$.
Ekkor a $f_1 : X \to R \cup \{+\infty\}$ kiterjesztett valós értékű függvény szintén konvex, és $D(f_1) = X$. Az eredeti f függvény értelmezési tartományát az f_1 függvény effektív tartományának nevezzük, és $\mathrm{dom} f_1$-gyel jelöljük, azaz $\mathrm{dom} f_1 := \{x \in X : f_1(x) < +\infty\}$. (Egy kiterjesztett valós értékű f_1 függvényt valódinak mondunk, ha a $\mathrm{dom} f_1$ nem üres.) A szubderiváltat nyilván csak a $\mathrm{dom} f_1$ pontjaiban értelmezhetjük, az azon kívüli pontokra üres halmaznak definiáljuk. A $\mathrm{dom} f_1$ határán az $f_1{}'(x, \cdot)X \to R \cup \{+\infty\}$ függvény viszont nem folytonos, így az 1.2. állítás 4. pontja már nem biztosítja, hogy a $\partial f_1(x)$ szubderivált halmaz nem üres. Azonban, különösen a szubderivált hagyományos 1.6. definíciójából jól látszik, hogy ha az x a $\mathrm{dom} f_1$ határán van, de az $x \in \mathrm{dom} f_1$, akkor a $\partial f_1(x)$ nem üres, sőt könnyen ellenőrizhető, hogy ott gyenge*-zárt konvex halmaz. A szubderivált ebben különbözik a többi derivált fogalomtól, mert azokat csak belső pontban lehet értelmezni.

2. A Clarke–féle általánosított derivált

A szubderivált konstrukcióját nézve természetesen vetődik fel a kérdés, hogy nem tudnánk-e ezt általánosabb függvény osztályra bevezetni. Ennek az az akadálya — ha a szubderivált 1.3. definíciójára gondolunk — hogy az iránymenti derivált nem létezik bármely függvény esetén. Így először egy olyan általánosított iránymenti deriváltat kellene bevezetni, amely tágabb függvény osztályon is létezik, és ez az iránymenti derivált függvény a második változójában, azaz az $f'(x,\cdot)$ függvény, szublineáris és folytonos.

Az iránymenti derivált legáltalánosabb fogalmát úgy lehet konstruálni, ha nem csak λ, hanem x és v szerint is vesszük a határértéket, vagyis mivel ez nem feltétlenül létezik, azért vagy a felső, vagy az alsó határértéket vesszük.

2.1. *Definíció.* Egy tetszőleges $f : X \to R$ függvény *általánosított iránymenti deriváltja* az $x \in X$ pontban a $v \in X$ irány mentén :

$$f^{\square}(x,v) := \limsup_{(\lambda,x',v')\to(+0,x,v)} \frac{f(x'+\lambda v') - f(x')}{\lambda}.$$

2.2. ÁLLÍTÁS. *Az $f^{\square}(x,\cdot) : X \to R$ függvény szublineáris.*

Bizonyítás. A pozitívan homogenitás bizonyítása ugyanúgy megy, mint az $f'(x,\cdot)$ hagyományos iránymenti derivált függvény esetén.
A szubadditivitás :

$$\begin{aligned} f^{\square}(x, v_1+v_2) &= \limsup_{(\lambda,v',v')\to(+0,x,v_1+v_2)} \frac{f(x'+\lambda v') - f(x')}{\lambda} = \\ &= \limsup_{(\lambda,x',v_1',v'-v_1')\to(+0,x,v_1,v_2)} \left(\frac{f(x'+\lambda v_1') - f(x')}{\lambda} + \right. \\ &\quad \left. + \frac{f[(x'+\lambda v_1') + \lambda(v'-v_1')] - f(x'+\lambda v_1')}{\lambda} \right) \le \\ &\le f^{\square}(x,v_1) + f^{\square}(x,v_2). \quad \square \end{aligned}$$

Az $f^{\square}(x,\cdot)$ függvény tehát szublineáris. Ugyanúgy, mint a konvex függvények esetén az iránymenti derivált függvénynek, az $f^{\square}(x,\cdot)$ függvénynek is vehetjük a támaszhalmazát, azaz az $M_{f^{\square}(x,\cdot)}$ halmazt, amely szintén tekinthető az f függvény x pontbeli deriváltjának. Ha az X véges dimenziós tér, úgy mivel az $f^{\square}(x,\cdot)$ függvény szublineáris, így konvex, azért folytonos is, ezért ekkor a 0.3. szerint a $M_{f^{\square}(x,\cdot)}$ $X^{\star}$-beli halmaz nem üres, konvex, gyenge*-kompakt.

Azonban, ha az X nem véges dimenziós, akkor az $f^{\square}(x,\cdot)$ függvény nem feltétlenül folytonos, ezért a 0.3. állítás nem biztosítja azt, hogy az $M_{f^{\square}(x,\cdot)}$ halmaz nem üres.
Kérdés, hogy milyen függvény osztály mellett lesz nem üres? Látni fogjuk, hogy a *lokálisan Lipschitz tulajdonságú* függvény esetén az $f^{\square}(X,\cdot)$ iránymenti derivált függvény folytonos, ezért az $M_{f^{\square}(x,\cdot)}$ halmaz nem üres, sőt $f^{\square}(x,\cdot)$–re ekkor valamivel egyszerűbb definíció is adható.

2.3. *Definíció.* Azt mondjuk, hogy egy $f : X \to R$ függvény kielégíti a Lipschitz feltételt egy $H \subseteq X$ halmazon, ha létezik egy $L \geq 0$ állandó, hogy minden $x_1, x_2 \in H$ esetén teljesül :

$$|f(x_1) - f(x_2)| \leq L\|x_1 - x_2\|.$$

Azt mondjuk, hogy az f *lokálisan Lipschitz* tulajdonságú az x pontban, ha x–nek van olyan U környezete, hogy f kielégíti a *Lipschitz feltételt* a $H = U$ halmazon.

Az f *lokálisan Lipschitz tulajdonságú,* ha $\forall x \in X$ pontban *Lipschitz tulajdonságú.*

Az f *globálisan Lipschitz tulajdonságú,* ha az f az egész $H = X$ halmazon kielégíti a *Lipschitz feltételt.*

A definícióból látható, hogy ha az $f : X \to R$ függvény *Lipschitz tulajdonságú* egy $x \in X$ pontban, akkor ott folytonos is.

2.4. *Definíció.* Legyen az $f : X \to R$ *lokálisan Lipschitz tulajdonságú* az $x \in X$ pontban. Az f függvény *Clarke-féle általánosított iránymenti deriváltja* az x pontban a v irány mentén :

$$f^{\circ}(x, v) := \limsup_{(\lambda, x') \to (+0, x)} \frac{f(x' + \lambda v) - f(x')}{\lambda}.$$

Az f *Lipschitzessége* miatt ez véges érték.

Az $f^{\circ}(x, v)$ fogalma az $f^{\square}(x, v)$–től abban különbözik, hogy itt v nem változik.

2.5. ÁLLÍTÁS. *Ha az f függvény Lipschitz tulajdonságú az x pontban, akkor az $f^{\square}(x, v)$ megegyezik az $f^{\circ}(x, v)$–vel, azaz ebben az esetben a két általánosított derivált fogalom ekvivalens.*

Bizonyítás. Egyrészt $f^{\circ}(x, v) \leq f^{\square}(x, v)$, másrészt

$$\begin{aligned} f^{\square}(x, v) = & \limsup_{(\lambda, x', v') \to (+0, x, v)} \left(\frac{f(x' + \lambda v') - f(x' + \lambda v)}{\lambda} + \right. \\ & \left. + \frac{f(x' + \lambda v) - f(x')}{\lambda} \right) \leq \limsup_{(\lambda, x', v') \to (+0, x, v)} L\|v - v'\| + \\ & + \limsup_{(\lambda, x') \to (+0, x)} \frac{f(x' + \lambda v) - f(x')}{\lambda} \leq f^{\circ}(x, v). \end{aligned}$$

Igy $f^{\square}(x, v) = f^{\circ}(x, v)$. □

2.6. ÁLLÍTÁS. *Az $f^{\circ}(x, \cdot) : X \to R$ általánosított iránymenti derivált függvény*

(1) szublineáris

(2) globálisan Lipschitz tulajdonságú ugyanazzal a Lipschitz állandóval, mint az f

függvény az x pontban, így folytonos az egész X-en.
(3) $|f^\circ(x,v)| \leq L\|v\|$
(4) $f^\circ(x,-v) = (-f)^\circ(x,v)$
(5) Az $f^\circ(\cdot,\cdot) : X \times X \to R$ *függvény felülről félig folytonos.*

Bizonyítás.
(1) Az $f^\circ(x,\cdot)$ függvény szublinearitása a 2.5. állítás alapján következik az $f^\square(x,\cdot)$ függvény szublinearitásából (lásd 2.2. állítás, egyébként a direkt bizonyítás ugyanúgy megy).
(2) Mivel f *Lipschitz tulajdonságú* az x pontban, azért $\forall v_1, v_2 \in X$ pont és elég kicsi $\lambda > 0$ esetén $f(x' + \lambda v_1') - f(x') \leq f(x' + \lambda v_2') - f(x') + L \cdot \lambda\|v_1 - v_2\|$, ebből $f^\circ(x,v_1) \leq f^\circ(x,v_2) + L\|v_1 - v_2\|$.
Mivel a v_1 és a v_2 szerepe felcserélhető, azért az állítás következik.
(3) Mivel $\forall x' \in X$ és $\forall \lambda > 0$ esetén $\dfrac{f(x'+\lambda v) - f(x')}{\lambda} \leq L\|v\|$, azért $|f^\circ(x,v)| \leq L\|v\|$.

(4)
$$\begin{aligned} f^\circ(x,-v) &= \limsup_{(\lambda,x')\to(+0,x)} \frac{f(x'-\lambda v) - f(x')}{\lambda} = \\ &= \limsup_{(\lambda,u')\to(+0,x)} \frac{(-f)(u'+\lambda v) - f(u')}{\lambda} = \\ &=(-f)^\circ(x,v), \end{aligned}$$

ahol $u' := x - \lambda v$.
(5) Legyenek (x_i) és $(y_i) \subseteq X$ tetszőleges $x_i \to x$, $v_i \to v$ sorozatok. A limesz szuperior definíciója szerint $\forall i$ esetén $\exists$ olyan $y_i \in X$ és $\lambda > 0$, hogy $\|y_i - x_i\| + \lambda_i < 1/i$, és

$$\begin{aligned} f^\circ(x_i,v_i) - 1/i \leq \frac{f(y_i + \lambda_i v_i) - f(y_i)}{\lambda_i} = \\ = \frac{f(y_i+\lambda_i v) - f(y_i)}{\lambda} + \frac{f(y_i + \lambda v_i) - f(y_i + \lambda_i v)}{\lambda_i}. \end{aligned}$$

A második kifejezés nem nagyobb mint $L\|v_i - v\|$ az f *Lipschitzessége* miatt. Mindkét oldal limesz szuperiorját véve, amint $i \to +\infty$, kapjuk, hogy $\limsup\limits_{i\to+\infty} f^\circ(x_i,v_i) \leq f^\circ(x,v)$, azaz az $f^\circ(\cdot,\cdot)$ felülről félig folytonos. □

A fentiekben a *lokálisan Lipschitz tulajdonságú* függvényekre sikerült konstruálni egy olyan általánosított iránymenti derivált fogalmat, amely a második változójában szublineáris és folytonos függvény. Ezért ugyanúgy, mint a konvex függvényeknél a szubderivált konstruálásánál, a 0.7. megfeleltetési tétel alapján, az $f^\circ(x,\cdot) : X \to R$ általánosított iránymenti derivált függvénynek kölcsönösen egyértelműen megfeleltethetjük az $X^\star$ egy nem üres, konvex, gyenge*-kompakt részhalmazát.

2.7. *Definíció.* Legyen az $f : X \to R$ függvény *Lipschitz tulajdonságú* az $x \in X$ pontban. Az f függvény x pontbeli *Clarke-féle általánosított derivált*jának nevezzük a következő halmazt:

$$\partial f(x) := M_{f^\circ(x,\cdot)} = \{x^\star \in X^\star : \langle x^\star, v\rangle \le f^\circ(x,v),\ \forall v \in X\}.$$

2.8. ÁLLÍTÁS. *A Clarke-féle deriváltra teljesül, hogy*

(1) $\partial f(x) \subseteq X^\star$, *nem üres, konvex, gyenge***-kompakt halmaz.*

(2) $f^\circ(x,v) = \sigma_{\partial f(x)} = \max\{\langle x^\star, x\rangle : x^\star \in \partial f(x)\}$

(3) $\|x^\star\|_\star \le L : \forall x^\star \in \partial f(x)$ *esetén*

(4) A $\partial f(\cdot) : X \to X^\star$ *halmazértékű leképzés zárt, abban az értelemben, hogy ha* $x_i{}^\star \in \partial f(x_i) : i = 1,2,\dots$, *és* $x_i \to x$, *továbbá* $x^\star$ *az* $(x_i{}^\star)$ *sorozat torlódási pontja a gyenge***-topológiában, akkor* $x^\star \in \partial f(x)$.

Bizonyítás.

(1) A 0.7. megfeleltetési tétel alapján a definícióból következik.

(2) A 0.6. állítás alapján a definícióból következik.

(3) A definíció szerint $\|x\|_\star = \sup\{\langle x^\star, v\rangle : \|v\| \le 1\}$. Ha $x^\star \in \partial f(x)$, akkor $\langle x^\star, v\rangle \le f^\circ(x,v)$, de a 2.9. állítás (3) szerint $|f^\circ(x,v)| \le L\|v\|$, így $\langle x^\star, v\rangle \le L\|v\|$. Ebből, ha $x^\star \in \partial f(x)$, akkor $\|x^\star\|_\star = \sup\{\langle x^\star, v\rangle : \|v\| \le 1,\ \langle x^\star, v\rangle \le L\|v\|\} \le L$.

(4) $\forall v \in X$ esetén van olyan részsorozata $(x_i{}^\star)$-nak (amit az általánosság megsértése nélkül vehetünk $x_i{}^\star$-nek), hogy $\langle x_i{}^\star, v\rangle \to \langle x^\star, v\rangle$. Definíció szerint $\langle x_i{}^\star, v\rangle \le f^\circ(x,v)$. Az $f^\circ(x,v)$ felülről félig folytonossága miatt (lásd 2.9. állítás (5)), következik, hogy $\langle x^\star, v\rangle \le f^\circ(x,v)$, azaz definíció szerint $x^\star \in \partial f(x)$. □

A *Clarke-féle általánosított derivált* felépítésében nagy súlyt helyeztünk arra, hogy ez is — mint ahogyan a szubderivált — tulajdonképpen egy fogalom pár: az általánosított iránymenti derivált és a *Clarke-féle derivált* ugyanannak a dolognak a kétféle reprezentációja.

2.9. *Példa.* Legyen k nemnegatív egész szám, és legyen az $f : R \to R$ függvény a következő:

$$f(x) := \begin{cases} -x + 1/2^k, & \text{ha} \quad 1/2^{k+1} < x < 1/2^k \\ 2x - 1/2^{2k+1}, & \text{ha} \quad 1/2^{k+2} < x < 1/2^{k+1} \\ x + 1/2^k, & \text{ha} \quad -1/2^k < x < -1/2^{k+1} \\ -2x - 1/2^{k+1}, & \text{ha} \quad -1/2^{k+1} < x < -1/2^{k+2} \\ 0, & \text{ha} \quad x = 0 \end{cases}$$

Ez a függvény nem deriválható a nullában, sőt még az iránymenti deriváltjai sem léteznek itt. Ugyanakkor *Lipschitz tulajdonságú* ebben a pontban ($L = 2$ *Lipschitz állandóval*), ezért vehetjük a *Clarke-féle általánosított iránymenti deriváltját*:

$$f^\circ(0,v) = \limsup_{(\lambda,x')\to(+0,0)} \frac{f(x'+\lambda v) - f(x')}{\lambda} = 2|v|,$$

valamint a *Clarke-féle deriváltját*:

$$\partial f(0) = \{x^\star : x^\star \cdot v \le 2|v|,\ \forall v \in R\} = [-2,2].$$

3. A Clarke–féle derivált kapcsolata a hagyományos deriválttal és a szubderiválttal

Mint az általánosított fogalmaktól, úgy a *Clarke–féle derivált* fogalomtól is elvárjuk, hogy jó kiterjesztése legyen a hagyományos derivált fogalomnak. Látni fogjuk, hogy folytonosan deriválható függvényekre valóban egybeesnek.

3.1. ÁLLÍTÁS. *Legyen az $f : X \to R$ függvény Gâteaux deriválható az $x \in X$ pontban, jelölje a deriváltját itt $Df(x)$. Akkor az f függvény Lipschitz tulajdonságú x–ben és $Df(x) \in \partial f(x)$.*

Bizonyítás. A *Gâteaux derivált* definíciója szerint létezik az $f'(x,v)$ és egyenlő $\langle Df(x), v\rangle$–vel. Mivel $f'(x,\cdot) \le f^\circ(x,\cdot)$, ezért $\langle Df(x),\cdot\rangle \le f^\circ(x,\cdot)$, azaz $Df(x) \in \partial f(x)$. □

Az állítás nyilván akkor is igaz, ha f x–ben *Fréchet deriválható.*

Az ellenkező tartalmazás viszont nem igaz. Tekintsük ugyanis a következő példát:

Legyen $f(x) := \begin{cases} x^2 \sin 1/x, & \text{ha } x \neq 0 \\ 0, & \text{ha } x = 0 \end{cases}.$

Ez a függvény deriválható, így iránymenti deriválható is a nullában:

$$Df(0) = \lim_{x\to 0} \frac{x^2 \sin 1/x - 0}{x} = \lim_{x \to 0} x \cdot \sin 1/x = 0.$$

Ez a függvény azonban *Lipschitz tulajdonságú* is a nullában, az általánosított iránymenti deriváltja:

$$f^\circ(0,v) = \limsup_{(x,\lambda)\to(0,+0)} \frac{(x+\lambda v)^2 \sin 1/x + \lambda v - x \cdot \sin 1/x}{\lambda} = |v|,$$

amiből $\partial f(0) = [-1, 1]$.

Vegyük észre, hogy a $Df : R \to R$ függvény nem folytonos a nullában, ugyanis, ha $x \neq 0$, akkor $Df(x) = 2x \cdot \sin 1/x - \cos 1/x$.

3.2. ÁLLÍTÁS. *Ha az $f : X \to R$ függvény Gâteaux deriválható az $x \in X$ pont egy U környezetében, és a $Df : X \to X^*$ derivált függvény folytonos x–ben, akkor a $\partial f(x)$ halmaz egy pontból áll, és $\partial f(x) = \{Df(x)\}$.*

Bizonyítás. Legyen tetszőleges $v \in X$ esetén $g(\lambda) := f(x' + \lambda v)$. A $g : R \to R$ függvény deriválható a nulla egy jobboldali $[0, \mu]$ környezetében:

$$g'(\lambda) = \langle Df(x' + \lambda v), v\rangle \qquad : \lambda \in [0, \mu]$$

A középérték tétel alapján létezik olyan $\lambda \in [0, \mu]$, hogy

$$\frac{f(x' + \mu v) - f(x')}{\mu} = \langle Df(x' + \lambda v), v\rangle$$

Ebből lim sup–ot véve $\mu \to +0$ mellett, a Df függvény folytonossága miatt $\forall v$ esetén $f^\circ(x,v) = \langle Df(x), v\rangle$, így a $\partial f(x) = M_{\langle Df(x),\cdot\rangle} = \{Df(x)\}$. □

A *Clarke-féle derivált* felépítése a szubderivált felépítésének az általánosított megismétlése volt. Természetesnek érezzük, hogy a *Clarke-féle derivált* a szubderivált fogalmának is jó kiterjesztése a konvex függvényekről a *lokálisan Lipschitz tulajdonságú* függvényekre.

Jelölje a továbbiakban megkülönböztetésül ∂_c a konvex függvényekre bevezetett szubderiváltat.

3.3. ÁLLÍTÁS. *Legyen az* $f : X \to R$ *konvex függvény Lipschitz tulajdonságú az* $x \in X$ *pontban, akkor az* f *Clarke-féle deriváltja az* x *pontban megegyezik a szubderiváltjával, azaz* $\partial f(x) = \partial_c f(x)$*, ami ekvivalens azzal, hogy az* f x *pontjához tartozó általánosított iránymenti derivált függvénye megegyezik az iránymenti derivált függvénnyel, azaz* $f^\circ(x, \cdot) = f'(x, \cdot)$.

Bizonyítás. Ez utóbbit igazoljuk. Nyilván $f'(x, \cdot) \leq f^\circ(x, v)$. Másrészt

$$f^\circ(x,v) = \lim_{\varepsilon \to +0} \sup_{\|x'-x\|<\varepsilon\delta} \sup_{0<\lambda<\varepsilon} \frac{f(x'+\lambda v) - f(x')}{\lambda},$$

ahol $\delta > 0$ tetszőleges rögzített szám. Könnyen ellenőrizhető, hogy ha az f függvény akkor a $t \to \dfrac{f(x'+\lambda v) - f(x')}{\lambda}$ függvény monoton növekedő. Ezért

$$f^\circ(x,v) = \lim_{\varepsilon \to +0} \sup_{\|x'-x\|<\varepsilon\delta} \frac{f(x'+\varepsilon v) - f(x')}{\varepsilon}.$$

Mivel az f *Lipschitz tulajdonságú* az x pontban, azért x-nek van olyan $x + \varepsilon\delta B$ környezete, amelyen az f kielégíti a *Lipschitz feltételt*, amiből

$$\left| \frac{f(x'+\varepsilon v) - f(x')}{\varepsilon} - \frac{f(x+\varepsilon v) - f(x)}{\varepsilon} \right| \leq 2\delta L,$$

így

$$f^\circ(x,v) \leq \lim_{\varepsilon \to +0} \frac{f(x+\varepsilon v) - f(x)}{\varepsilon} + 2\delta L = f'(x,v) + 2\delta L.$$

Mivel δ tetszőleges, azért $f^\circ(x,v) \leq f'(x,v)$. □

4. A Clarke-féle derivált ekvivalens definíciói

Az eddigiekből is látszik, hogy milyen szoros a kapcsolat a *Clarke-féle derivált* és a szubderivált között. Nem meglepő ezért, hogy a szubderivált segítségével további ekvivalens definíciók adhatók.

4.1. ÁLLÍTÁS. (A *Clarke-féle derivált* ekvivalens definíciója)
Az $f : X \to R$ lokálisan Lipschitz tulajdonságú függvény $x \in X$ pontjához tartozó $\partial f(x)$ Clarke-féle deriváltja megegyezik az x pontbeli $f^{\circ}(x, \cdot) : X \to R$ általánosított iránymenti derivált függvény $O \in X$ pontbeli szubderiváltjával, azaz $\partial f(x) = \partial_c f^{\circ}(x, O)$.

Bizonyítás. Mivel az $f^{\circ}(x, \cdot) : X \to R$ függvény szublineáris, azért konvex is. A 0 ponthoz tartozó szubderiváltja az 1.7. definíció szerint:

$$\partial_c f^{\circ}(x, O) = \{x^{\star} : \langle x^{\star}, v - O\rangle \leq f^{\circ}(x, v) - f^{\circ}(x, O),\ \forall v \in X\} =$$

$$= \{x^{\star} : \langle x^{\star}, v\rangle \leq f^{\circ}(x, v),\ \forall v \in X\} = M_{f^{\circ}(x,\cdot)} = \partial f(x). \quad \square$$

Végtelen dimenziós téren értelmezett függvényekre CLARKE eredetileg így vezette be az általánosított derivált fogalmat.

A következőkben a *Clarke-féle deriváltnak* egy, a véges dimenziós téren értelmezett függvényekre érvényes ekvivalens definícióját adjuk. Ehhez szükséges a véges dimenziós téren értelmezett *lokálisan Lipschitz tulajdonságú* függvényekre igaz *Rademacher tétele:*

4.2. ÁLLÍTÁS. *Ha az $f : R^n \to R$ függvény Lipschitz tulajdonságú egy $C \subseteq R^n$ halmaz $\forall$ pontjában, akkor f (Lebesgue mérték szerint) m.m. differenciálható a C halmazon.*

Bizonyítás. Lásd például [8].

4.3. ÁLLÍTÁS. (A *Clarke-féle derivált* véges dimenzió–beli ekvivalens definíciója.)
Legyen az $f : R^n \to R$ függvény Lipschitz tulajdonságú az $x \in X$ pontban. Jelölje Ω azokat a pontokat, ahol az f függvény nem differenciálható. Legyen továbbá $S \subseteq R^n$ tetszőleges nulla Lebesgue mértékű halmaz. Akkor

$$\partial f(x) = \mathrm{co}\{\mathrm{Lim} f'(x_i) : x_i \to x,\ x_i \notin S \cup \Omega\},$$

ahol Lim a torlódási pontok halmazát jelöli.

Bizonyítás. A 4.2. állítás (*Rademacher tétele*) szerint a Ω halmaz mértéke nulla az x pont egy környezetében, így az $\Omega \cup S$ mértéke is az, ezért van olyan $(x_i) \subseteq R^n$ sorozat, amely $x_i \to x$ és létezik $f'(x_i)$. A 3.1. állítás szerint $f'(x_i) \in \partial f(x_i)$ $\forall i$ esetén. Mivel a $\partial f(x)$ halmaz gyenge*-kompakt halmaz (lásd 2.8. állítás (1)), azért lokálisan korlátos is. Ezért a *Bolzano-Weierstrass tétel* szerint a $(f'(x_i)) \subseteq R^n$ sorozatnak van konvergens részsorozata. Ennek határértéke a 2.8. állítás 4. alapján benne van a $\partial f(x)$-ben. Eszerint a $\{\mathrm{Lim} f'(x_i) : x_i \to x,\ x_i \notin S \cup \Omega\}$ halmaz benne van a $\partial f(x)$-ben. A fentiek szerint ez a halmaz nem üres, korlátos és zárt, azaz kompakt, így a konvex burka is az, ugyanis kompakt halmaz konvex burka kompakt. Mivel a $\partial f(x)$ halmaz konvex azért a fenti halmaz konvex burkát is tartalmazza.

A fordított tartalmazáshoz azt látjuk be, hogy a $\partial f(x)$ támaszfüggvénye $f^\circ(x,\cdot)$ kisebb vagy egyenlő a $\mathrm{co}\{\mathrm{Lim} f'(x_i) : x_i \to x,\ x_i \notin S \cup \Omega\}$ halmaz támaszfüggvényénél, azaz $\forall\, 0 \neq v \in R^n$ és $\forall \varepsilon > 0$ esetén

$$f^\circ(x,v) - \varepsilon \leq \limsup\{\langle f'(x_i), v\rangle : x_i \to x,\ x_i \notin S \cup \Omega\}.$$

Az egyenlőtlenség jobboldalát jelölje α. Definíció szerint létezik olyan $\delta > 0$, hogy bármely $y \in x+\delta B$, $y \notin S\cup\Omega$ esetén $\langle f'(y), v\rangle \leq \alpha+\varepsilon$. Tekintsük az $L_y := \{y+\mu v : 0 < \mu < \delta/(2|v|)\}$ szakaszt. Mivel az $(S \cup \Omega) \cap (x + \delta B)$ mértéke nulla, azért a *Fubini tétel* alapján $\lambda-$ m.m. $y \in x + (\delta/2)B$ esetén az L_y szakasz az $S\cup\Omega$ halmazt egy dimenziós nulla mértékű halmazban metszi. Ezért ezen az L_y szakaszon az f $\lambda-$ m.m. differenciálható, azaz létezik az f', így

$$f(y+\lambda v) - f(y) = \int_0^\mu \langle f'(y+\beta v), v\rangle \, d\beta.$$

Mivel $y + \beta v - x \in \delta B$, ahol $0 < \beta < \mu$, azért $\langle f'(y+\beta v), v\rangle \leq \alpha + \varepsilon$. Ez teljesül majdnem minden $y \in x + (\delta/2)B$ és minden $\mu \in (0, \delta/(2|v|))$ esetén. Mivel az f függvény folytonos, azért ez minden y és minden μ esetén teljesül, amiből következik, hogy $f^\circ(x,v) \leq \alpha + \varepsilon$. □

Ez az ekvivalens definíció két dolog miatt is fontos. Egyrészt CLARKE eredetileg így vezette be ezt a fogalmat, lásd [2]. Az ötletet valószínűleg ROCKAFELLAR [7] könyvének 25.6. tétele adhatta, amely a szubderiváltat hasonlóan jellemzi.

Másrészt a gyakorlatban legtöbbször ennek a segítségével számolható ki a *Clarke-féle derivált.*

5. A Clarke–féle derivált tulajdonságai

Az előzőekben már szerepelt, hogy a *Clarke-féle derivált* és az általánosított iránymenti derivált ugyanannak a dolognak a kétféle reprezentációja. A 0.7. megfeleltetési tétel szerint egyrészt egymásnak kölcsönösen egyértelműen megfeleltethetők, azaz bármelyiknek az ismerete egyben a másik ismeretét is jelenti, másrészt $\partial f_1(x) \subseteq \partial f_2(x)$ pontosan akkor, ha $f_1^\circ(x,\cdot) \leq f_2^\circ(x,\cdot)$, azaz bármelyik valamely tulajdonságának az ismerete egyben a másik hasonló tulajdonságának az ismeretét is jelenti. A következő tulajdonságokat ezért mind a *Clarke-féle deriváltra*, mind az általánosított iránymenti deriváltra megfogalmazzuk, de csak az egyikre bizonyítjuk, általában az iránymenti deriváltra, mert erre kényelmesebb.

5.1. ÁLLÍTÁS.
1. Legyen az $f : X \to R$ függvény Lipschitz tulajdonságú az $x \in X$ pontban. Akkor $\forall \lambda \in R$ esetén a λf függvény is Lipschitz tulajdonságú az x pontban, és teljesül, hogy

$$\partial f(\lambda x) = \lambda \partial f(x), \text{ azaz } (\lambda f)^\circ(x,\cdot) = \lambda f^\circ(x,\cdot).$$

2. Legyenek az $f_i : X \to R$ $(i = 1, 2, \dots, n)$ függvények Lipschitz tulajdonságúak az $x \in X$ pontban. Akkor a $\sum_{i=1}^{n} f_i$ függvény is Lipschitz tulajdonságú az x pontban, és teljesül, hogy

$$\partial\left(\sum_{i=1}^{n} f_i\right)(x) \subseteq \sum_{i=1}^{n} \partial f_i(x), \text{ azaz}$$

$$\left(\sum_{i=1}^{n} f_i\right)^{\circ}(x, \cdot) \leq \sum_{i=1}^{n} f^{\circ}(x, \cdot).$$

3. Ha teljesül még az, hogy az f_i függvények legfeljebb egy kivételével folytonosan differenciálhatók, akkor egyenlőség áll fenn.

Összevonva 1.-et és 2.-t: $\forall \lambda_i \in R$ $(i = 1, 2, \dots)$ esetén

$$\partial\left(\sum_{i=1}^{n} \lambda_i f_i\right)(x) \subseteq \sum_{i=1}^{n} \lambda_i \partial f_i(x), \text{ azaz}$$

$$\left(\sum_{i=1}^{n} \lambda_i f_i\right)^{\circ}(x, \cdot) \leq \sum_{i=1}^{n} \lambda_i f^{\circ}(x, \cdot).$$

illetve egyenlőség van, ha a 3. feltételei teljesülnek.

Bizonyítás.

1. Nyilván λf is *Lipschitz tulajdonságú* az x pontban.

Ha $\lambda \geq 0$, akkor

$$\begin{aligned}(\lambda f)^{\circ}(x, v) &= \limsup_{(\mu, x') \to (+0, x)} \frac{\lambda f(x' + \mu v) - \lambda f(x')}{\mu} = \\ &= \limsup_{(\mu, x') \to (+0, x)} \lambda \frac{f(x' + \mu v) - f(x')}{\mu} = \\ &= \lambda \limsup_{(\mu, x') \to (+0, x)} \frac{f(x' + \mu v) - f(x')}{\mu} = \lambda f^{\circ}(x, v).\end{aligned}$$

Ezek után elég az állítást $\lambda = -1$ esetén bizonyítani. Most nem tudjuk közvetlenül azt bizonyítani, hogy $(-f)^{\circ}(x, v) = -f^{\circ}(x, v)$. Ehelyett legyen $x^{\star} \in \partial(-f)(x)$, ez definíció szerint azt jelenti, hogy $\langle x^{\star}, v\rangle \leq (-f)^{\circ}(x, v)$, $\forall v \in X$ esetén. A 2.6. állítás (4) szerint $(-f)^{\circ}(x, v) = f^{\circ}(x, -v)$, így $\langle x^{\star}, v\rangle \leq f^{\circ}(x, -v)$, $\forall v \in X$, azaz $\langle -x^{\star}, -v\rangle \leq f^{\circ}(x, -v)$, $\forall v \in X$, ami azt jelenti, hogy $-x^{\star} \in \partial f(x)$, azaz $x^{\star} \in -\partial f(x)$.

2. Elég az állítást $n = 2$-re bizonyítani, az általános eset teljes indukcióval adódik. Az pedig, hogy

$$(f_1 + f_2)^{\circ}(x, v) \leq f_1^{\circ}(x, v) + f_2^{\circ}(x, v) \quad (\forall v \in X),$$

az $f^\circ(x,v)$ 2.4. definíciójából a lim sup tulajdonsága alapján közvetlenül látszik.
3. Mivel folytonosan differenciálható függvények összege is folytonosan differenciálható, azért elég az állítást csak két olyan függvényre bizonyítani, amelyek közül legalább az egyik folytonosan differenciálható. Ekkor egy olyan öszegnek vesszük a lim sup-ját amely egyik tagjának létezik a limesze, így $(f_1+f_2)^\circ(x,v) = f_1'(x,v) + f_2^\circ(x,v) = f_1^\circ(x,v) + f_2^\circ(x,v)$ $(\forall v \in X)$. □

5.2. *Megjegyzés.* A konvex függvényekre vonatkozó, az 5.1. állítás 2.-höz hasonló állításban, a *Moreau-Rockafellar tételben* (lásd pl. [9] 0.3.§ 0.3.3. tétel) a tartalmazás pont fordított. Ellentmondás nincs, mert ott folytonos konvex függvényekre (, így *Lipschitz tulajdonságú* konvex függvényekre is) egyenlőség áll fenn.

5.3. ÁLLÍTÁS. (A lokális szélsőérték szükséges feltétele.)
Legyen az $f : X \to R$ függvény Lipschitz tulajdonságú az $x \in X$ pontban. Ha az f függvénynek az x pontban lokális szélsőértéke van, akkor $\underset{\sim}{0} \in \partial f(x)$.

(A $\underset{\sim}{0}$ az X^* zérus elemét, azaz az X-en értelmezett azonosan nulla lineáris funkcionált jelöli.)

Bizonyítás. Mivel az 5.1. állítás 1. szerint $\partial(-f) = -\partial f$, azért elég az állítást abban az esetben bizonyítani, ha az x pont lokális minimum hely. Ebben az esetben $\forall v \in X$ esetén $f^\circ(x,v) \geq 0$, azért a *Clarke-féle derivált* 2.11. definíciója szerint $\underset{\sim}{0} \in \partial f(x)$. □

5.4. ÁLLÍTÁS. (Közvetett függvény deriválása.)
Legyen az $f : X \to R$ függvény Lipschitz tulajdonságú az $x \in X$ pontban, a $h : R \to R$ függvény folytonosan differenciálható az $f(x) \in R$ pontban. Akkor a $h \circ f : X \to R$ függvény is Lipschitz tulajdonságú az $x \in X$ pontban, és

$$\partial(h\circ f)(x) \subseteq h'(f(x)) \cdot \partial f(x), \text{ azaz}$$
$$(h\circ f)^\circ(x,v) \leq h'(f(x)) \cdot f^\circ(x,v).$$

Bizonyítás.

$$(h\circ f)^\circ(x,v) = \limsup_{(\lambda,x')\to(+0,x)} \frac{h(f(x'+\lambda v)) - h(f(x'))}{\lambda} =$$
$$= \limsup_{(\lambda,x')\to(+0,x)} \frac{h(f(x'+\lambda v)) - h(f(x'))}{f(x'+\lambda v) - f(x')} \cdot \frac{f(x'+\lambda v) - f(x')}{\lambda} \leq$$
$$\leq \limsup_{(\lambda,x')\to(+0,x)} \frac{h(f(x'+\lambda v)) - h(f(x'))}{f(x'+\lambda v) - f(x')} \limsup_{(\lambda,x')\to(+0,x)} \frac{f(x'+\lambda v) - f(x')}{\lambda} \leq$$
$$\leq \lim_{f(y)\to f(x)} \frac{h(f(y)) - h(f(x))}{f(y) - f(x)} \limsup_{(\lambda,x')\to(+0,x)} \frac{f(x'+\lambda v) - fx')}{\lambda} =$$
$$= h'(f(x)) \cdot f^\circ(x,v). \quad \square$$

5.5. KÖVETKEZMÉNY.
1. Legyen az f_1 és az $f_2 : X \to R$ Lipschitz tulajdonságú az $x \in X$ pontban. Akkor az $f_1 \cdot f_2$ függvény is Lipschitz tulajdonságú az x pontban, és

$$\partial(f_1 \cdot f_2) \subseteq f_2(x)\partial f_1(x) + f_1(x)\partial f_2.$$

2. Az előzőeken kívűl tegyük fel, hogy $f_2(x) \neq 0$. Akkor az $\frac{f_1}{f_2}$ függvény is Lipschitz tulajdonságú az x pontban, és

$$\partial\left(\frac{f_1}{f_2}\right) \subseteq \frac{f_2(x)\partial f_1(x) - f_1(x)\partial f_2(x)}{f_2^2(x)}.$$

Bizonyítás.
1. Legyen $g : R^2 \to R$ a következő függvény: $g(u_1, u_2) := u_1 \cdot u_2$, és legyen $h : X \to R^2$ a következő függvény: $h(x) := [f_1(x), f_2(x)]$. Ekkor $f_1 \cdot f_2 = g \circ h$. Ezek után alkalmazzuk az 5.4. állítást.
2. Az előzőhöz teljesen hasonlóan megy. □

5.6. TÉTEL. (A burkolófüggvény deriváltjáról.)
Legyenek az $f_i : X \to R$ $(i = 1, 2, \ldots, n)$ függvények Lipschitz tulajdonságúak az $x \in X$ pontban. Jelölje f a burkoló (a pontonkénti maximum) függvényüket, azaz

$$f(z) := \max\{f_i(z) : \ i = 1, 2, \ldots, n\} \qquad (\forall\, z \in X).$$

Jelölje $I(z) := \{i : f_i(z) = f(z)\}$, azaz azoknak az indexeknek a halmazát, amelyekre a maximum eléretik. Akkor az f függvény Lipschitz tulajdonságú az x pontban, és

$$\partial f(x) \subseteq \mathrm{co}\{\partial f_i(x) : i \in I(x)\}.$$

Bizonyítás. Azt fogjuk belátni, hogy a bal oldali halmaz támaszfüggvénye nem nagyobb a jobb oldali halmaz támaszfüggvényénél. A bal oldali halmaz támaszfüggvénye:

$$\sigma_{\partial f(x)} = f^{\circ}(x, \cdot) = (\max\{f_i(\cdot) : i = 1, 2, \ldots, n\})^{\circ}(x, \cdot).$$

A jobb oldali halmaz támaszfüggvénye:

$$\begin{aligned}\sigma_{\mathrm{co}\{f_i(x):i\in I(x)\}} &= \max\{\langle x^{\star}, \cdot\rangle \ : x^{\star} \in \mathrm{co}\{\partial f_i(x) : i \in I(x)\}\} = \\ &= \max\{f_i^{\circ}(x, \cdot) : i \in I(x)\}.\end{aligned}$$

Jelölje az $I(x)$ index halmazhoz tartozó függvények felső burkolófüggvényét h, azaz $h(z) := \max\{f_i(z) : i \in I(x)\}$, és jelölje az $I(x)^c$ index halmazhoz tartozó függvények felső burkolófüggvényét g, azaz $g(z) := \max\{f_i(z) : i \notin I(x)\}$.
Mivel h és g folytonosak és $g(x) < h(x)$, azért létezik olyan $x + \delta B$ környezete x-nek, hogy $h|_{x+\delta B} = f|_{x+\delta B}$, ezért minden $v \in X$ esetén $h^{\circ}(x, v) = f^{\circ}(x, v)$. Ezért elég

azt bebizonyítani, hogy $h^\circ(x,\cdot) \leq \max\{f_i^\circ(x,\cdot) : i \in I(x)\}$, ehhez pedig elég azt belátni, hogy létezik olyan $\ell \in I(x)$ index, hogy $h^\circ(x,\cdot) \leq f_\ell^\circ(x,\cdot)$.

Legyenek a $\lambda_n \to +0$ és $x_n \to x$ olyan sorozatok, hogy $\dfrac{h(x_n + \lambda_n v) - h(x_n)}{\lambda_n} \to h^\circ(x,v)$. Ekkor minden n esetén van olyan $i_n \in I(x)$ index, hogy $f_{i_n}(x_n + \lambda_n v) = h(x_n + \lambda_n v)$. Mivel $h(z) = \max\{f_i(z) : i \in I(x)\}$, azért $f_{i_n}(x_n) \leq h(x_n)$, így $\dfrac{h(x_n + \lambda_n v) - h(x_n)}{\lambda_n} \leq \dfrac{f_{i_n}(x_n + \lambda_n v) - f_{i_n}(x_n)}{\lambda_n}$.

Mivel az $I(x)$ index halmaz véges, azért van olyan $\ell \in I(x)$ index, hogy végtelen sok n_k esetén $\ell = i_{n_k} \cdot h^\circ(x,v) = \limsup\limits_{(\lambda_n, x_{n_k}) \to (+0,x)} \dfrac{h(x_{n_k} + \lambda_n v) - h(x_{n_k})}{\lambda} \leq$

$\leq \limsup\limits_{(\lambda_n, x_{n_k}) \to (+0,x)} \dfrac{f_1(x_{n_k} + \lambda_n v) - f_1(x_{n_k})}{\lambda} = f_1^\circ(x,v)$, ez teljesül $\forall v \in X$ esetén.

□

5.6. Megjegyzés. Ez a tétel a *Clarke-féle elmélet* egyik legjelentősebb eredménye. Ugyanis a gyakorlati feladatokban szereplő függvények ha nem differenciálhatók és nem konvexek, akkor általában nem is *lokálisan Lipschitz tulajdonságúak*. Ezt mutatja az is, hogy a *Clarke-féle deriváltat* szemléltető 2.9. példa elég mesterkélt volt. Azonban a gyakorlatban gyakran keressük egy adott függvényhalmaz burkolófüggvényének a szélsőérték helyét. Ennek illusztrálására tekintsük a következő példákat:

1. Jelölje egy termék keresletét az ár függvényében $q(p)$, és legyen Q az a mennyiség amennyit egy adott vállalat ebből a termékből elő tud állítani. Ekkor az a mennyiség amennyit a vállalat el tud adni: $\min\{q(p), Q\}$.

2. A tőkenövekedési modellekben fel szokták tenni, hogy a termelési tényezőket adott arányban használják fel. Például, ha feltesszük, hogy K és L a tőke és a munka mennyisége valamilyen egységekben, amelyeket egyenlő mértékben lehet felhasználni, akkor a termeléshez felhasználható input az egyik–egyik fajtából: $\min\{K, L\}$.

3. Szintén burkoló-függvényhez vezetnek a gazdasági és a műszaki élet küszöbjelenségeinek tanulmányozásai. Szemléletes példa erre egy tartály, amelynek a magassága h_0, a tartályban lévő víz mennyisége a t időpillanatban $h(t)$, akkor az a víztömeg amely a tartályból ki fog folyni arányos a következő függvénnyel: $\max\{h(t) - h_0, 0\}$.

A tételnek nem csak az ilyen irányú alkalmazásai miatt van jelentősége. Alapvető szerepet játszik a szélsőérték feladatok megoldásai szükséges feltételeinek a bizonyításában, így például a matematikai programozási feladatra vonatkozó *Lagrange multiplikátor szabály* bizonyításában is.

6. Matematikai programozás

A legegyszerűbb típusú matematikai programozási feladatra vonatkozó szükséges feltétel bizonyítása a fenti eszközök segítségével nagyon egyszerű.

6.1. *A matematikai programozási feladat I.*
Legyenek az f és $g_i : X \to R$ $(i = 1, 2, \dots, n)$ *lokálisan Lipschitz* (azaz az X halmaz minden pontjában *Lipschitz*) *tulajdonságú* függvények. Tekintsük a következő feladatot:

$$\text{(6.1)} \qquad \begin{cases} f(x) \to min \\ g_i(x) \leq 0 \quad (i = 1, 2, \dots, n) \end{cases}$$

6.2. TÉTEL. (Lagrange multiplikátor szabály.)
Ha az x_0 megoldása a (6.1) matematikai programozási feladatnak, akkor léteznek olyan λ, ν_i $(i = 1, 2, \dots, n)$ nem mind nulla valós számok (szorzók), hogy

$$\text{(1)} \qquad \underset{\sim}{0} \in \lambda \partial f(x_0) + \sum_{i=1}^{n} \nu_i \partial g_i(x_0)$$

$$\text{(2)} \qquad \lambda, \nu_i \geq 0 \quad (i = 1, 2, \dots, n)$$

$$\text{(3)} \qquad \nu_i \cdot g_i(x_0) = 0 \quad (i = 1, 2, \dots, n).$$

Bizonyítás. Jelölje $g_0(x) := f(x) - f(x_0)$, legyen

$$F(x) := \max\{f(x) - f(x_0), g_1(x), \dots, g_n(x)\} = \max\{g_0(x), g_1(x), \dots, g_n(x)\}.$$

Könnyen látható, hogy $F(x_0) = 0$ és $F(x) \geq 0 \; \forall x \in X$ esetén (, ha $F(x) < 0$ volna, akkor az x megengedett és $f(x) < f(x_0)$ lenne). Így x_0 az F minimum helye, ezért az 5.3. állítás alapján $\underset{\sim}{0} \in \partial f(x_0)$.
A burkolófüggvény deriválására vonatkozó 5.6. tétel szerint $\partial F(x_0) \subseteq \text{co}\, \{\partial g_i(x_0) : i \in I(x_0)\}$, ahol $I(x_0)$ azon $i = 1, 2, \dots, n$ indexek halmaza amelyekre a maximum eléretik az x_0 pontban. Ezek szerint a $\partial F(x_0)$ minden pontja előáll ezen függvények konvex kombinációjaként, így a $\underset{\sim}{0}$ pont is, azaz léteznek olyan $\lambda \geq 0, \nu_i \geq 0$ $(i = 1, 2, \dots, n)$, $\lambda + \nu_1 + \ldots + \nu_n = 1$ számok, hogy $\underset{\sim}{0} \in \lambda \partial g_0(x_0) + \sum_{i=1}^{n} \nu_i \partial g_i(x_0)$.
Mivel az 5.1. állítás alapján $\partial g_0(x_0) = \partial[f - f(x_0)](x_0) = \partial f(x_0)$, azért $\underset{\sim}{0} \in \lambda \partial f(x_0) + \sum_{i=1}^{n} \nu_i \partial g_i(x_0)$.
Azoknak a g_i $(i = 1, 2, \dots, n)$ függvényeknek az indexei, amelyekre $g_i(x_0) < 0$ nincsenek benne az $I(x_0)$ indexhalmazban, mert $F(x_0) = 0$. Ezért az ezekre vonatkozó ν_i együtthatók nullák, így $\nu_i \cdot g_i(x_0) = 0 \quad \forall i = 1, 2, \dots, n$ esetén. □

6.3. *Megjegyzés.* Ha a matematikai programozási feladatban feltesszük még, hogy az x pont benne van egy adott halmazban, akkor a távolságfüggvény fogalmának a segítségével az előbbihez hasonló tétel aránylag könnyen megfogalmazható. Azonban, ha még egyenlőségi korlátozások is szerepelnek, akkor már egy kicsivel kevesebbet mondhatunk, nem lehet biztosítani a multiplikátorok nemnegativitását, a bizonyítás pedig nagyon körülményessé válik. A bizonyításhoz felhasználjuk IVAR EKELANDnak egy 1974–ben publikált híres tételét, amelynek számos területen való hasznos alkalmazása ismert (lásd [5]). Ebben az esetben tulajdonképpen a sima esetben használt *Ljusztyernyik tételt* helyettesíti.

6.4. Definíció. Egy $x \in X$ pontnak egy adott $C \subseteq X$ nem üres halmaztól való távolsága:

$$d_C(x) := \inf\{\|x - c\| : c \in C\}.$$

A $d_C : X \to R$ függvényt távolságfüggvénynek nevezzük.

A d_C távolságfüggvény nem differenciálható a C határán, de itt is létezik az általánosított deriváltja, ugyanis:

6.5. ÁLLÍTÁS. *A $d_C : X \to R$ távolságfüggvény globálisan Lipschitz tulajdonságú az $L = 1$ állandóval, azaz $\forall\, x, y \in X$ esetén $|d_C(x) - d_C(y)| \leq \|x - y\|$.*

Bizonyítás. $\forall \varepsilon > 0$-hoz $\exists$ olyan $c \in C$, hogy $\|y - c\| \leq d_C(y) + \varepsilon$. Ekkor $d_C(x) \leq \|x - c\| \leq \|x - y\| + \|y - c\| \leq \|x - y\| + d_C(y) + \varepsilon$.
Itt x és y szerepe felcserélhető. □

6.6. ÁLLÍTÁS. *Legyen $f : X \to R$ Lipschitz tulajdonságú egy S halmazon L Lipschitz állandóval. Legyen $C \subseteq S$ és tegyük fel hogy az f függvény eléri a minimumát a C halmaz felett egy x pontban. Akkor bármely $L' \geq L$ esetén a $g(z) := f(z) + L'd_C(z)$ függvény eléri a minimumát S felett az x pontban.*

Bizonyítás. Indirekt módon tegyük fel, hogy létezik olyan $y \in S$ és $\varepsilon > 0$, hogy $f(y) + L'd_C(x) < f(x) + L'\varepsilon$. Legyen $c \in C$ olyan pont, hogy $\|y - c\| \leq d_C(y) + \varepsilon$. Ekkor $f(c) \leq f(y) + L'\|y - c\| \leq f(y) + L'(d_C(y) + \varepsilon) < f(x)$, ami ellentmondás.

6.7. KÖVETKEZMÉNY. *Ha az $f : X \to R$ függvény eléri a minimumát egy C halmaz felett egy x pontban, és itt lokálisan Lipschitz tulajdonságú L állandóval, akkor*
1. x az $f + Ld_{C\cap S}$ függvény minimum helye,
2. $\underset{\sim}{0} \in \partial f(x) + L\partial d_C(x)$.

Bizonyítás.
1. Legyen S olyan környezete az x pontnak, amelyen a függvény az L állandóval kielégíti a *Lipschitz feltételt.*
Az előző állítás szerint x az $f + Ld_{C\cap S}$ függvény így az $f + Ld_C$ függvény minimum helye is, mivel $d_{C\cap S}(x) = d_C(x)$.
2. Ebből az 5.2. és az 5.4. állítás alapján következik, hogy $\underset{\sim}{0} \in \partial(f + Ld_C)(x) \subseteq \partial f(x) + L\partial d_C(x)$. □

6.8. TÉTEL. (EKELAND) *Legyen (E, ρ) teljes metrikus tér, és legyen $F : E \to R \cup \{+\infty\}$ alulról félig folytonos és alulról korlátos függvény. Legyenek $\varepsilon \geq 0$ és $x \in E$ olyanok, hogy $F(x) \leq \inf_E F + \varepsilon$. Akkor létezik olyan $y \in E$ pont hogy $\forall \lambda > 0$ esetén*
(1) $F(y) \leq F(x)$
(2) $\rho(x, y) \leq 1/\lambda$
(3) $\forall z \neq y$ esetén $F(z) > F(y) - \varepsilon\lambda d(y, z)$.

Bizonyítás. Lásd [5].

6.9. *A matematikai programozási feladat II.*
Legyen $C \subseteq X$ nem üres zárt halmaz, legyenek az $f, g_i, h_j : X \to R$ $(i = 1, 2, \ldots, n;$ $j = 1, 2, \ldots, m)$ a C halmazon *lokálisan Lipschitz* (azaz a C halmaz $\forall$ pontjában *Lipschitz*) *tulajdonságú* függvények. Tekintsük a következő feladatot:

$$(6.2) \qquad \begin{cases} f(x) \to \min \\ g_i(x) \leq 0 \quad (i = 1, \ldots, n), \ h_j(x) = 0 \quad (j = 1, \ldots, m), \quad x \in C \end{cases}$$

6.10. TÉTEL. (Lagrange multiplikátor szabály)
Ha az x_0 megoldása a (6.2) matematikai programozási feladatnak, akkor léteznek olyan λ, ν_i, μ_j $(i = 1, 2, \ldots, n;$ $j = 1, 2, \ldots, m)$ nem mind nulla valós számok (szorzók), hogy

$$(1) \qquad \underset{\sim}{0} \in \lambda \partial f(x_0) + \sum_{i=1}^{n} \nu_i \partial g_i(x_0) + \sum_{i=1}^{m} \mu_j \partial h_j(x_0) + (L+1) \partial_C(x_0)$$

ahol L az összes szereplő függvény együttes Lipschitz állandója,
(2) $\lambda, \nu_i \geq 0 \quad (i = 1, 2, \ldots, n)$,
(3) $\nu_i \cdot g_i(x_0) = 0 \quad (i = 1, 2, \ldots, n)$.

Bizonyítás. A fenti F függvény helyett vezessük be a következő $F_\varepsilon : X \to R$ függvényt:

$$F_\varepsilon(x) := \max\{f(x) - f(x_0) + \varepsilon, g_1(x), \ldots, g_n(x), |h_1(x)|, |h_2(x)|, \ldots, |h_m(x)|\}.$$

Az F_ε függvény *lokálisan Lipschitz tulajdonságú,* így folytonos, továbbá alulról korlátos, mert $F_\varepsilon(x) > 0 \; \forall x \in C$. Az x_0 nem biztos, hogy az F_ε-t minimalizálja, de nyilván $F_\varepsilon(x_0) = \varepsilon \leq \inf\limits_C F_\varepsilon(x) + \varepsilon$. Alkalmazva EKELAND 6.8. tételét $\lambda := 1/\sqrt{\varepsilon}$ mellett kapjuk, hogy létezik olyan $x_\varepsilon \in X$, hogy $\|x_0 - x_\varepsilon\| \leq \sqrt{\varepsilon}$ és x_ε minimalizálja az $F_\varepsilon(x) + \sqrt{\varepsilon}\|x - x_\varepsilon\|$ függvényt a C halmaz felett. Mivel ez a függvény is *Lipschitz tulajdonságú* $L+1$ állandóval, azért a 6.7. következmény 1 pontja szerint x_ε a lokális minimum helye a $G_\varepsilon(x) := F_\varepsilon(x) + \sqrt{\varepsilon}\|x - x_\varepsilon\| + (L+1)d_C(x)$ függvénynek. Ezért a lokális szélsőértékre vonatkozó 5.4. állítás és az 5.2. állítás alapján $\underset{\sim}{0} \in \partial G_\varepsilon(x_\varepsilon) =$ $\partial[F_\varepsilon + \sqrt{\varepsilon}\|\cdot - x_\varepsilon\| + (L+1) \cdot d_C](x_\varepsilon) \subseteq \partial F_\varepsilon(x_\varepsilon) + \sqrt{\varepsilon}\partial\,\|\cdot\|\,(0) + (L+1) \cdot \partial d_C(x_\varepsilon)$. Mivel a norma *Lipschitz tulajdonságú* konvex függvény, azért a 3.3. állítás szerint a *Clarke-féle deriváltja* megegyezik a szubderiváltjával, ez pedig az $0 \in X$ pontban az 1.5. példa alapján az $X^\star$-beli nulla körüli $B^\star$ egységgömb, azért

$$0 \in \partial F_\varepsilon(x_\varepsilon) + \sqrt{\varepsilon}B^\star + (L+1) \cdot \partial d_C(x_\varepsilon).$$

A burkolófüggvény deriválására vonatkozó 5.7. tétel szerint $\partial F_\varepsilon(x_\varepsilon)$ benne van azon függvények *Clarke-féle deriváltjainak* konvex burkában, amelyekben a maximum eléretik. Ezért léteznek olyan $\lambda_\varepsilon, \nu_{\varepsilon,i}, \mu'_{\varepsilon,j} \geq 0$ $(i = 1, 2, \dots, n;\ j = 1, 2, \dots, m)$, $\lambda_\varepsilon + \nu_{\varepsilon,1} + \dots + \nu_{\varepsilon,n} + \mu'_{\varepsilon,1} + \dots + \mu'_{\varepsilon,m} = 1$ számok, hogy

$$\underset{\sim}{0} \in \lambda_\varepsilon \partial f(x_\varepsilon) + \sum_{i=1}^{n} \nu_{\varepsilon,i} \partial g_i(x_\varepsilon) + \sum_{j=1}^{m} \mu'_{\varepsilon,j} \partial |h_j(x_\varepsilon)| + \sqrt{\varepsilon} B^\star + (L+1) \cdot \partial d_C(x_\varepsilon),$$

(az első tag $\partial(f - f(x_\varepsilon) + \varepsilon)(x_\varepsilon) = \partial f(x_\varepsilon)$).

Az $F_\varepsilon(x_\varepsilon) > 0$, különben az x_ε kielégítene minden feltételt és $f(x_\varepsilon) < f(x_0)$ lenne, ami ellentmondana x_0 optimalitásának. Ezért a fenti előállításban csak azok a $|h_j(x_\varepsilon)|$-ok szerepelnek, amelyekre $h_j(x_\varepsilon) > 0$. Mivel az abszolútérték függvény a nulla helyet kivéve folytonosan differenciálható, azért a közvetett függvény deriválására vonatkozó 5.4. állítás alapján

$$\partial |h_j(x_\varepsilon)| \subseteq [h_j(x_\varepsilon)/|h_j(x_\varepsilon)|] \cdot \partial h_j(x_\varepsilon) = [\mathrm{sgn}\ h_j(x_\varepsilon)] \cdot \partial h_j(x_\varepsilon).$$

Legyen $\mu_{\varepsilon,j} := \mu'_{\varepsilon,j}[\mathrm{sgn}\ h_j(x_\varepsilon)]$, erre már nem teljesül, hogy $\mu_{\varepsilon,j} \geq 0$. Ekkor a fentiek alapján $\underset{\sim}{0} \in \lambda_\varepsilon \partial f(x_\varepsilon) + \sum\limits_{i=1}^{n} \nu_{\varepsilon,i} \partial g_i(x_\varepsilon) + \sum\limits_{j=1}^{m} \mu_{\varepsilon,j} \partial h_j(x_\varepsilon) + \sqrt{\varepsilon} B^\star + (L+1) \cdot \partial d_C(x_\varepsilon)$, ez azt jelenti, hogy létezik olyan $u_\varepsilon \in \partial f(x_\varepsilon)$, $v_{\varepsilon,i} \in \partial g(x_\varepsilon)$, $w_{\varepsilon,j} \in \partial h_j(x_\varepsilon)$, $y_\varepsilon \in \sqrt{\varepsilon} B^\star$, $z_\varepsilon \in (L+1) \cdot \partial d_C(x_\varepsilon)$, amelyekre bármely $\varepsilon > 0$ esetén

$$\lambda_\varepsilon u_\varepsilon + \sum_{i=1}^{n} \nu_{\varepsilon,i} \cdot v_{\varepsilon,i} + \sum_{j=1}^{m} \mu_{\varepsilon,j} \cdot w_{\varepsilon,j} + y_\varepsilon + z_\varepsilon = \underset{\sim}{0}.$$

Ha $\varepsilon \to 0$, akkor $x_\varepsilon \to x_0$. Ekkor egyrészt a 2.12. állítás (4) alapján az $(u_\varepsilon, v_{\varepsilon,i}, w_{\varepsilon,j}, y_\varepsilon, z_\varepsilon)$ sorozatnak bármely $(u, v_i, w_j, 0, z)$ torlódási pontjára $u \in \partial f(x_0), v_i \in \partial g(x_0), w_j \in \partial h_j(x_0)$, $y = 0$, $z \in (L+1) \cdot \partial d_C(x_0)$, másrészt a $(\lambda_\varepsilon, \nu_{\varepsilon,i}, \mu_{\varepsilon,j})$ sorozatnak létezik olyan (λ, ν_i, μ_j) torlódási pontja, és az $(u_\varepsilon, v_{\varepsilon,i}, w_{\varepsilon,j}, y_\varepsilon, z_\varepsilon)$ sorozatnak létezik olyan $(u, v_i, w_j, 0, z)$ torlódási pontja, amelyre $\underset{\sim}{0} \in \lambda \partial f(x_0) + \sum\limits_{i=1}^{n} \nu_i \partial g_i(x_0) + \sum\limits_{j=1}^{m} \mu_j \partial h_j(x_0) + (L+1) \cdot \partial d_C(x_0)$.

Mivel a $(\lambda_\varepsilon, \nu_{\varepsilon,i}, \mu'_{\varepsilon,j})$-k egy $n+m$ dimenziós szimplex pontjai, azért a szimplex kompaktsága miatt létezik (λ, ν_i, μ'_j) torlódási pontjuk a szimplexben, így a $(\lambda_\varepsilon, \nu_{\varepsilon,i}, \mu_{\varepsilon,j})$ sorozat (λ, ν_i, μ_j) torlódási pontjában $\lambda, \nu_i \geq 0$.

Ha a $g_i(x_0) < 0$, akkor létezik egy olyan környezete az x_0-nak amelyben $g_i(x) < 0$. Mivel az $F_\varepsilon(x_\varepsilon) > 0$, azért az ilyen $g_i(x_\varepsilon)$-ok nem szerepelnek a $\underset{\sim}{0}$-t konvex kombinációként előállító függvények között, azaz ekkor $\nu_{\varepsilon,i} = 0$, így $\nu_i = 0$, azaz $\nu_i \cdot g_i(x_0) = 0$, $i = 1, 2, \dots, n$. □

6.11. *Megjegyzés.* A tételt nem lehet a 6.2. tételhez hasonlóan bizonyítani. Legyen ugyanis ahhoz hasonlóan legyen $F(x) := \max\{f(x) - f(x_0), g_1(x), \dots, g_n(x), |h_1(x)|, \dots, h_m(x)|\}$. Ugyanúgy most is x_0 az F minimum helye, így $\underset{\sim}{0} \in \partial F(x_0)$, továbbá a burkoló függvény deriválására vonatkozó tétel szerint léteznek olyan $\lambda > 0$, $\nu_i > 0$, $\mu_j > 0$ $(i = 1,2,\dots,n,\ j = 1,2,\dots,m)$, $\lambda + \nu_1 + \ldots + \nu_n + \mu_1, \dots, \mu_m$ számok, hogy

$$\underset{\sim}{0} \in \lambda\partial f(x_0) + \sum_{i=1}^{n} \nu_i \partial g_i(x_0) + \sum_{j=1}^{m} \mu_j \partial |h_j(x_0)|.$$

Most azonban a $\partial|h_j(x_0)|$ kiszámításához nem alkalmazhatjuk a közvetett függvény deriválására vonatkozó 5.4. állítást, mert $h_j(x_0)$ nulla is lehet, és itt az abszolútérték függvény nem differenciálható. Ez az akadály azonban még áthidalható, ugyanis van olyan állítás (lásd [4] 2.3.9 tétel), amely szerint $\partial|h_j(x_0)| \subseteq \bigcup_{|\alpha|\le 1} \alpha \cdot \partial h_j(x_0)$, amiből

$$\underset{\sim}{0} \in \lambda\partial f(x_0) + \sum_{i=1}^{n} \nu_i \partial g_i(x_0) + \sum_{j=1}^{m} \mu_j \cdot \alpha_j \partial h_j(x_0).$$

Úgy látszik, hogy elértük a célunkat, de sajnos nem, mert előfordulhat, hogy $\lambda = 0$, $\mu_i = 0$ $(i = 1,2,\dots,n)$ és $\nu_j \neq 0$ valamilyen j-re, de közben $\alpha_j = 0$, azaz $\nu_j \cdot \alpha_j = 0$, így nem tudtuk biztosítani, hogy a multiplikátorok nem mind nullák.

Ezúton szeretnék köszönetet mondani DANCS ISTVÁNnak, akitől a dolgozat megírása során állandó támogatást és segítséget kaptam.

IRODALOM

[1] CLARKE, F. H., "Necessary conditions for nonsmooth problems in optimal control and the calculus of variations ", Ph. D. Thesis, Univ. of Washington, 1973.

[2] CLARKE, F. H., "Generalized gradients and applications ", *Trans. Am. Math. Soc.* **205** (1975), 247–262.

[3] CLARKE, F. H., "A new approach to Lagrange multipliers ", *Math. Oper. Res.* **1** (1976), 165–174.

[4] CLARKE, F. H., *Optimization and Nonsmooth Analysis* (John Wiley & Sons, 1983).

[5] EKELAND, I., "Nonconvex minimization problems ", *Bull. Am. Math. Soc. (N. S.)* **1** (1979), 443–474.

[6] HARNOS, Zs., *Differenciálszámítás általánosításai* (Magyar Rendszeranalízis Bizottság Kiadvány, 1979).

[7] ROCKAFELLAR, R. T., *Convex Analysis*, Princeton Mathematics Ser., vol. 28 (Princeton Univ. Press, 1970).

[8] STEIN, E. M., *Singular Integrals and Differentiability Properties of Functions*, Princeton Math. Ser., vol. 30 (Princeton Univ. Press, 1970).

[9] JOFFE, A. D., TIHOMIROV, V. M., *Teoriia ekstremalnyh zadač* (Nauka, Moszkva, 1974).

(Beérkezett: 1989. február 6.)

SZABÓ IMRE
ELTE SZÁMÍTÓKÖZPONT, OPERÁCIÓKUTATÁSI OSZTÁLY
1117 BUDAPEST, BOGDÁNFY ÚT 10/B.

CLARKE'S DERIVATIVE

I. SZABÓ

In this paper we present the *Clarke's derivative.* With the help of the dual concept of the support function and the support set give a unified treatment of the *Clarke's derivative* and the subderivative. From this one can see the close relation between the generalized derivative and the generalization of the directional derivative.

By means of the *Lagrange multiplier rule* an example will be given to formulate necessary conditions for the constrained extremum problems with the help of the *Clarke's derivative.*

Alkalmazott Matematikai Lapok 15 (1990–1991) 115–132

MÁTRIXOK KOPOZITIVITÁSÁNAK A VIZSGÁLATÁN ALAPULÓ MÓDSZEREK ALKALMAZÁSA NEMKONVEX KVADRATIKUS OPTIMALIZÁLÁSRA

KÉRI GERZSON

Budapest

A dolgozatban az általános kvadratikus programozási probléma kezelésére alkalmazható olyan módszereket tárgyalunk, amelyek parametrizált mátrixok kopozitivitásának a vizsgálatán alapulnak. A kopozitív, ill. nem kopozitív mátrixok bizonyos tulajdonságaira vonatkozó elméleti eredmények alapján véges algoritmust dolgozunk ki egy kvadratikus programozással kapcsolatos optimalizálási probléma megoldására. A dolgozat végén beszámolunk az eddigi számítástechnikai tapasztalatokról.

1. Bevezetés. A fontosabb fogalmak, jelölések és rövidítések

A dolgozatban egy új megközelítési módot javaslunk az általános kvadratikus programozási (QP) probléma kezelésére. Ennek lényege, hogy az eredeti probléma bizonyos szubproblémáira egy transzformációt alkalmazva, az ennek eredményeként kapott, és a szubproblémával ekvivalens leszármaztatott problémára elvégzünk egy véges sok lépésből álló megoldási algoritmust. Ez az algoritmus négyzetes mátrixok kopozitivitásának tesztelésére a COTTLE-HABETLER-LEMKE [2] cikkben megadott kritériumok alkalmazásán alapul. A dolgozat néhány alapgondolata már felmerült a szerző [4] cikkében. Ott azonban még problémaként szerepelt olyan eljárás kidolgozása, mint amilyent a jelenlegi dolgozat már teljesen kidolgozott algoritmusként tárgyal.

Jelöljük $\mathcal{R}^r$-rel az r dimenziós euklideszi teret, $\mathcal{R}^r_+$-rel pedig az összes r dimenziós nemnegatív vektor halmazát.

1.1. *Definíció.* Egy $r \times r$ méretű valós szimmetrikus $\mathbf{A}$ mátrix pozitív szemidefinit (a továbbiakban PSD), ill. kopozitív (a továbbiakban CP), ha $\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0$ teljesül minden $\mathbf{x} \in \mathcal{R}^r$, ill. minden $\mathbf{x} \in \mathcal{R}^r_+$ vektorra.

Nem kopozitív mátrixra az NCP rövidítéssel fogunk utalni.

1.2. *Definíció.* Egy valós szimmetrikus $\mathbf{A}$ mátrix k-adrendben pozitív szemidefinit (a továbbiakban kPSD), ill. k-adrendben kopozitív (a továbbiakban kCP), ha $\mathbf{A}$-nak minden k-adrendű fő szubmátrixa PSD, ill. CP.

Azért, hogy a következő definíciónak $r = 1$ esetén is legyen értelme, tekintsük konvenciónak, hogy minden valós szimmetrikus mátrix 0CP.

1.3. *Defininíció.* Egy $r \times r$ méretű valós szimmetrikus $\mathbf{A}$ mátrix pontosan k-adrendben kopozitív, ha e mátrix kCP, de nem $(k+1)$CP.

A most bevezetett fogalomnak arra a fontos speciális esetére, amikor $k = r - 1$, külön jelölést, a [2] cikk szerzőinek kezdőbetűiből kialakított CHL szót fogjuk használni. Tehát egy $r \times r$ méretű valós szimmetrikus $\mathbf{A}$ mátrix CHL mátrix, ha ez a mátrix $(r-1)$CP, de nem CP.

Egy négyzetes $\mathbf{A}$ mátrix determinánsát $|\mathbf{A}|$-val, adjungáltját $\operatorname{adj}\mathbf{A}$-val fogjuk jelölni. $\mathbf{A}_{ij}$-vel jelöljük az $\operatorname{adj}\mathbf{A}$ mátrix transzponáltjának (i,j) koordinátájú elemét, ami nem más mint az $\mathbf{A}$ mátrix a_{ij} eleméhez tartozó algebrai komplementum (más szóval: ko–faktor).

A $\mathbf{0}$, ill. $\mathbf{1}$ szám vektorként használva csupa 0, ill. csupa 1 értékű komponensből álló vektort for jelenteni.

2. Az általános QP probléma és a szubprobléma megfogalmazása

Az általános kvadratikus programozási probléma alatt a következő feladatot értjük:

$$\text{(2.1)} \qquad \begin{aligned} &\text{minimalizálandó} \quad f(\mathbf{x}) = \mathbf{x}'\mathbf{D}\mathbf{x} + 2\mathbf{g}'\mathbf{x} \\ &\text{feltéve, hogy} \quad \mathbf{M}\mathbf{x} \le \mathbf{b} \quad \text{és} \quad \mathbf{x} \ge \mathbf{0}. \end{aligned}$$

Itt $\mathbf{D}$ egy adott $n \times n$ méretű mátrix, $\mathbf{g}$ egy adott n dimenziós vektor, $\mathbf{M}$ egy adott $m \times n$ méretű mátrix, $\mathbf{b}$ pedig egy adott m dimenziós vektor.

A QP probléma célfüggvényében szereplő $\mathbf{D}$ mátrixról semmi további kikötést nem teszünk fel. Az $f(\mathbf{x})$ célfüggvény tehát lehet se nem konvex, se nem konkáv.

Rátérünk a szubprobléma megfogalmazására, amelyet e dolgozatban a QP probléma megengedett irányai és megengedett pontjai halmazának véges részhalmazaihoz rendelünk hozzá. Megengedett irányok alatt az $\{\mathbf{x} : \mathbf{M}\mathbf{x} \le \mathbf{0},\ \mathbf{x} \ge \mathbf{0}\}$ halmaz vektorait, megengedett pontok alatt pedig az $\{\mathbf{x} : \mathbf{M}\mathbf{x} \le \mathbf{b},\ \mathbf{x} \ge \mathbf{0}\}$ halmaz elemeit értjük.

Tekintsük most megengedett irányok valamely $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_{r_1}$ véges sorozatát, és megengedett pontok valamely $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_{r_2}$ véges sorozatát. (Feltesszük, hogy $r_1 \ge 0$ és $r_2 \ge 1$.) Ekkor $\mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_{r_1}]$ egy $n \times r_1$ méretű mátrix, $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_{r_2}]$ pedig egy $n \times r_2$ méretű mátrix.

A (2.1)-nél megadott QP probléma valamely szubproblémája alatt a következő módon leírható problémát értjük:

$$\text{(2.2)} \qquad \begin{aligned} &\text{minimalizálandó} \quad f(\mathbf{x}) = \mathbf{x}'\mathbf{D}\mathbf{x} + 2\mathbf{g}'\mathbf{x} \\ &\text{feltéve, hogy} \quad \mathbf{x} \in \mathcal{K} = \{\mathbf{Q}\mathbf{y} + \mathbf{S}\mathbf{z} : \mathbf{y} \ge \mathbf{0},\ \mathbf{z} \ge \mathbf{0},\ \mathbf{1}'\mathbf{z} = 1\}, \end{aligned}$$

ahol $\mathbf{Q}$ és $\mathbf{S}$ a fenti módon megadott mátrixok. A (2.2)-ben feltételi halmazként szereplő $\mathcal{K}$ halmaz nyilvánvalóan részhalmaza az eredeti (2.1) probléma feltételi halmazának. Ezért egy (2.2) alakú probléma optimális megoldásait nevezhetjük szuboptimálisnak a (2.1) problémára nézve. Egy vagy több ilyen szubproblémának a megoldása hasznos információt adhat az általános QP problémára vonatkozóan. Szubproblémák generálására alkalmazhatunk extremális pontok keresésére szolgáló eljárásokat, *Monte-Carlo módszereket* vagy egyéb eljárásokat. Ilyen eljárások részletes kidolgozását azonban nem tekintjük e dolgozat feladatának. A továbbiakban a (2.2) szubproblémával foglalkozunk, ennek megoldását tűzzük ki célul.

3. A szubprobléma átalakítása és a leszármaztatott probléma

Legyen

$$\mathbf{A} = \begin{bmatrix} \mathbf{Q}' & \mathbf{0} \\ \mathbf{S}' & \mathbf{1} \end{bmatrix} \begin{bmatrix} \mathbf{D} & \mathbf{g} \\ \mathbf{g}' & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{Q} & \mathbf{S} \\ \mathbf{0}' & \mathbf{1}' \end{bmatrix}, \tag{3.1}$$

vagyis

$$\mathbf{A} = \begin{bmatrix} \mathbf{Q}'\mathbf{D}\mathbf{Q} & \mathbf{Q}'\mathbf{D}\mathbf{S} + \mathbf{Q}'\mathbf{g}\mathbf{1}' \\ \mathbf{S}'\mathbf{D}\mathbf{Q} + \mathbf{1}\mathbf{g}'\mathbf{Q} & \mathbf{S}'\mathbf{D}\mathbf{S} + \mathbf{1}\mathbf{g}'\mathbf{S} + \mathbf{S}'\mathbf{g}\mathbf{1}' \end{bmatrix}. \tag{3.2}$$

Ekkor az $r = r_1 + r_2$ jelöléssel $\mathbf{A}$ egy $r \times r$ méretű valós szimmetrikus mátrix.

Legyen ezenkívül

$$\mathcal{I} = \{r_1 + 1,\ r_1 + 2, \ldots,\ r_1 + r_2\}, \tag{3.3}$$

$\mathbf{A}(h)$-val pedig jelöljük azt a parametrizált mátrixot, melynek elemei

$$a_{ij}(h) = \begin{cases} a_{ij}, & \text{ha } i \notin \mathcal{I} \text{ vagy } j \notin \mathcal{I} \\ a_{ij} + h, & \text{ha } i \in \mathcal{I} \text{ és } j \in \mathcal{I}. \end{cases} \tag{3.4}$$

Az újonnan bevezetett jelölésekkel a (2.2) szubprobléma célfüggvénye a következőképpen alakítható át:

$$\begin{aligned} f(\mathbf{x}) &= [\mathbf{x}' \quad 1] \begin{bmatrix} \mathbf{D} & \mathbf{g} \\ \mathbf{g}' & 0 \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ 1 \end{bmatrix} = [\mathbf{y}'\mathbf{Q}' + \mathbf{z}'\mathbf{S}' \quad 1] \begin{bmatrix} \mathbf{D} & \mathbf{g} \\ \mathbf{g}' & 0 \end{bmatrix} \begin{bmatrix} \mathbf{Q}\mathbf{y} + \mathbf{S}\mathbf{z} \\ 1 \end{bmatrix} = \\ &= [\mathbf{y}' \quad \mathbf{z}'] \begin{bmatrix} \mathbf{Q}' & \mathbf{0} \\ \mathbf{S}' & \mathbf{1} \end{bmatrix} \begin{bmatrix} \mathbf{D} & \mathbf{g} \\ \mathbf{g}' & 0 \end{bmatrix} \begin{bmatrix} \mathbf{Q} & \mathbf{S} \\ \mathbf{0}' & \mathbf{1}' \end{bmatrix} \begin{bmatrix} \mathbf{y} \\ \mathbf{z} \end{bmatrix} = [\mathbf{y}' \quad \mathbf{z}'] \ \mathbf{A} \begin{bmatrix} \mathbf{y} \\ \mathbf{z} \end{bmatrix} = \\ &= \mathbf{v}'\mathbf{A}\mathbf{v}. \end{aligned} \tag{3.5}$$

Az átalakítás végén a

$$\mathbf{v} = \begin{bmatrix} \mathbf{y} \\ \mathbf{z} \end{bmatrix} \tag{3.6}$$

jelölést alkalmaztuk. A (3.5) átalakításhoz hasonlóan adódik, hogy

$$f(\mathbf{x}) + h = [\mathbf{x}' \quad 1] \begin{bmatrix} \mathbf{D} & \mathbf{g} \\ \mathbf{g}' & h \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ 1 \end{bmatrix} = \mathbf{v}'\mathbf{A}(h)\mathbf{v}, \tag{3.7}$$

tetszőleges valós h esetén. Az utóbbi formula alapján a következő észrevételt fogalmazhatjuk meg: Tetszőleges valós h esetén fennáll, hogy $f(\mathbf{x}) + h \geq 0$ minden $\mathbf{x} \in \mathcal{K}$-ra akkor és csak akkor teljesül, ha $\mathbf{v}'\mathbf{A}(h)\mathbf{v} \geq 0$ minden $\mathbf{v} \in \mathcal{R}^r_+$ vektorra, vagyis ha az $\mathbf{A}(h)$ mátrix CP. Ebből viszont az alábbi két állítás is következik:

3.1. ÁLLÍTÁS. *A*

$$h^* = \inf\{h : \text{az } \mathbf{A}(h) \text{ mátrix } CP\}, \tag{3.8}$$

vagyis

$$h^* = \sup\{h : \text{az } \mathbf{A}(h) \text{ mátrix } NCP\} \tag{3.9}$$

mennyiség akkor és csak akkor véges értékű, ha az $f(\mathbf{x})$ függvény alulról korlátos a $\mathcal{K}$ halmazon.

3.2. ÁLLÍTÁS. *Ha a (3.8)–(3.9)-nél értelmezett h^* mennyiség véges értékű, akkor a (2.2) szubprobléma optimum értéke $-h^*$.*

A 3.2. állítás alapján következik, hogy valamely $\mathbf{x}^* \in \mathcal{K}$ vektor akkor és csak akkor optimális megoldása a (2.2) szubproblémának, ha $f(\mathbf{x}^*) + h^* = 0$, vagyis ha az $\mathbf{x}^*$-nak megfelelő $\mathbf{v}^*$ vektorra $\mathbf{v}^{*\prime}\mathbf{A}(h^*)\mathbf{v}^* = 0$. (Figyelembe véve a (3.7) egyenlőséget.) Igaz tehát a következő állítás is:

3.3. ÁLLÍTÁS. *Ha a (3.8)–(3.9) szerint definiált h^* véges értékű, és valamely $\mathbf{v}^* = \begin{bmatrix} \mathbf{y}^* \\ \mathbf{z}^* \end{bmatrix}$ vektorra teljesül*

$$\begin{gathered} \mathbf{v}^* \in \mathcal{R}^r_+, \\ \sum_{i \in \mathcal{I}} v_i^* = 1, \\ \mathbf{v}^{*\prime}\mathbf{A}(h^*)\mathbf{v}^* = 0, \end{gathered} \tag{3.10}$$

akkor az

$$\mathbf{x}^* = \mathbf{Q}\mathbf{y}^* + \mathbf{S}\mathbf{z}^* \tag{3.11}$$

vektor optimális megoldása a (2.2) szubproblémának. Fordítva, ha a (3.11)-nél adott vektor optimális megoldása a szubproblémának, akkor a $\mathbf{v}^* = \begin{bmatrix} \mathbf{y}^* \\ \mathbf{z}^* \end{bmatrix}$ vektorra fennáll (3.10).

A fentiek szerint a (2.2) szubproblémával ekvivalens a következő, a) és b) részből álló *leszármaztatott probléma:*

a) *Határozzuk meg a (3.8) alatt értelmezett h^* mennyiség értékét. (Döntsük el, hogy h^* véges értékű-e. Ha igen, akkor számítsuk ki az értékét.)*

b) *Amennyiben h^* véges értékű, akkor keressünk egy olyan $\mathbf{v}^*$ vektort, melyre teljesülnek a (3.10) feltételek.*

4. Véges algoritmus a leszármaztatott probléma megoldására

Jelölje $\mathbf{B}_1(h), \mathbf{B}_2(h), \ldots, \mathbf{B}_N(h)$ az $\mathbf{A}(h)$ mátrix fő szubmátrixait rendjük nemcsökkenő sorrendjében, tehát a sorrend elején levő $\mathbf{B}_1(h), \mathbf{B}_2(h), \ldots, \mathbf{B}_r(h)$ mátrixok rendje 1, a sor végén levő $\mathbf{B}_N(h)$ mátrixra pedig $\mathbf{B}_N(h) = \mathbf{A}(h)$. Az azonos rendű szubmátrixok egymás közötti sorrendje tetszőleges lehet.

Az algoritmus menete a következő: A h_{new} és k értékek kezdő beállítása után ciklikusan, k növekvő értékeire elvégzünk egy összetett vizsgálatot, miközben a vizsgálat eredményétől függően h_{new} értékét változatlanul hagyjuk vagy megnöveljük. Az utóbbi esetben a $\mathbf{v}_{new}$ vektort is megadjuk, ill. újraértelmezzük. Az algoritmus végrehajtásának eredményeként vagy kiderül, hogy a probléma nem korlátos (h^* értéke $+\infty$), vagy pedig az utolsóként kapott $(h_{new}, \mathbf{v}_{new})$ pár a probléma megoldását adja ($h^* = h_{new}$ és $\mathbf{v}^* = \mathbf{v}_{new}$).

A ciklikusan ismétlődő vizsgálatot azzal kezdjük, hogy a $\mathbf{B}_k(h_{new})$ mátrixról eldöntjük: CHL mátrix-e. Ehhez COTTLE-HABETLER-LEMKE tétele [2, 3.1 tétel] szerint a szóbanforgó mátrix determinánsának és adjungáltja elemeinek az előjelét kell ellenőrizni. KELLER tétele [2, 4.2. tétel] szerint pedig az adjungált elemeinek az ellenőrzését az adjungált egyetlen oszlopára lehet redukálni.

Az algoritmus folyamatának leírására a PASCAL programozási nyelv néhány egyszerűbb elemét is alkalmazzuk, ezeket ismertnek feltételezzük. A felhasznált PASCAL kulcsszavakat vastag betűkkel írjuk, kiemelés céljából. Az algoritmus folyamata a következő:

$k := 1$;
$h_{new} := -\min\{a_{ii} : i \in \mathcal{I}\} - 1$;
repeat
  $h_{old} := h_{new}$;
  if CHL_test $(\mathbf{B}_k(h_{old}))$ **then** (*nincs teendő*)
    else if $|\mathbf{B}_k(h_{old})| = |\mathbf{B}_k(h_{old} + 1)|$
      then exit (*a probléma nem korlátos*)
      else
        begin
          $h_{new} := h_{old} - \frac{|\mathbf{B}_k(h_{old})|}{|\mathbf{B}_k(h_{old}+1)| - |\mathbf{B}_k(h_{old})|}$;
          $\mathbf{u} :=$ az adj$\mathbf{B}_k(h_{new})$ mátrix bármely nemzéró oszlopa;
          $\mathbf{v} :=$ az $\mathbf{u}$ vektor r dimenziós kiterjesztése;
          $\mathbf{v}_{new} := \frac{\mathbf{v}}{\sum_{i \in \mathcal{I}} v_i}$ (*normálás*)
        end;
  $k := k + 1$
until leállási_feltétel

Kiegészítő magyarázatok az algoritmus folyamatleírásához:

CHL_test egy *Boole típusú függvény*, melynek értéke igaz, ill. hamis attól függően, hogy az argumentumban adott mátrix nem CHL mátrix, ill. CHL mátrix. A függvény értékének kiszámítását KELLER tételének alkalmazásával, a következő eljárással végezhetjük:

function CHL_test (**B**): **Boolean**;
if **B**-nek van nemnegatív sora
  then CHL_test := **true**
  else if $|\mathbf{B}| \geq 0$ **or** adj**B** első oszlopában van negatív elem
    then CHL_test := **true**
    else CHL_test := **false**

A **repeat** szó után rögtön követlező **if** szelekció **then** ága üres; ha tehát $\mathbf{B}_k(h_{new})$ nem CHL mátrix, akkor rögtön k értékének növelésére, és a következő fő szubmátrix vizsgálatára térünk át.

Az $\mathbf{u}$ és $\mathbf{v}_{new}$ vektorok kijelöléséhez az alábbi magyarázatot fűzzük: Az algoritmus bizonyítása folyamán meg fogjuk mutatni, hogy ha $|\mathbf{B}_k(h_{old})| \neq |\mathbf{B}_k(h_{old}+1)|$, akkor az adj$\mathbf{B}_k(h_{new})$ mátrixnak van zéró vektortól különböző oszlopvektora. Ha $i_1, i_2, \ldots, i_s$-nek nevezzük a $\mathbf{B}_k$ szubmátrix sorainak az $\mathbf{A}$ mátrixban elfoglalt sorindexeit, akkor az $\mathbf{u}$ vektor r dimenziós kiterjesztésén azt a $\mathbf{v}$ vektort értjük, melyre

$$v_{i_j} = u_j \quad (j = 1, 2, \ldots, s), \tag{4.1}$$

és

$$v_i = 0, \text{ ha } i \notin \{i_1, i_2, \ldots, i_s\}. \tag{4.2}$$

A kiterjesztés által kapott $\mathbf{v}$ vektorból az algoritmus szerint történő normálással adódik a $\mathbf{v}_{new}$ vektor. A normálás célja annak biztosítása, hogy a $\mathbf{v}_{new}$ vektorra (3.10) második sora teljesüljön. (Be fogjuk látni, hogy a nevezőben levő $\sum_{i \in I} v_i$ értéke nem lehet zéró.)

A $\mathbf{v}_{new}$ vektor meghatározása után érdemes kiszámítani az

$$\mathbf{x} = [\mathbf{Q} \quad \mathbf{S}]\,\mathbf{v}_{new}, \tag{4.3}$$

vektort nemcsak az eljárás végén, hanem közben is, minden olyan alkalommal, amikor az algoritmus menete új $\mathbf{v}_{new}$ vektor előállítását eredményezi. Nyilvánvaló ugyanis, hogy valamennyi így kapott $\mathbf{x}$ vektor a QP szubprobléma megengedett megoldása. Az algoritmus bizonyítása során pedig az is ki fog derülni, hogy az egymás után képzett $\mathbf{x}$ vektorokhoz egyre javuló $-h_{new}$ célfüggvényértékek tartoznak. Ezért az algoritmus számításainak bármilyen okból történő félbeszakadása esetén az addig esetleg már megkapott részeredmények is felhasználhatók, ha az algoritmust kiegészítjük a QP szubprobléma megengedett megoldásainak (4.3) szerint történő generálásával.

A következőkben rátérünk a leállási feltétel értelmezésére. Az **until** szó után írhattuk volna konkrétan hogy $k > N$, amely a legtermészetesebb és legegyszerűbb lehetséges leállási feltétel. Ilyenkor ténylegesen sorravesszük az $\mathbf{A}$ mátrix valamenynyi $\mathbf{B}_k$ fő szubmátrixát. Az így elvégzett algoritmus tehát exponenciális lépésszámú.

Megadhatók azonban más egyéb leállási feltételek, amelyek a lépésszám redukcióját eredményezik. Ezek alkalmazása esetén a fő szubmátrixok közül csak a viszonylag kisebb méretűeket kell sorravenni. Vezessük be mátrixok méretének jellemzésére a következő definícióban szereplő kifejezéseket.

4.1. *Definíció (és jelölések).* Jelöljük egy tetszőleges valós szimmetrikus $\mathbf{A}$ mátrix esetén $\mathrm{ordo}\mathbf{A}$-val az $\mathbf{A}$ mátrix rendjét, $\mathrm{rank}\mathbf{A}$-val pedig a rangját. Nevezzük $\mathrm{negordo}\mathbf{A}$-nak az $\mathbf{A}$ mátrix azon sorainak számát, melyekben van negatív elem, $\mathrm{negrank}\mathbf{A}$-nak pedig annak a mátrixnak a rangját, amely $\mathbf{A}$-ból a csupa nemnegatív elemet tartalmazó sorok és az ugyanilyen oszlopok elhagyásával keletkezik.

Az algoritmus leállási feltételeként a következő feltételek bármelyike választható:

$$\begin{aligned}
&\mathrm{ordo}\mathbf{B}_k > \mathrm{rank}\mathbf{A}(h_{new}),\\
&\mathrm{ordo}\mathbf{B}_k > \mathrm{negordo}\mathbf{A}(h_{new}),\\
&\mathrm{ordo}\mathbf{B}_k > \mathrm{negrank}\mathbf{A}(h_{new}),\\
&\mathrm{ordo}\mathbf{B}_k > \mathrm{ordo}\mathbf{D} + 1 \quad (= n + 1),\\
&\mathrm{ordo}\mathbf{B}_k > \mathrm{rank}\mathbf{D} + 2.
\end{aligned}$$

A felsoroltak közül a leghamarabb történő leállást a harmadik feltétel biztosítja, a többiek viszont könnyebben ellenőrizhetők. A felsorolt leállási feltételek jobb oldalai között az alábbi relációk állnak fenn:

$$\begin{aligned}
\mathrm{negrank}\mathbf{A}(h_{new}) &\le \mathrm{rank}\mathbf{A}(h_{new}) \le \mathrm{ordo}\mathbf{D} + 1,\\
\mathrm{rank}\mathbf{A}(h_{new}) &\le \mathrm{rank}\mathbf{D} + 2,\\
\mathrm{negrank}\mathbf{A}(h_{new}) &\le \mathrm{negordo}\mathbf{A}(h_{new}).
\end{aligned}$$

A szükséges bizonyításokat a következő két szakaszban adjuk meg.

5. A véges algoritmus alapjául szolgáló tételek

Először felsorolunk CP és CHL mátrixok fontosabb tulajdonságaival kapcsolatos, majd valós négyzetes mátrix adjungáltjával kapcsolatos néhány állítást. Ezek általában nagyon egyszerűen, elemi módon bizonyíthatók, ezért bizonyításukat elhagyjuk.

5.1. ÁLLÍTÁS. *Az azonos méretű CP mátrixok halmaza konvex zárt kúp.*

5.2. ÁLLÍTÁS. *Egy CP mátrixhoz nemnegatív mátrixot hozzáadva újra CP mátrixhoz jutunk.*

5.3. ÁLLÍTÁS. *Egy CP mátrix minden fő szubmátrixa szintén CP. (Speciális esetként főátlójának az elemei nemnegatívak.)*

5.4. ÁLLÍTÁS. *Egy CHL mátrixnak minden sorában van negatív elem.*

5.5. ÁLLÍTÁS. *Egy CP (ill. CHL) mátrixból szimmetrikus sor-oszlop permutációval keletkező mátrix szintén CP (ill. CHL).*

5.6. ÁLLÍTÁS. *Szinguláris négyzetes mátrixot az adjungáltjával megszorozva, zéró mátrixot kapunk eredményül.*

5.7. ÁLLÍTÁS. *Szinguláris négyzetes mátrix adjungáltjának a rangja legfeljebb 1 lehet, vagyis ilyenkor az adjungált valamennyi oszlopa arányos egymással.*

5.8. ÁLLÍTÁS. *Tetszőleges $r \times r$ méretű valós négyzetes mátrix és tetszőleges r dimenziós $\mathbf{u}, \mathbf{v}$ vektorok esetén fennáll*

$$|\mathbf{A} + \mathbf{u}\mathbf{v}'| = |\mathbf{A}| + \mathbf{v}'(\mathrm{adj}\,\mathbf{A})\mathbf{u}.$$

A felsorolt tulajdonságok alapján most még a 3.1. állításhoz kapcsolódóan fogalmazunk meg két megjegyzést.

5.1. *Megjegyzés.* $h^* = -\infty$ soha nem állhat fenn, mivel az 5.3. állítás szerint tetszőleges $i \in \mathcal{I}$ index esetén az $\mathbf{A}(-a_{ii} - 1)$ mátrix NCP.

5.2. *Megjegyzés.* Az 5.1. állítás folytán véges h^* esetén az $\mathbf{A}(h^*)$ mátrix CP.

Rátérünk a tételként kimondott állítások felsorolására.

5.1. TÉTEL. (COTTLE-HABETLER-LEMKE tétele). *Tegyük fel, hogy egy $r \times r$ méretű valós szimmetrikus $\mathbf{A}$ mátrix $(r-1)$CP. Ebben az esetben $\mathbf{A}$ akkor és csak akkor NCP, ha teljesülnek a következők:*

$$(i) \quad |\mathbf{A}| < 0;$$
$$(ii) \quad \mathrm{adj}\,\mathbf{A} \geq \mathbf{0}.$$

A bizonyítást mellőzzük, mivel az megtalálható a [2, 3.1. tétel] megfelelő szövegrészében.

5.2. TÉTEL. (KELLER tétele). *A tétel állításához az 5.1. tétel szövegét úgy módosítjuk, hogy a (ii) feltétel sorát a*

(ii') adj$\mathbf{A}$ *első oszlopa nemnegatív* *sorral helyettesítjük.*

Eredeti formájában [2, 4.2. tétel] a tétel úgy szól, hogy *„Egy valós szimmetrikus mátrix akkor és csak akkor* CP, *ha minden olyan fő szubmátrixának a determinánsa negatív, amelyben az utolsó sorhoz tartozó összes ko-faktor nemnegatív."* A 5.5 állításra hivatkozva, az előző mondatban az *„utolsó sor"* szavak helyére „első sor" vagy *„tetszőleges sor"* is helyettesíthető. A mátrix szimmetriája miatt pedig a *„sor"* szó helyére *„oszlop"* szót is írhatunk. Az 5.2. tétel megfogalmazásakor azért választottuk az *első oszlop* vizsgálatát, mert ezzel az algoritmus gépi programjának elkészítése során előjövő kényelmi szempontot veszünk figyelembe.

Itt említjük meg, hogy négyzetes mátrixok kopozitivitására vonatkozó további kritériumokat tárgyal HADELER [3] és VÄLIAHO [8], [9].

5.3. LEMMA. *Legyen* $\mathbf{A}$ *egy* $r \times r$ *méretű valós mátrix,* $\mathcal{I}$, $\mathcal{J}$ *pedig az* $\{1,2,\ldots,r\}$ *halmaz részhalmazai. Jelöljük* $\mathbf{C}$*-vel azt az* $r \times r$ *méretű mátrixot, melynek elemei*

$$c_{ij} = \begin{cases} 0, & \text{ha } i \notin \mathcal{I} \text{ vagy } j \notin \mathcal{J} \\ 1, & \text{ha } i \in \mathcal{I} \text{ és } j \in \mathcal{J}, \end{cases} \tag{5.1}$$

és legyen

$$\mathbf{A}(h) = \mathbf{A} + h\mathbf{C}.$$

Ekkor tetszőleges valós h_1, h_2*-re és tetszőleges* $\Delta h \neq 0$*-ra fennáll*

$$\frac{|\mathbf{A}(h_1 + \Delta h)| - |\mathbf{A}(h_1)|}{\Delta h} = \sum_{i\in\mathcal{I}} \sum_{j\in\mathcal{J}} \mathbf{A}_{ij}(h_2). \tag{5.3}$$

Bizonyítás. Elemi úton bizonyítható, vagy az 5.8. állítás speciális eseteként is megfogalmazható, hogy tetszőleges valós h esetén

$$|\mathbf{A}(h)| = |\mathbf{A} + h\mathbf{C}| = |\mathbf{A}| + h \sum_{i\in\mathcal{I}} \sum_{j\in\mathcal{J}} \mathbf{A}_{ij}. \tag{5.4}$$

Ha most az (5.4) összefüggésben $\mathbf{A}$ helyére az $\mathbf{A}(h_1)$ mátrixot, h helyére pedig Δh-t írjuk, akkor azt kapjuk, hogy

$$|\mathbf{A}(h_1 + \Delta h)| = |\mathbf{A}(h_1)| + \Delta h \sum_{i\in\mathcal{I}} \sum_{j\in\mathcal{J}} \mathbf{A}_{ij}(h_1). \tag{5.5}$$

Ebből $\Delta h \neq 0$ esetén következik, hogy

$$\frac{|\mathbf{A}(h_1 + \Delta h)| - |\mathbf{A}(h_1)|}{\Delta h} = \sum_{i\in\mathcal{I}} \sum_{j\in\mathcal{J}} \mathbf{A}_{ij}(h_1). \tag{5.6}$$

(5.4) szerint $|\mathbf{A}(h)|$ mint h függvénye, lineáris, és h szerinti deriváltja (meredeksége) éppen a

$$\sum_{i\in\mathcal{I}} \sum_{j\in\mathcal{J}} \mathbf{A}_{ij} \tag{5.7}$$

kifejezés. Így az (5.6) jobb oldalán álló összeg értéke nem függ h_1 értékétől, tehát érvényes (5.3) is.

5.4. LEMMA. *Legyen* $\mathbf{A}$ *egy* $r \times r$ *méretű valós szimmetrikus mátrix,* $\mathcal{I}$, $\mathcal{J}$ *pedig az* $\{1,2,\ldots,r\}$ *halmaz azonos részhalmazai. Jelöljük* $\mathbf{A}(h)$*-val az (5.1)–(5.2) szerint előállított parametrizált mátrixot, amely ebben az esetben szintén szimmetrikus. Ekkor a*

$$\mathcal{H} = \{h : \text{ az } \mathbf{A}(h) \text{ mátrix CHL}\} \tag{5.8}$$

halmaz konvex (azaz véges vagy végtelen intervallum).

Bizonyítás. Tekintsük a

$$\mathcal{H}_1 = \{h : \text{ az } \mathbf{A}(h) \text{ mátrix } (r-1)\text{CP}\}, \tag{5.9}$$

és a

$$\mathcal{H}_2 = \{h : \text{ az } \mathbf{A}(h) \text{ mátrix CP}\} \tag{5.10}$$

halmazokat. Ezekre fennáll

$$\mathcal{H} = \mathcal{H}_1 - \mathcal{H}_2, \tag{5.11}$$

a CHL mátrix definíciója szerint. Az 5.1. állításban megfogalmazott tulajdonság következtében $\mathcal{H}_1$ és $\mathcal{H}_2$ nyilvánvalóan intervallumok a számegyenesen, és mindkettő jobbról végtelen. Következésképpen $\mathcal{H}$ is intervallum, tehát a számegyenes konvex részhalmaza.

5.5. TÉTEL. *Tegyük fel, hogy az 5.4. lemmában szereplő parametrizált* $\mathbf{A}(h)$ *mátrixra az (5.8) formulával értelmezett* $\mathcal{H}$ *halmaz nem üres, és* $h_0 = \sup \mathcal{H}$ *véges értékű. Ekkor az* $\mathbf{A}(h_0)$ *mátrixra fennáll a következő három tulajdonság:*

$$\begin{gathered} |\mathbf{A}(h_0)| = 0, \\ \operatorname{adj}\mathbf{A}(h_0) \geq \mathbf{0}, \\ \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{I}} \mathbf{A}_{ij}(h_0) = \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{I}} (\operatorname{adj}\mathbf{A}(h_0))_{ji} \neq 0. \end{gathered} \tag{5.12}$$

Bizonyítás. Jelölje h_1 a $\mathcal{H}$ halmaz egy tetszőleges elemét. Az 5.4. lemma szerint $h_1 \leq h < h_0$ esetén az $\mathbf{A}(h)$ mátrix CHL, ezért az 5.1. tétel szerint az ilyen h értékekre $|\mathbf{A}(h)| < 0$ és $\operatorname{adj}\mathbf{A}(h) \geq \mathbf{0}$. Ebből folytonosság miatt következik, hogy igaz (5.12) második sora, és fennáll

$$|\mathbf{A}(h_0)| \leq 0. \tag{5.13}$$

Mivel $h > h_0$ esetén az $\mathbf{A}(h)$ mátrix $(r-1)$CP, de nem CHL, következésképpen ilyen h-kra az $\mathbf{A}(h)$ mátrix CP. Az 5.1. állítás folytán az $\mathbf{A}(h_0)$ mátrix is CP, ezért az 5.1. tétel szerint $|\mathbf{A}(h_0)| < 0$ nem állhat fenn, és eszerint érvényes (5.12) első sora is.

Az 5.3. lemma szerint fennáll

$$|\mathbf{A}(h_1)| = |\mathbf{A}(h_0)| + (h_1 - h_0)\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{I}} \mathbf{A}_{ij}(h_0). \tag{5.14}$$

Mivel $|\mathbf{A}(h_1)| < 0$, $|\mathbf{A}(h_0)| = 0$ és $h_1 - h_0 < 0$, ezért (5.14)-ből következik, hogy

$$\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{I}} \mathbf{A}_{ij}(h_0) > 0, \tag{5.15}$$

tehát fennáll (5.12) harmadik sora is.

5.6. TÉTEL. *Tegyük fel ismét, hogy az 5.4. lemmában szereplő parametrizált* $\mathbf{A}(h)$ *mátrixra az (5.8) formulával értelmezett* $\mathcal{H}$ *halmaz nem üres. Ekkor tetszőleges valós* h *esetén igazak a következők:*

(i) $\sup\mathcal{H}$ *akkor és csak akkor véges értékű, ha*

$$|\mathbf{A}(h+1)| - |\mathbf{A}(h)| > 0; \tag{5.16}$$

(ii) $\sup\mathcal{H} = +\infty$ *akkor és csak akkor, ha*

$$|\mathbf{A}(h+1)| - |\mathbf{A}(h)| = 0; \tag{5.17}$$

(iii) ha $\sup\mathcal{H}$ *értéke véges, akkor*

$$\sup\mathcal{H} = h - \frac{|\mathbf{A}(h)|}{|\mathbf{A}(h+1)| - |\mathbf{A}(h)|}. \tag{5.18}$$

Bizonyítás. Az 5.3. lemma szerint a

$$\Phi = |\mathbf{A}(h+1)| - |\mathbf{A}(h)| \tag{5.19}$$

kifejezés értéke nem függ h-tól. Az 5.1. tétel szerint $\mathrm{adj}\mathbf{A}(\hat{h}) \geq \mathbf{0}$ tetszőleges $\hat{h} \in \mathcal{H}$ esetén, következésképpen

$$\Phi = |\mathbf{A}(h+1)| - |\mathbf{A}(h)| = \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{I}} \mathbf{A}_{ij}(\hat{h}) \geq 0. \tag{5.20}$$

Az (5.20) egyenlőtlenség és a tétel feltevéséből következően fennálló $\sup\mathcal{H} > -\infty$ egyenlőtlenség miatt elég az (i) és (ii) állításoknak a „csak akkor" részét bizonyítani.

(i) Tegyük fel, hogy $h_0 = \sup\mathcal{H}$ véges értékű. Az 5.3. lemma szerint tetszőleges $\hat{h} \in \mathcal{H}$ értékre

$$\frac{|\mathbf{A}(h_0)| - |\mathbf{A}(\hat{h})|}{h_0 - \hat{h}} = \Phi. \tag{5.21}$$

Az 5.1. tétel szerint $|\mathbf{A}(\hat{h})| < 0$, az 5.5. tétel szerint pedig $|\mathbf{A}(h_0)| = 0$. Nyilvánvalóan $h_0 - \hat{h} > 0$, ennélfogva (5.21) bal oldala pozitív, tehát fennáll (5.16).

(ii) Tegyük fel, hogy $\sup \mathcal{H} = +\infty$, és megint válasszunk $\mathcal{H}$-ból egy tetszőleges $\hat{h}$ elemet. Ekkor az 5.3. lemma szerint

$$|\mathbf{A}(h)| = |\mathbf{A}(\hat{h})| + (h - \hat{h})\Phi \tag{5.22}$$

tetszőleges valós h-ra. Az 5.1. tétel alapján tudjuk, hogy $h \geq \hat{h}$ esetén $|\mathbf{A}(h)| < 0$, ezért

$$(h - \hat{h})\Phi < -|\mathbf{A}(\hat{h})| \tag{5.23}$$

minden $\hat{h}$-nál nagyobb h-ra. Ez csak úgy lehetséges, ha fennáll $\Phi \leq 0$, de ez (5.20) miatt csak egyenlőséggel teljesülhet.

(iii) Az 5.3. lemma szerint

$$0 = |\mathbf{A}(h_0)| = |\mathbf{A}(h)| + (h_0 - h)\Phi, \tag{5.24}$$

vagyis átrendezéssel

$$h_0 = h - \frac{|\mathbf{A}(h)|}{\Phi}. \tag{5.25}$$

Ezzel éppen az (5.18) formulát kaptuk más jelölésekkel.

6. A véges algoritmus bizonyítása

Az ismertetett algoritmus végessége következik abból, hogy az A mátrix összes fő szubmátrixát legfeljebb egyszer vizsgáljuk.

Megmutatjuk, hogy minden olyan k érték esetén, melyre az algoritmus ciklusát végrehajtjuk, és nem jelentkezik a „probléma nem korlátos" észrevétel, $j = 1, 2, \ldots, k$-ra igaz, hogy a $\mathbf{B}_j(h_{new})$ mátrix CP. Mindig h_{new}-nak a k-adik lépésben kapott értékével érvényes az állítás. $k = N$ esetén ez az állítás épp azt jelenti, hogy az $\mathbf{A}(h_{new})$ mátrix CP. (Feltéve, hogy az algoritmus a $k > N$ leállási feltétellel fejeződik be.)

Az állítást indukcióval bizonyítjuk. Legyen először $k = 1$. Ha a $\mathbf{B}_1(h_{old})$ mátrix nem CHL, akkor ebben az iterációban h_{new} értéke nem változik meg. A $\mathbf{B}_1(h_{new})$ mátrix nem CHL és rendje 1, következésképpen e mátrix CP. Ha viszont a $\mathbf{B}_1(h_{old})$ mátrix CHL, és h_{new} értéke az algoritmusban megadott formula szerint módosul, akkor az 5.6 tétel szerint

$$h_{new} = \sup\{h : \text{ a } \mathbf{B}_1(h) \text{ mátrix CHL}\}, \tag{6.1}$$

tehát az 5.1. állítás folytán a $\mathbf{B}_1(h_{new})$ mátrix CP.

Az indukciós feltevést úgy fogalmazhatjuk meg, hogy a $\mathbf{B}_j(h_{old})$ mátrix kopozitív $j = 1, 2, \ldots, k-1$ esetén. Ebből arra kell tudnunk következtetni, hogy a $\mathbf{B}_j(h_{new})$ mátrix kopozitív $j = 1, 2, \ldots, k$ esetén. Megint két esetet kell végiggondolnunk, attól függően, hogy a $\mathbf{B}_k(h_{old})$ mátrix nem CHL vagy CHL. Ha a $\mathbf{B}_k(h_{old})$ mátrix nem CHL, akkor a k-adik iterációban h_{new} értéke nem változik meg, tehát a $\mathbf{B}_j(h_{new})$ mátrix kopozitív $j = 1, 2, \ldots, k-1$ esetén. Ez azt is jelenti, hogy a $\mathbf{B}_k(h_{new})$ mátrix minden nála kisebb méretű fő szubmátrixa CP. Mivel még azt is tudjuk, hogy a $\mathbf{B}_k(h_{new})$ mátrix nem CHL, mindebből következik, hogy ez a mátrix CP. Most nézzük azt az esetet, amikor a $\mathbf{B}_k(h_{old})$ mátrix CHL, és h_{new} értéke az algoritmusban adott formula szerint módosul. A $k = 1$ esetben már alkalmazott gondolatmenettel, az 5.6. tételre hivatkozva írhatjuk, hogy

$$h_{new} = \sup\{h : \text{ a } \mathbf{B}_k(h) \text{ mátrix CHL}\}, \tag{6.2}$$

tehát a $\mathbf{B}_k(h_{new})$ mátrix CP. Mivel h_{old} eleme a (6.2)-nél szereplő halmaznak, ennélfogva $h_{old} < h_{new}$. Ezt figyelembe véve, továbbá az 5.2. állításra is hivatkozva, az indukciós feltevésből következik, hogy a $\mathbf{B}_j(h_{new})$ mátrix is kopozitív $j = 1, 2, \ldots, k-1$ esetén.

Folytatjuk az algoritmus helyességének a bizonyítását. Ha az eljárás úgy ér véget, hogy valamely k esetén a $\mathbf{B}_k(h_{old})$ mátrix CHL és fennáll

$$|\mathbf{B}_k(h_{old})| = |\mathbf{B}_k(h_{old} + 1)|,$$

akkor az 5.6. tétel (ii) állítása következtében a $\mathbf{B}_k(h)$ mátrix CHL minden $h \geq h_{old}$ esetén, ennélfogva az $\mathbf{A}(h)$ mátrix semmilyen valós h esetén nem lehet CP. Ilyen esetben $h^* = +\infty$, tehát a probléma valóban nem korlátos.

Ha a $\mathbf{B}_k(h_{old})$ mátrix CHL és $|\mathbf{B}_k(h_{old})| \neq |\mathbf{B}_k(h_{old} + 1)|$, akkor be kell még látnunk, hogy az $\mathrm{adj}\mathbf{B}_k(h_{new})$ mátrixnak van zéró vektortól különböző oszlopvektora. Ez valóban így van, ugyanis az 5.5. tétel szerint a $\mathbf{B}_k(h_{new})$ mátrixra fennáll

$$\sum_{i \in \mathcal{I}^*} \sum_{j \in \mathcal{I}^*} (\mathrm{adj}\mathbf{B}_k(h_{new}))_{ji} \neq 0, \tag{6.3}$$

ahol $\mathcal{I}^*$ a $\mathbf{B}_k$ mátrix azon sorainak az indexhalmazát jelöli, amelyek az $\mathbf{A}$ mátrixban $\mathcal{I}$-hez tartoznak.

A bizonyítás további részét csak a $k > N$ leállási feltétel alkalmazása esetére részletezzük. Ha az eljárás e leállási feltétellel fejeződik be, akkor az algoritmus során h_{new} értékének legalább egy alkalommal változnia kellett. A kezdő h_{new} értékkel ugyanis az $\mathbf{A}(h_{new})$ mátrix NCP, az 5.3. állítás következtében. Ezért a h_{new} kezdőértékével parametrizált $\mathbf{B}_k(h_{new})$ szubmátrixok között kell lenni legalább egy CHL mátrixnak, más szóval olyan eset nem következhet be, hogy a CHL_test függvény az algoritmus lefutása során végig igaz értéket vegyen fel. Ebből következik, hogy az elvégzett iterációk valamelyike (mondjuk a k^*-adik iteráció) során fennáll

$$h_{old} < h_{new},$$

viszont a k^*-nál későbbi iterációk során h_{new} értéke már nem változik. Az algoritmus menetéből és a bizonyítás eddigi részéből látható, hogy az $\mathbf{A}(h)$ mátrix NCP, ha $h_{old} \leq h < h_{new}$, de az $\mathbf{A}(h_{new})$ mátrix már CP. Az algoritmusnak a $k > N$ leállási feltétellel történő befejezésekor tehát h_{new} valóban a (3.8) szerint értelmezett h^* értéket veszi fel.

Az utolsóként generált $\mathbf{v}_{new}$ vektor a k^* iterációhoz tartozó $\mathbf{B}_{k^*}(h_{new})$ szubmátrixból származik, oly módon, hogy az $\mathrm{adj}\mathbf{B}_{k^*}(h_{new})$ mátrix valamely nemzéró $\mathbf{u}$ oszlopvektorának $\mathbf{v}$ kiterjesztését lenormáljuk az algoritmus folyamatleírásában megadott módon. Az 5.5. tétel szerint a $\mathbf{B}_{k^*}(h_{new})$ mátrixra fennáll

$$\begin{gathered} |\mathbf{B}_{k^*}(h_{new})| = 0, \\ \mathrm{adj}\mathbf{B}_{k^*}(h_{new}) \geq \mathbf{0}, \\ \sum_{i\in\mathcal{I}^*}\sum_{j\in\mathcal{I}^*}(\mathrm{adj}\mathbf{B}_{k^*}(h_{new}))_{ji} \neq 0, \end{gathered} \tag{6.4}$$

ahol $\mathcal{I}^*$ értelmezése hasonló, mint kicsit feljebb, a (6.3) formulánál. A (6.4) tulajdonságokból az 5.7. állítás figyelembe vételével következik, hogy az $\mathrm{adj}\mathbf{B}_{k^*}(h_{new})$ mátrix rangja 1. Ennek alapján (6.4)-ből az is következik, hogy az $\mathrm{adj}\mathbf{B}_{k^*}(h_{new})$ mátrix bármely nemzéró $\mathbf{u}$ oszlopvektorára fennáll

$$\begin{gathered} \mathbf{u} \geq \mathbf{0}, \\ \sum_{i\in\mathcal{I}^*} u_i \neq 0, \\ \mathbf{u}'\mathbf{B}_{k^*}(h_{new})\mathbf{u} = 0. \end{gathered} \tag{6.5}$$

(Az 5.6. állítás alapján adódik (6.5) utolsó sora.)

Az $\mathbf{u}$ vektorra felírt (6.5) tulajdonságokból következik, hogy az $\mathbf{u}$ vektor r dimenzióssá történő kiterjesztése, majd a kiterjesztett vektor alkalmas normálása után kapott $\mathbf{v}_{new}$ vektor kielégíti a (3.10) feltételeket, a $h^* = h_{new}$ érték esetén. (A (6.5) képletcsoport középső sora biztosítja, hogy a normálás elvégzésekor zérótól különböző értékkel osztjuk le a $\mathbf{v}$ vektor komponenseit.) Az algoritmus eredményeként kapott $(h_{new}, \mathbf{v}_{new})$ pár tehát valóban a leszármaztatott probléma megoldása.

A CHL_test logikai függvény értéke kiszámításának gyorsítására a KELLER tételén alapuló vizsgálatot megelőzően ellenőrizzük, hogy a vizsgált szubmátrix sorai között nincs-e csupa nemnegatív elemet tartalmazó sor. Abban az esetben ugyanis az 5.4 állítás szerint ez a mátrix nem lehet CHL, tehát fölösleges a determináns és az adjungált kiszámítása.

Egy-egy mondattal vázoljuk az algoritmus leírása után felsorolt különböző leállási feltételek indoklását. Az 5.1. tétel szerint egy CHL mátrix nem lehet szinguláris, ebből kiindulva bizonyítható az $\mathrm{ordo}\mathbf{B}_k > \mathrm{rank}\mathbf{A}(h_{new})$ leállási feltétel helyessége. Az ezt követő $\mathrm{ordo}\mathbf{B}_k > \mathrm{negordo}\mathbf{A}(h_{new})$ fennállása esetén történő leállást az 5.4. állításban megfogalmazott tulajdonság teszi lehetővé. A két utóbbi gondolat kombinálása eredményezi az ordo$\mathbf{B}_k >$ negrank$\mathbf{A}(h_{new})$ leállási feltételt. Végül az

$$\mathbf{A}(h) = \begin{bmatrix} \mathbf{Q}'\mathbf{D} \\ \mathbf{S}'\mathbf{D} + \mathbf{1g}' \end{bmatrix} [\mathbf{Q} \quad \mathbf{S}] + \begin{bmatrix} \mathbf{Q}'\mathbf{g} \\ \mathbf{S}'\mathbf{g} + h\mathbf{1} \end{bmatrix} [\mathbf{0}' \quad \mathbf{1}']$$

alakú felírásból látható, hogy az $\mathbf{A}(h)$ mátrixra tetszőleges valós h esetén fennáll

$$\text{rank}\mathbf{A}(h) \le \text{ordo}\mathbf{D} + 1,$$

és

$$\text{rank}\mathbf{A}(h) \le \text{rank}\mathbf{D} + 2.$$

Ez a két egyenlőtlenség teszi lehetővé az utolsóként felsorolt két leállási feltétel alkalmazását.

A két utolsó leállási feletételtől eltekintve, a bizonyítás során sehol nem használtuk ki az $\mathbf{A}(h)$ mátrix speciális szerkezetét, amely abból adódott, hogy ezt a mátrixot a (2.2) QP szubprobléma adataiból állítottuk elő. Az algoritmus érvényes tehát általánosabb körülmények között is, az 5.4. lemma értelmezése szerinti $\mathbf{A}(h)$ mátrixra, azzal az egy további kikötéssel, hogy az $\mathcal{I} = \mathcal{J}$ halmaz ne legyen üres.

Végül megemlíthetjük, hogy az algoritmus egy új, konstruktív bizonyítást szolgáltat arra a kvadratikus programozási egzisztenciatételre, mely szerint *„ha egy* QP *probléma célfüggvénye alulról korlátos a feltételi poliedrikus halmazon, akkor fel is veszi ott a minimumát"*.

7. Számítástechnikai tapasztalatok

A 4. szakaszban ismertetett algoritmus számítógépes programját FORTRAN és PASCAL programnyelvi változatokban készítettük el és PC AT gépen próbáltuk ki néhány, viszonylag kisebb méretű QP probléma megoldására. Az algoritmust egyelőre olyan esetekre alkalmaztuk, amikor a (2.2) szubprobléma feltételi halmaza azonos az eredeti (2.1) probléma feltételi halmazával, és így a szubprobléma megoldásával megkapjuk az eredeti probléma globális optimumát is. Ehhez azonban először fel kell tudnunk írni a (2.1) QP probléma feltételi halmazának a *Motzkin-féle felbontási tételen* alapuló belső reprezentációját, vagyis elő kell állítanunk a feltételi halmaz extremális pontjait és extremális irányait.

Az extremális pontok és extremális irányok meghatározását egy külön programmal végezzük. Az előbbiekhez MANAŠ–NEDOMA [7] cikkében ismertetett algoritmust alkalmazzuk. Ugyanezt az algoritmust alkalmazhatjuk az extremális irányok meghatározására is, ha az eredeti feltételrendszerben a jobb oldalakat 0-ra módosítjuk, és az így kapott feltételrendszert kiegészítjük a $\sum x_i = 1$ feltétellel.

Az algoritmusunkkal megoldott QP problémák célfüggvénye — egy kivétellel — se nem konvex, se nem konkáv. Az egyes problémák megnevezése és eredete:
B jelű probléma: BALAS [1] cikkében tárgyalt mintafeladat:

$$\min \ -0,66667x_1^2 + 0,5x_2^2 + x_1 - 0,5x_2$$
$$\text{feltéve, hogy } 2x_1 - x_2 \le 3,$$
$$x_1, x_2 \ge 0.$$

M jelű probléma: MAJTHAY et al. [6] cikkében tárgyalt mintafeladat:

$$\begin{aligned} \max\ & 0,5x_1^2 - 0,5x_2^2 + x_1 + 2x_2 \\ \text{feltéve, hogy}\ & x_1 + x_2 \le 4, \\ & 2x_1 - x_2 \le 5, \\ & x_1, x_2 \ge 0. \end{aligned}$$

K1, K2 jelű problémák: KOUGH [5] cikkében szereplő első két mintafeladat:

$$\begin{aligned} \max\quad & x_1^2 + x_2^2 - y_1^2 - y_2^2 \\ \text{feltéve, hogy}\quad & \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{y} + \mathbf{C} \ge \mathbf{0}, \\ & \mathbf{x}, \mathbf{y} \ge \mathbf{0}, \end{aligned}$$

ahol $\mathbf{A}, \mathbf{B}$ és $\mathbf{C}$ adott 5 soros mátrixok a K1 probléma esetén, de ugyanezek 8 sorosak a K2 probléma esetén.

K3 jelű probléma: KOUGH [5] harmadik mintafeladata: Ennek feltételrendszere azonos a K2 jelű probléma feltételrendszerével, célfüggvénye azonban

$$\max\quad x_1^2 + x_2^2 - y_1^2 - y_2^2 - 50x_1 - 100x_2.$$

B+M jelű probléma: A B és M jelű problémák egyesítéséből adódó QP probléma, melynek feltételhalmaza a B és M jelű problémák feltételhalmazának *Descartes szorzata*.

B+B jelű probléma: Két azonos alakú probléma formális összetétele. A feltételhalmaz a B jelű probléma feltételhalmazának saját magával képzett *Descartes szorzata*.

P1, P2, P3 jelű problémák: A feltételi poliéder belső reprezentációjával megadott problémák: Feltételrendszerük a változók nemnegativitásának kikötésén kívül csak egy további feltételt tartalmaz, amely a P1 és P2 jelű problémák esetén $\sum x_i \le 1$ alakú, a P3 jelű probléma esetén pedig $\sum x_i = 1$ alakú.

P4, P5 jelű problémák: A szerző által kitűzött két további mintafeladat.

A számítógépen lefuttatott mintafeladatok főbb jellemzőit és a futási időket a következő táblázat tartalmazza:

probl. jele	min vagy max	változók száma	$\leq, \geq$ felt. száma	= felt. száma	extr. ir. száma	extr. pontok száma	futási idő (mp)	ebből Q, S felépítése
B	min	2	1	0	2	2	4	3
M	max	2	2	0	0	4	3	3
K1	max	4	5	0	8	6	18	4
K2	max	4	8	0	0	28	10	6
K3	max	4	8	0	0	28	9	6
B+M	min	4	3	0	2	8	33	3
B+B	min	4	2	0	4	4	16	3
P1	min	6	1	0	0	7	13	-
P2	min	6	1	0	0	7	7	-
P3	min	8	0	1	0	8	70	-
P4	min	5	6	0	1	8	8	5
P5	max	8	6	0	0	85	110	16

Néhány lefuttatott problémára bemutatjuk az algoritmus végrahajtása során talált megengedett megoldások és a hozzájuk tartozó célfüggvényértékek sorozatát. Az utolsó megoldás mindegyik probléma esetén az optimális.

B:

x_1	x_2	célf.
0,0	0,0	0,0
1,5	0,0	-0,00001
0,0	0,5	-0,12500
1,5	0,5	-0,12501
2,25	1,5	-0,75002

M:

x_1	x_2	célf.
0,0	0,0	0,0
2,5	0,0	5,625
3,0	1,0	9,0

K1:

x_1	x_2	y_1	y_2	célf.
1,9953	0,9965	0,9953	0,9953	2,99301
4,9847	0,0	3,9861	0,4973	8,71076
4,9948	0,0	3,9978	0,4996	8,71619
4,9974	0,0	3,9988	0,4997	8,73366
6,1248	0,8556	5,1265	1,6381	9,28082
6,1319	0,8514	5,1336	1,6350	9,29814
12,9882	1,1864	11,9926	3,3149	15,28990
12,9924	1,1852	11,9968	3,3144	15,29889

K2:

x_1	x_2	y_1	y_2	célf.
16,1257	15,5901	13,1200	14,9099	108,6520
16,2149	15,5380	13,1118	14,8599	111,6169
16,8117	15,2646	13,1900	14,3474	135,8190
17,4622	14,7453	13,5071	14,0164	143,4512

K3:

x_1	x_2	y_1	y_2	célf.
16,1257	15,5901	13,1200	14,9099	-2256,646
16,2149	15,5380	13,1118	14,8599	-2252,927
11,9839	0,0	29,2130	15,5627	-1551,177
16,5911	0,0	27,8055	11,9603	-1470,485

IRODALOM

[1] BALAS, E., "Nonconvex quadratic programming via generalized polars ", *SIAM Journal on Applied Mathematics* **28** (1975), 335–339.

[2] COTTLE, R. W., HABETLER, G.J., and LEMKE, C. E., "On classes of copositive matrices ", *Linear Algebra and its Applications* **3** (1970), 295–310.

[3] HADELER, K. P., "On copositive matrices ", *Linear Algebra and its Applications* **49** (1983), 79–89.

[4] KÉRI, G., "On a class of quadratic forms ", in A. Prékopa (ed.), Survey of mathematical programming, Proceedings of the 9th International Mathematical Programming Symposium (North-Holland Publishing Company, 1979), pp. 231–247.

[5] KOUGH, P. F., "The indefinite quadratic programming problem ", *Operatins Research* **27** (1979), 516–533.

[6] MAJTHAY, A., WHINSTON, A., and COFFMAN, J., "Local optimization for nonconvex quadratic programming ", *Naval Research Logistics Quarterly* **21** (1974), 465–490.

[7] MANAŠ, M. and NEDOMA, J., "Finding all vertices of a convex polyhedron ", *Numerische Mathematik* **12** (1968), 226–229.

[8] VÄLIAHO, H., "Criteria for Copositive Matrices ", *Linear Algebra and its Applications* **81** (1986), 19–34.

[9] VÄLIAHO, H., "Testing the Definiteness of Matrices on Polyhedral Cones ", *Linear Algebra and its Applications* **101** (1988), 135–165.

(Beérkezett: 1989. február 8.)

KÉRI GERZSON
ÁLLAMIGAZGATÁSI SZÁMÍTÓGÉPES SZOLGÁLAT VÁLLALAT
1119 BUDAPEST, ANDOR U. 47–49.

METHODS OF NONCONVEX QUADRATIC OPTIMIZATION BASED ON THE MATRIX COPOSITIVITY INVESTIGATION

G. KÉRI

In the paper methods for the solution of general quadratic programming problems are investigated. These methods are based on copositivity investigations of parametrized matrices. There is given a finite algorithm for the solution of an optimization problem related to quadratic programming. The algorithm is based on theoretical results according to some properties of copositive and noncopositive matrices. There are also given some computational experiences.

Alkalmazott Matematikai Lapok 15 (1990–91) 133–162

A KARMARKAR TÍPUSÚ ALGORITMUSOKRÓL

TERLAKY TAMÁS*

Budapest

Dolgozatunkban a lineáris programozás belső pontos algoritmusaival foglalkozunk. A belső pontos algoritmusok KARMARKAR 1984-ben megjelent dolgozata óta az érdeklődés középpontjában állnak mind elméleti mind implementációs szempontból. Cikkünkben a belső pontos algoritmusok egy rövid áttekintését adjuk.

A *Karmarkar típusú algoritmusok* három fő típusát különböztetjük meg: a projektív, potenciálfüggvényes változatot (KARMARKAR), az affin– (DIKIN-BARNES) és a trajektória követő algoritmusokat. Mind a három típus egy–egy egyszerű, könnyen érthető változatát közöljük. Ismertetjük az induló megoldást kereső (első fázis) feladatot, valamint az algoritmusok során használatos segédfeladatok megoldását.

1. Bevezetés

A lineáris programozás mindmáig legismertebb, legelterjedtebb megoldási módszere a szimplex módszer, amely DANTZIG-tól [3] származik. Már a kezdeti időktől vizsgálták a szimplex módszer elméleti és gyakorlati hatékonyságát. A szimplex módszer a gyakorlatban nagyon hatékonynak bizonyult, megbízható, gyors implementációt tesz lehetővé. A módszer elméleti hatékonysága sokáig nyitott, tisztázatlan kérdés volt. Először KLEE és MINTY [11] adott példát arra, hogy a legrosszabb esetben a szimplex módszer a dimenzió függvényében exponenciálisan sokat léphet. Így nagy szakadék mutatkozott a szimplex módszer elméleti és gyakorlati („átlagos") viselkedése között. Ezt a szakadékot hidalta át (többek között) SMALE [21], aki bizonyította, hogy – alkalmas feltételek mellett – a szimplex módszer várható lépésszáma polinomiális.

Nem kívánunk itt kitérni a szimplex módszerrel, a lineáris programozással kapcsolatos kiterjedt irodalom áttekintésére, az új pivot szabályok ismertetésére, csak a *Karmarkar algoritmussal* kapcsolatos problémákat, eredményeket ismertetjük.

Így meg kell említenünk, hogy a fenti eredmények, problémák mellett sokáig nyitott kérdés maradt, hogy létezik–e polinomiális algoritmus a lineáris programozási feladat megoldására. A problémát KHACSIJÁN oldotta meg, az ellipszoid módszer megalkotásával. Az ellipszoid módszer egy változata található KLAFSZKY és TERLAKY [10] cikkében, illetve SCHRIJVER [20] könyvében. Az ellipszoid módszer, illetve az a tény, hogy létezik polinomiális algoritmus a lineáris programozási feladatok megoldására, nagy izgalmat váltott ki, elsősorban a kombinatórikus optimalizálás területén. Sok feladatosztályról sikerült belátni, hogy megoldására létezik polinomiális algoritmus. Az ezzel kapcsolatos eredmények egy lényegében teljes

*A kutatás az OTKA 1044.sz. és az EGPO 59/86 pályázatok részleges támogatásával történt.

áttekintése található SCHRIJVER [20] könyvében. Az ellipszoid módszer nagyon hamar újabb problémákat vetett fel. A módszer elméleti hatékonysága számtalan implementációt indukált. Azonban nagyon hamar kiderült, hogy az ellipszoid módszer elméleti hatékonysága ellenére gyakorlatilag használhatatlan, már kisméretű feladatok esetén sem adott kielégítő megoldást. Ez azt a problémát vetette fel, hogy van–e olyan polinomiális algoritmus, amely elméletileg és gyakorlatilag is hatékonyan oldja meg a lineáris programozási feladatokat.

A fenti probléma megoldásának útján (ha a megoldás kérdése még nem is dőlt el véglegesen) mindenképpen nagyon jelentős lépés KARMARKAR [8] algoritmusa. Míg a szimplex módszer a megengedett tartomány (poliéder) csúcsain, a *Khacsiján algoritmus* a poliéderen kívül iterálva, a poliédert tartalmazó csökkenő térfogatú ellipszoid sorozatot definiálva oldja meg a lineáris programozási feladatokat, addig a *Karmarkar féle projektív algoritmus* belső pontokon, a belső pontok köré gömböket (ellipszoidokat) írva közelíti a feladat optimális megoldását.

KARMARKAR cikke óriási pezsgést indított el a matematikai programozási társadalomban. Különösen KARMARKAR azon többször is elhangzott állítása váltott ki heves vitát, amely szerint algoritmusa jelentékenyen (50–szer, 100–szor, sőt nagy méretekben ennél is többször) hatékonyabb, mint a szimplex módszeren alapuló, kereskedelmi forgalomban is kapható LP programcsomagok. Habár ez még nem megnyugtatóan bizonyított, annyi azért általánosan elfogadott, hogy lényegesen hatékonyabban implementálható, mint az ellipszoid módszer, és legalábbis a szimplex módszerrel is összehasonlítható eredményeket ad. Általában meglepően kevés (az elméleti korlátnál sokkal kevesebb, lényegében konstans számú) iteráció szükséges a feladat megoldásához, azonban az iterációk nagyon költségesek, számítógépigényesek. Nem célunk itt az implementációs kérdések, eredmények áttekintése, csak néhány hivatkozást említünk az érdeklődő olvasó könnyebb tájékozódása céljából. KARMARKAR [8] cikkében már hatékony implementációra is utal.TOMLIN [25] negatív tapasztalatait közli. ROOS [17, 18, 19] is közöl tapasztalati eredményeket, GILL és társai [5, 6] számos implementációs kérdést tárgyalnak. Ezen szerzők eredményei azonban meg sem közelítik KARMARKAR fent említett állításait, melyeket [9] cikkében megerősít, sőt fokoz.

A *Karmarkar algoritmussal* kapcsolatos belső pontos eljárások irodalmának teljes áttekintése szinte lehetetlen feladat a cikkek nagy száma, sokfélesége, valamint az újabb és újabb cikkek megjelenése miatt. Itt csak a fő irányzatok tömör áttekintését kíséreljük meg.

A belső pontos algoritmusok három fő osztályát különböztetjük meg (természetesen ezek közti átmenetek is előfordulnak), ennek megfelelően cikkünkben a három fő algoritmusirányt mutatjuk be a további fejezetekben.

Az első csoportba a „projektív–potenciálfüggvényes” változatok tartoznak. Ide tartozik KARMARKAR [8] eredeti cikke, ANSTREICHER [1], IRI és IMAI [7], KOJIMA [12], RENEGAR [16] és ROOS [17] algoritmusa.

A második csoport az „affin”, illetve „affin skálázási” algoritmusok. Ide BARNES [2] és DIKIN [4] algoritmusai tartoznak. Itt meg kell említenünk, hogy DIKIN és BARNES algoritmusa azonos, csak az az érdemi eltérés, hogy BARNES cikke sokkal részletesebb konvergencia analízist tartalmaz. Igy BARNES mintegy húsz évvel később újra felfedezte DIKIN algoritmusát, ami azt is mutatja, hogy a belső pontos megközelítés a lineáris programozásban sem teljesen új, csak újabb technikák, komplexitást vizsgáló eszközök birtokában új megvilágításba került, és újabb eredményekre vezetett.

A harmadik típusba a „trajektória követő" algoritmusok tartoznak. Itt tulajdonképpen egy paraméterezett feladatsereg „(analitikus) centrumai" által leírt görbét közelítik az iterációs pontok. Ebbe az osztályba tartoznak ROOS [18, 19], SONNEVEND és STOER [22], YE és TODD [27] algoritmusai.

Mindhárom algoritmusosztályba, mint említettük, számos algoritmus, algoritmusváltozat tartozik, melyek tárgyalásától terjedelmi okokból eltekintünk.

Meg kell említenünk azon eredményeket is, melyek a belső pontos algoritmusok szimplex módszerhez való viszonyát vizsgálják (TODD [24] és MEGIDDO [14]), illetve azt, hogyan kapható egy optimális bázismegoldás a belsőpontos algoritmusok eredményéből (MEGIDDO [13, 15]).

Szoros kapcsolat fedezhető fel a külső pontos (ellipszoid) módszerek és a belső pontos módszerek között, melyet többek között YE [26] illetve YE és TODD [27] vizsgált.

A lineáris programozásnál bővebb feladatosztályokra is kiterjesztették (kiterjesztik) a belső pontos megközelítést. Igy YE [28] konvex kvadratikus programozásra, KOJIMA [12] lineáris komplementaritási feladatokra, míg SONNEVEND és STOER [22] konvex kvadratikus feltételes kvadratikus programozási feladatokról bizonyította, hogy van polinomiális algoritmus megoldásukra.

Végül az irodalom rövid áttekintésének zárásaként megemlítjük, hogy – egyelőre polinomialitás bizonyítása nélkül – megkezdődött a fenti algoritmusok alkalmazása bővebb feladatosztályokra, úgymint nemkonvex programozási feladatokra és kombinatórikus optimalizálási feladatokra. Sajnos ebből a témakörből még nem áll rendelkezésre publikáció (csak konferencia előadások hangzottak el).

Cikkünkben KARMARKAR [8] „projektív" algoritmusának egy olyan változatát közöljük, mely ROOS [17] interpretációjához áll legközelebb. Az „affin" változatot BARNES [2] cikke alapján, míg a „trajektória követő" algoritmust ROOS és VIAL [19] dolgozata alapján ismertetjük.

Függelékekben foglaltuk össze az algoritmusok megértéséhez, felépítéséhez szükséges előismereteket, segédfeladatokat és azok megoldását, a belső pontot előállító „első fázis" feladatát, a szükséges transzformációkat. A függelékben közöltek nagy része a [8, 10, 17, 20] publikációkban található. A témakörben tájékozatlan olvasónak célszerű a Függelék tanulmányozásával kezdeni.

Jelöléseinkről megjegyezzük, hogy a mátrixokat latin nagybetűkkel, a vektorokat latin kisbetűkkel, a skalárokat és a vektorok illetve mátrixok koordinátáit görög betűkkel jelöljük. A ^{T} felső index jelenti a mátrix transzponáltját. Általában a $\min\{cx : Ax = b,\, x \geq 0\}$ lineáris programozási feladatot tekintjük, ahol az általánosság megszorítása nélkül feltehető, hogy $A : m \times n$ méretű mátrix, $c, x \in \mathbf{R}^n$,

$b \in \mathbf{R}^m$, rang $(A) = m$, és $1 \leq m < n$. (A redundáns feltételek kiszűrése *Gauss-eliminációval* – nyilván polinomiális algoritmus – elvégezhető.). $\mathbf{R}^n_+$ a szigorúan pozitív, míg $\mathbf{R}^n_\oplus$ a nemnegatív vektorok halmazát jelöli. Az alábbiakban $S^n = \{x : \mathbf{1}x = n,\, x \geq 0\}$ jelöli az n-dimenziós tér $(0, \dots, 0, n, 0, \dots, 0)$ pontjai által definiált szimplexet, ahol $\mathbf{1} = (1, \dots, 1)$.

Feltételezzük, hogy a paraméterek egész értékűek, ami a paraméterek racionalitásával egyenértékű.

Cikkünk fő célja a *Karmarkar algoritmus* fő variánsainak és leglényegesebb részeinek bemutatása ismeretterjesztő céllal. A numerikus, számítógépes implementáció problémáit, eredményeit egy másik cikk fogja tárgyalni, melyet MAROS ISTVÁN ír. Így remélünk egy lényegében teljes áttekintést adni erről az új területről.

2. Karmarkar projektív algoritmusa

Ebben a fejezetben az eredeti KARMARKAR [8] algoritmust ismertetjük. Tárgyalásunk légyegében a projektív algoritmus ROOS [17] féle interpretációját követi, ahhoz áll a legközelebb. Így tárgyalásmódunk célja nem a leghatékonyabb változat közlése (ez még nem is eldöntött kérdés), hanem az algoritmus fő lépéseinek, lényeges pontjainak lehető legegyszerűbb bemutatása.

Tekintsük az általános

$$\text{(2.1)} \qquad \text{Primál} \quad \begin{aligned} &\min cx \\ &Ax \geq b \\ &x \geq 0 \end{aligned} \qquad \text{Duál} \quad \begin{aligned} &\max yb \\ &A^T y \leq c \\ &y \geq 0 \end{aligned}$$

lineáris programozási feladatpárt. Mint ahogy F.5. és F.6. függelékekben is ismertetjük, (2.1) feladatpár megoldása ekvivalens a

$$\text{(2.2)} \qquad \begin{aligned} \min e_1 x& \\ Ax &= 0 \\ \mathbf{1}x &= n \\ x &\geq 0 \end{aligned}$$

feladat megoldásával, ahol az A mátrix teljes rangú, az $\mathbf{1} = (1, \dots, 1)$ pont pozitív induló megoldása a (2.2) feladatnak, és a célfüggvény optimális értéke zérus. Ehhez csak azt kellett feltételeznünk, hogy a (2.1) feladatnak van optimális megoldása, és a primál optimális halmaz korlátos. Ezek a feltételek tetszőleges polinomiális algoritmussal, polinomiális időben ellenőrizhetőek.

Segédfeladat

Legyen $a > 0$ a (2.2) feladat egy megengedett megoldása. Ennél szeretnénk egy jobb megoldást kapni. Az F.3. függelék szerint affin halmaz és ellipszoid (gömb) metszetén (a középpontból indulva) meghatározható egy lineáris célfüggvény minimuma. Ahhoz, hogy az ott leírtakat alkalmazhassuk, az a pontot egy gömb, esetünkben $E(\mathbf{1}, \alpha r)$ középpontjába kell transzformálnunk. Eközben szeretnénk megőrizni a (2.2) feladat specialitását. Ez megtehető az alábbi, az F.6. függelékben leírtakhoz hasonló, projektív transzformációval.

$$T'(x) = \frac{n}{\sum_{j=1}^{n} \frac{\xi_j}{\alpha_j}} \left(\frac{\xi_1}{\alpha_1}, \dots, \frac{\xi_n}{\alpha_n} \right) = \frac{n}{\mathbf{1}D^{-1}x} D^{-1}x = \bar{x},$$

ahol $D = \mathrm{diag}(\alpha_j)$. Könnyen ellenőrizhető (mint F.6. függelékben), hogy:

(1) $$T'^{-1}(\bar{x}) = \frac{n}{\sum_{j=1}^{n} \bar{\xi}_j \alpha_j} (\bar{\xi}_1 \alpha_1, \dots, \bar{\xi}_n \alpha_n) = \frac{n}{\mathbf{1}D^{-1}\bar{x}} D\bar{x},$$

(2) $$T'(a) = \mathbf{1},$$

(3) $$Ax = AT'^{-1}(\bar{x}) = 0 \iff AD\bar{x} = 0,$$

(4) $$(T'^{-1}(\bar{x}))_1 = \frac{n\alpha_1 \bar{\xi}_1}{\sum_{j=1}^{n} \bar{\xi}_j \alpha_j} = 0 \iff \bar{\xi}_1 = 0.$$

Így ha $\bar{x}$ megoldása a (legyen $0 < \alpha < 1$ mint F.4. függelékben)

(2.3) $$\begin{aligned} &\min \bar{\xi}_1 \\ &AD\bar{x} = 0 \\ &\mathbf{1}\bar{x} = n \\ &(\bar{x} - \mathbf{1})^2 \le \alpha^2 r^2 \end{aligned}$$

feladatnak, akkor $x = T'^{-1}(\bar{x})$ jobb megoldása (2.2)-nek mint a. Ekkor az F.3. függelékben leírtak szerint

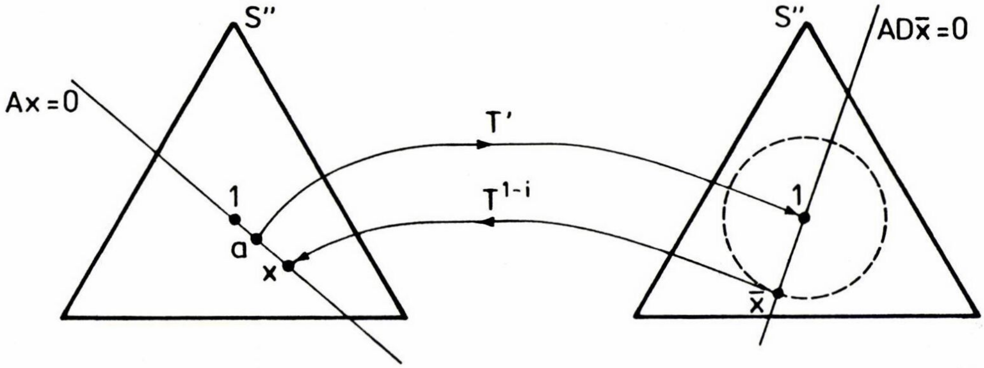
(2.4) $$\bar{x} = e_1 - \frac{1}{n}\mathbf{1} - DA^T(AD^2A^T)^{-1}ADe_1.$$

A leírt javító eljárást az 1. ábra szemlélteti.

A fenti javító eljárás ismételt alkalmazásával nyerjük *Karmarkar projektív algoritmusát.*

Algoritmus (K)

Inicializálás. - Adott az $x^0 = \mathbf{1} > 0$ pont, mely megoldása a (2.2) feladatnak. Legyen $k = 0$. Adott $\varepsilon > 0$.



1. ábra

Általános lépés. Adott $x^k > 0$, melyre $Ax^k = 0, \mathbf{1}x^k = n$. Legyen

$$D_k = \mathrm{diag}(\xi_j^k).$$

— Ha $\xi_1^k < \varepsilon$ akkor STOP.
— A segédfeladatban leírtak szerint legyen

$$x^{k+1} = T_k'^{-1}(\bar{x}^k)$$

— $k = k + 1$.
Megjegyezzük, hogy $D_{k+1} = \mathrm{diag}(x^{k+1}) = \dfrac{n}{\bar{x}^k x^k} D_k \; \mathrm{diag}(\bar{x}^k)$.

Az alábbiakban egy úgynevezett potenciálfüggvény segítségével belátjuk, hogy az algoritmus $\sigma(n \log \varepsilon)$ lépésben megáll (az optimális megoldásnál), azaz polinomiális.

2.1. Definició. Az $F(x) = \dfrac{\xi_1^n}{\prod_{j=1}^{n} \xi_j} := \dfrac{\xi_1^n}{\prod x}$ függvényt potenciálfüggvénynek nevezzük.

Megjegyezzük, hogy valójában az F függvény logaritmusa KARMARKAR potenciálfüggvénye, és mint azt az alábbiakban látni fogjuk, $\log(F)$ tényleg potenciálfüggvényként viselkedik. Vegyük észre továbbá, hogy korlátos tartomány esetén ξ_1 akkor és csak akkor „kicsi", ha $F(x)$ „kicsi".

Most a T' projektív transzformáció és az F potenciálfüggvény kapcsolatát vizsgáljuk.

2.1. LEMMA. *Ha $x, x' \in S^n$ és $x, x' > 0$ akkor*

$$\frac{F(T_k'(x))}{F(T_k'(x'))} = \frac{F(x)}{F(x')}.$$

Bizonyítás. Nyilván $F(T_k'(x)) = F\left(\frac{\xi_1}{\xi_1^k}, \dots, \frac{\xi_n}{\xi_n^k}\right) = \frac{F(x)}{F(x^k)}$. Hasonlóan $F(T_k'(x')) = \frac{F(x')}{F(x^k)}$, amiből következik az állítás.

KÖVETKEZMÉNY. *Minden $k \geq 0$ esetén* $\frac{F(x^{k+1})}{F(x^k)} = F(\bar{x}^k)$.

Bizonyítás. Mivel $T_k'(x^k) = \mathbf{1}$ és $x^{k+1} = T_k'^{-1}(\bar{x}^k)$, így $T_k'(x^{k+1}) = \bar{x}^k$, amiből $\frac{F(x^{k+1})}{F(x^k)} = \frac{F(T_k'(x^{k+1}))}{F(T_k'(x^k))} = \frac{F(\bar{x}^k)}{F(\mathbf{1})} = F(\bar{x}^k)$.

Így most már a potenciafüggvény javulás mértékére is becslést tudunk adni.

2.2. LEMMA. *Ha* $\alpha = \frac{1}{2}$, *akkor* $\frac{F(x^{k+1})}{F(x^k)} < \frac{2}{e}$.

Bizonyítás. Az előzőek alapján elég megmutatni, hogy $F(\bar{x}^k) < \frac{2}{e}$. Definíció szerint $F(\bar{x}^k) = \frac{(\bar{\xi}_1^k)^n}{\prod \bar{x}^k} \leq \frac{(1 - \frac{\alpha}{n-1})^n}{\prod \bar{x}^k}$, ahol az utóbbi egyenlőtlenség az F.4. függelékben bizonyított célfüggvénybecslés miatt áll fenn.

Mivel a $\prod \bar{x}^k$ szorzat az $E(\mathbf{1}, \alpha r) \cap S^n$ gömbön (lásd F.8. függelék) az $\mathbf{1} + \alpha\left(-1, \frac{1}{n-1}, \dots, \frac{1}{n-1}\right)$ pontban veszi fel a minimumát, így $\prod \bar{x}^k \geq (1-\alpha)$ $\left(1 + \frac{\alpha}{n-1}\right)^{n-1}$. Így $F(\bar{x}^k) \leq \frac{(1 - \frac{\alpha}{n-1})}{1-\alpha}\left(\frac{1 - \frac{\alpha}{n-1}}{1 + \frac{\alpha}{n-1}}\right)^{n-1}$, ahonnan az $\left(\frac{1-t}{1+t}\right)^{\frac{1}{t}}$ $\leq e^{-2}$ $(t \geq 0)$ egyenlőtlenség és $1 - \frac{\alpha}{n-1} < 1$ miatt az $F(\bar{x}^k) < \frac{e^{-2\alpha}}{1-\alpha}$ egyenlőtlenséget nyerjük. Ez utóbbi pedig $\alpha = 1/2$ esetén a kívánt egyenlőtlenséget adja.

KÖVETKEZMÉNY. *Ha* $\alpha = \frac{1}{2}$, *akkor* $F(x^k) < \left(\frac{2}{e}\right)^k$.

Bizonyítás. $F(x^0) = F(\mathbf{1}) = 1$ miatt az előző lemmákból adódik.

A potenciálfüggvény a fentiek szerint zérushoz konvergál, a konvergencia sebessége lineáris. Így elérkeztünk ahhoz a ponthoz, amikor beláthatjuk, hogy az algoritmus polinomiális lépésszámban véget ér.

2.1. TÉTEL. *Az algoritmus $\sigma(n \log \epsilon)$ lépés után megáll, azaz polinomiális.*

Bizonyítás. Az előző következmény miatt $\frac{(\xi_1^k)^n}{\prod x^k} < \left(\frac{2}{e}\right)^k$. Továbbá a számtani–mértani egyenlőtlenség miatt $\sqrt[n]{\prod x^k} \leq \frac{\sum_{j=1}^{n} \xi_j^k}{n} = 1$. Így $(\xi_1^k)^n < \left(\frac{2}{e}\right)^k$, amiből

$\xi_1^k < \varepsilon$, ha $\left(\frac{2}{e}\right)^{\frac{k}{n}} \leq \varepsilon$, azaz ha

$$k \geq n \frac{\log \varepsilon}{\log 2 - \log e}.$$

KÖVETKEZMÉNY. *Így ha $\varepsilon < 1/\sigma$, ahol σ a paraméterek terjedelme, ahogy azt az F.2. függelékben bevezettük, akkor az ott leírtak szerint optimális megoldásnál áll meg az algoritmus. Ekkor*

$$k \geq n \frac{\log \sigma}{1 - \log 2}.$$

Tehát a potenciálfüggvény segítségével beláttuk, hogy KARMARKAR projektív algoritmusa polinomiális lépésszámban megoldja a lineáris programozási feladatot.

3. A Dikin–Barnes algoritmus

Mint a bevezetőben is említettük, a KARMARKAR algoritmus affin változatát BARNES [2] közölte, de ez az algoritmus már DIKIN [4] munkáiban is megtalálható. Talán ez a legegyszerűbb, legkönnyebben érthető variáns a belsőpontos algoritmusok között.

Tekintsük a

$$\text{(3.1)} \qquad \text{Primál:} \quad \begin{array}{c} \min cx \\ Ax = b \\ x \geq 0 \end{array} \qquad \text{Duál:} \quad \begin{array}{c} \max y^T b \\ A^T y \leq c \end{array}$$

lineáris proramozási feladatpárt.

Legyen $\bar{x} > 0$ pozitív megoldása a primál feladatnak. Ilyen az F.1. függelékben leírtak szerint előállítható. Legyen $0 < \rho < 1$ tetszőleges rögzített szám.

3.1. LEMMA. *Minden $0 < \rho < 1$ esetén*

$$E(\bar{x}, \rho) = \left\{ x \;\middle|\; \sum_{j=1}^{n} \left(\frac{\xi_j - \bar{\xi}_j}{\bar{\xi}_j} \right)^2 \leq \rho^2 \right\} \subset \mathbf{R}^n_+.$$

Bizonyítás. Nyilvánvaló, hiszen $\xi_j \leq 0$ esetén $\dfrac{\xi_j - \bar{\xi}_j}{\bar{\xi}_j} \geq 1$.

Így a (3.1) primál feladat helyett (iterációs lépésként) oldjuk meg az alábbi egyszerűbb feladatot.

$$
\begin{aligned}
&\min c^T x\\
(3.2)\qquad &Ax = b\\
&\sum_{j=1}^{n} \frac{1}{\bar{\xi}_j^2}(\xi_j - \bar{\xi}_j)^2 \le \rho^2.
\end{aligned}
$$

Vegyük észre, hogy (3.2) feladat struktúrálisan megegyezik az F.3. függelékben tárgyalt (F3.1) feladattal, az $A = A$, $c = c$, $b = b$, $a = \mathbf{1}$, $D = \mathrm{diag}(\bar{\xi}_j)$ szereposztással. Így esetünkben

$$u = -(AD^2A^T)^{-1}AD^2c \tag{3.3}$$

$$\zeta = \frac{1}{\rho}\|D(c + A^T u)\| \tag{3.4}$$

$$x = a - \frac{D^2(c + A^T u)}{\|D(c + A^T u)\|}\rho \tag{3.5}$$

$$z = \zeta D^{-1}(x - a) \tag{3.6}$$

$$cx = c\bar{x} - \rho\|D(c + A^T u)\| \tag{3.7}$$

Így most már megfogalmazhatjuk a *Dikin–Barnes féle algoritmust.*

Algoritmus (D–B)

Inicializálás. Adott $x^0 > 0$, amelyre $Ax^0 = b$ és $\mathrm{rang}(A) = m$.

Általános lépés.

$-x^k > 0, Ax^k = b$ adott.

$-$Legyen $x^{k+1} = x^k - \rho \dfrac{D_k^2(c + A^T u^k)}{\|D_k(c + A^T u^k)\|}$

(3.8) ahol $D_k = \mathrm{diag}\ (\xi_j^k)$ és

(3.9) $u^k = -(AD_k^2A^T)^{-1}AD_k^2c$

$-k = k + 1.$

Az alábbi tétel az algoritmus konvergenciáját biztosítja.

3.1. TÉTEL. *Ha a (3.1) primál-duál feladatpár nem degenerált és van optimális megoldása, akkor az $\{x^k\}$ sorozat a primál, a $\{-u^k\}$ sorozat pedig a duál feladat optimális megoldásához konvergál.*

Bizonyítás. A nemdegeneráltsági feltételből következik, hogy a primál és a duál feladatnak is egyértelmű optimális megoldása létezik, ami csúcs. Feltehető, hogy ekkor A első m oszlopa adja a bázist (optimális bázist).

Az F.2. függelékben ismertetettek szerint az x primál és y duál optimális megoldások esetén $\xi_1 > \frac{1}{\sigma}, \ldots, \xi_m > \frac{1}{\sigma}, \xi_{m+1} = 0, \ldots, \xi_n = 0$ és $\gamma_1 - ya_1 = 0, \ldots, \gamma_m - ya_m = 0, \gamma_{m+1} - ya_{m+1} > \frac{1}{\sigma}, \ldots, \gamma_n - ya_n > \frac{1}{\sigma}$.

A (3.7) képlet szerint $cx^{k+1} = cx^k - \rho\|D_k(c + A^T u^k)\|$, vagyis a célfüggvény érték szigorúan monoton csökkenő sorozatot alkot, ez a sorozat alulról korlátos (mivel van optimális megoldása a feladatnak), tehát konvergens. Így $\lim_{k\to\infty} \|D_k(c + A^T u^k)\| = 0$. Legyen $\varepsilon < \frac{1}{\sigma}$, és legyen p olyan nagy, hogy

$$\|D_k(c + A^T u^k)\| < \varepsilon^2, \quad \text{ha} \quad k \geq p.$$

Ekkor $\varepsilon^2 > \sqrt{\sum_{j=1}^{n} \left(\xi_j^k\right)^2 (\gamma_j - a_j y^k)^2} \geq \sqrt{\left(\xi_j^k\right)^2 (\gamma_j - a_j y^k)^2} = \xi_j^k |\gamma_j - a_j y^k|$.

Nemdegeneráltság miatt van $n - m$ darab olyan koordináta, melyre (feltehető $j = m+1, \ldots n$) $|\gamma_j - a_j y^k| > \varepsilon$. Így $\xi_j^k < \varepsilon$, $j = m+1, \ldots, n$, és így nemdegeneráltság miatt $\xi_j > \varepsilon$, $j = 1, \ldots, m$ és $|\gamma_j - a_j y^k| < \varepsilon$, $j = 1, \ldots, m$.

Ezekből a feltételekből következik, hogy $a_1, \ldots, a_m$ lineárisan független rendszer, így bázis. Legyen x illetve y az ehhez a bázishoz tartozó megoldás. (Ekkor x^k „közel van" x–hez, $(c + A^T u^k)$ „közel van" $(c - A^T y)$–hoz.)

Az F.2. függelékben leírtak szerint két különböző bázismegoldás távolsága legalább $\sqrt{2}\varepsilon$, így egyértelmű az az x csúcspont, melyhez x^k „közel van". A

$$\left(\frac{\xi_j^{r+1} - \xi_j^r}{\xi_j^r}\right)^2 \leq \sum_{j=1}^{n} \left(\frac{\xi_j^{r+1} - \xi_j^r}{\xi_j^r}\right)^2 \leq \rho^2$$

egyenlőtlenségből következik, hogy $(\xi_j^{r+1} - \xi_j^r)^2 \leq \rho^2(\xi_j^r)^2$, amiből $\xi_j^r - \xi_j^{r+1} \leq \rho\xi_j^r$, azaz $\xi_j^{r+1} \geq (1-\rho)\xi_j^r$. Így $\varepsilon < (1-\rho)\frac{1}{\sigma}$ esetén $\xi_j^r > \frac{1}{\sigma}$–ból következik $\xi_j^{r+1} > \varepsilon$, azaz elég nagy k esetén az $\{x^k\}$ sorozat mindig az x csúcs egy környezetében marad. Tehát $\lim_{k\to\infty} x^k = x$.

Hasonlóan, meg kell mutatnunk, hogy $\lim_{k\to\infty}(-u^k) = y = c_B B^{-1}$, ahol $B = (a_1, \ldots, a_m)$, $c_B = (\gamma_1, \ldots, \gamma_m)$. Legyen $\bar{D}_k = \mathrm{diag}(\xi_1^k, \ldots, \xi_m^k)$. Így

$$AD_k^2 A^T = \sum_{j=1}^{n} (\xi_j^k)^2 a_j a_j^T = B\bar{D}_k^2 B + \sum_{j=m+1}^{n} (\xi_j^k)^2 a_j a_j^T = B\bar{D}_k^2 B + \varepsilon_1,$$

ahol $\varepsilon_1 = o(\varepsilon^2)$, melyből

$$\begin{aligned}(AD_k^2 A^T)^{-1} &= (I + (B\bar{D}_k^2 B)^{-1}\varepsilon_1)^{-1}(B\bar{D}_k^2 B)^{-1} = \\ &= \{I - (B\bar{D}_k^2 B)^{-1}\varepsilon_1 + (B\bar{D}_k^2 B)^{-2}\varepsilon_1^2 - \ldots\}(B\bar{D}_k^2 B)^{-1} = \\ &= (B\bar{D}_k^2 B)^{-1} + \varepsilon_2,\end{aligned}$$

ahol $\varepsilon_2 = o(\varepsilon^2)$. Az u^k vektor meghatározásához még az $AD_k^2 c$ szorzatot kell kiszámolni. A fentiekhez hasonlóan

$$AD_k^2 c = B\bar{D}_k^2 c_B + \sum_{k=m+1}^{n} a_j \gamma_j (\xi_j^k)^2.$$

A fenti két formula alapján

$$-u^k = (AD_k^2 A^T)^{-1} AD_k^2 c = (B\bar{D}_k^2 B^T)^{-1} B\bar{D}_k^2 c_B + \varepsilon_3 = (B^T)^{-1} c_B + \varepsilon_3,$$

ahol $\varepsilon_3 = o(\varepsilon^2)$. Így nyilván, ahogy állítottuk,

$$\lim_{k\to\infty} (-u^k) = y = c_B B^{-1}.$$

Tehát eddig beláttuk, hogy az $\{x^k\}$ és a $\{-u^k\}$ sorozatok ugyanahhoz a (primál illetve duál) bázismegoldáshoz konvergálnak. Mivel belső pontokon halad az algoritmus, az x megoldás primál megengedettsége nyilvánvaló. Be kell látnunk, hogy az y megoldás duál megengedett, továbbá az optimalitáshoz be kell még látnunk a komplementaritási feltételek teljesülését. A (3.5) képlet szerint

$$\xi_j^{k+1} = \xi_j^k - \frac{(\xi_j^k)^2 (\gamma_j - a_j y^k)}{\|D_k(c - A^T y^k)\|}$$

melyből következik, hogy

$$\begin{aligned} \gamma_j - a_j y^k < 0 \quad &\Rightarrow \quad \xi_j^{k+1} > \xi_j^k \\ \gamma_j - a_j y^k > 0 \quad &\Rightarrow \quad \xi_j^{k+1} < \xi_j^k \end{aligned}$$

(Hasonlítsuk össze a szimplex módszernél tapasztalt változó növekménnyel.)

Így a bázisváltozókra $\gamma_j - a_j y^k \to 0$, mivel $\xi_j^k \to \xi_j > \varepsilon$. Így y duál megengedettségéhez azt kell belátni, hogy $\gamma_j - a_j y^k > 0$, $k = m+1, \ldots, n$. A nembázis változók esetén pedig $\gamma_j - a_j y^k > 0$, ugyanis ez a kifejezés csak úgy válthatna előjelet, ha közben zérus értéket is felvenne, ami a nemdegeneráltsági feltétel miatt lehetetlen. Így a ξ_j^k koordináták $j = m+1, \ldots, n$ esetén monoton tartanak zérushoz, továbbá

$$\lim_{k\to\infty} (\gamma_j - a_j y^k) = \gamma_j - a_j y \geq 0, \quad j = m+1, \ldots, n.$$

Nyilván $cx = c_B x_B = c_B B^{-1} b = yb$, azaz a célfüggvény értékek megegyeznek, mivel azonos bázishoz tartozó megoldásokról van szó. Így nyilván a komplementaritási feltétel is teljesül, azaz $x^T(c - A^T y) = 0$, ugyanis $\gamma_j - a_j y = 0$, $j = 1, \ldots, m$ és $\xi_j = 0$, $j = m+1, \ldots, n$.

Összefoglalva, $Ax = b$, $x \geq 0$, $A^T y \leq c$, $(c - yA)x = 0$, azaz az $\{x^k\} \to x$ és $\{y^k\} \to y$ sorozatok a feladat optimális megoldásaihoz konvergálnak.

Konvergenciasebesség

Az iterációs képletek és a célfüggvény javulását leíró képletek alapján megbecsülhetjük a *Dikin–Barnes algoritmus* konvergenciasebességét.

3.2. TÉTEL. *A (D–B) algoritmus által generált $\{x^k\}$ sorozat esetén*

$$cx^{k+1} - cx \le \left(1 - \frac{\rho}{\sqrt{n-m+\varepsilon_k}}\right)(cx^k - cx),$$

ahol x a feladat optimális megoldása.

Bizonyítás. Az előző tétel bizonyításából tudjuk, hogy $x = (\xi_1, \dots, \xi_m, 0, \dots, 0)$. Így

$$\begin{aligned} cx^k - cx &= [D_k(c - A^T y^k)]^T D_k^{-1}(x^k - x) \le \|D_k(c - A^T y^k)\| \sqrt{n-m+\varepsilon_k} = \\ &= \frac{1}{\rho}(cx^k - cx^{k+1})\sqrt{n-m+\varepsilon_k} \end{aligned}$$

ahol $\varepsilon_k = \sum_{j=1}^{m} \left(\frac{\xi_j^k - \xi_j}{\xi_j^k}\right)^2 \to 0$, ha $k \to \infty$. Ebből

$$(cx^{k+1} - cx) - (cx^k - cx) \le -\frac{\rho}{\sqrt{n-m+\varepsilon_k}}(cx^k - cx)$$

$$cx^{k+1} - cx \le \left(1 - \frac{\rho}{\sqrt{n-m+\varepsilon_k}}\right)(cx^k - cx)$$

Ahogy állítottuk.

4. Roos és Vial trajektóriakövető algoritmusa

Tekintsük a standard formájú LP feladatpárt:

$$\text{(4.1)} \qquad \text{Primál:} \quad \begin{aligned} &\min cx \\ &Ax = b \\ &x \ge 0 \end{aligned} \qquad \text{Duál:} \quad \begin{aligned} &\max by \\ &A^T y \le c \end{aligned}$$

Legyen $D = \mathrm{diag}(\xi_i)$, és vizsgáljuk az alábbi paraméteres lineáris komplementaritási feladatot:

$$\text{(4.2)} \qquad \begin{aligned} Ax &= b \\ A^T y + z &= c \\ Dz &= \beta \mathbf{1} \end{aligned} \qquad \begin{aligned} x &> 0 \\ z &> 0 \end{aligned}$$

Nyilván $\beta = 0$ esetén (4.2) megoldása ekvivalens (4.1) megoldásával. Megjegyezzük, hogy (4.2) nem más, mint a primál feladat β paraméterű $\min\{cx - \beta\sum\log\xi_j \mid Ax = b,\quad x \geq 0\}$ logaritmikus barrier függvényes feladatának *Kuhn-Tucker rendszere*.

Amennyiben adott β és adott $x > 0$ értékek mellett szeretnénk a (4.2) feltételeket lehető legjobban kielégítő (y, z) vektort előállítani, akkor az (F8.1) feladatot kapjuk, melynek megoldása (F8.2),...,(F8.6) képletekkel explicite előállítható.

Így az ott leírtak szerint:

$$x^* = x - D\left(\frac{Dz}{\beta} - \mathbb{1}\right) = 2x - \frac{D^2 z}{\beta}.$$

Az $s = \dfrac{Dz}{\beta}$ jelölést bevezetve tehát $x^* = 2x - Ds$, és $\delta(x,\beta) = \|s - \mathbb{1}\|$.

4.1. LEMMA. *Ha $\delta(x,\beta) < 1$, akkor $x^* > 0$, $Ax = Ax^*$ és $\delta(x^*,\beta) \leq \delta(x,\beta)^2$.*

Bizonyítás. $\delta(x,\beta) < 1$ - ből az $(s - \mathbb{1})^2 < 1$ egyenlőtlenséget kapjuk, így $s = (\sigma_1, \ldots, \sigma_n)$ jelöléssel $0 < \sigma_i < 2$, ahonnan

$$\xi_i^* = 2\xi_i - \xi_i\sigma_i = \xi_i(2 - \sigma_i) > 0, \quad \text{azaz} \quad x^* > 0.$$

Továbbá (F8.2) felhasználálával s definíciója miatt

$$Ax^* = 2Ax - ADs = 2Ax - Ax = Ax$$

(F8.5) miatt (y, z) megengedett megoldása az (F8.1) feladatnak, így

$$\delta^2(x^*,\beta) \leq \left(\frac{D^* z}{\beta} - \mathbb{1}\right)^2 = (D^* D^{-1} s - \mathbb{1})^2.$$

Mivel $D^* D^{-1} s = (2D - DS)D^{-1}s = 2s - Ss$, ahol $S = \operatorname{diag}(\sigma_i)$, így

$$\begin{aligned}\delta^2(x^*,\beta) &\leq (2s - Ss - \mathbb{1})^2 = \sum_{i=1}^{n}(2\sigma_i - \sigma_i^2 - 1)^2 = \sum_{i=1}^{n}(\sigma_i - 1)^4 \\ &\leq \left(\sum_{i=1}^{n}(\sigma_i - 1)^2\right)^2 = \delta(x,\beta)^4.\end{aligned}$$

A fenti lemmában csak az $x = D\mathbb{1}$ paramétereket módosítottuk, a következő lemmában arra adunk becslést, hogyan változik $\delta(x,\beta)$, ha β–t csökkentjük.

4.2. LEMMA. *Legyen $\beta^* = (1-\vartheta)\beta$, ahol $0 < \vartheta < 1$. Ekkor*

$$\delta(x,\beta^*) \le \frac{1}{1-\vartheta}(\delta(x,\beta) + \vartheta\sqrt{n}).$$

Bizonyítás. Mint 4.1 lemmában is láttuk, (y,z) megengedett megoldása az (F8.1) feladatnak. Így

$$\delta(x,\beta^*)^2 \le \left(\frac{\beta}{\beta^*}s - \mathbb{1}\right)^2 = \left(\frac{1}{1-\vartheta}s - \mathbb{1}\right)^2, \quad \text{ahonnan}$$

$$\delta(x,\beta^*) \le \left\|\frac{1}{1-\vartheta}(s-\mathbb{1}) + \frac{\vartheta}{1-\vartheta}\mathbb{1}\right\| \le \frac{1}{1-\vartheta}\|s-\mathbb{1}\| + \frac{\vartheta}{1-\vartheta}\sqrt{n}$$

$$\delta(x,\beta^*) \le \frac{1}{1-\vartheta}(\delta(x,\beta) + \vartheta\sqrt{n}),$$

ahogy állítottuk.

A fenti két lemma eredményét a következő lemmában foglaljuk össze, amikor arra adunk becslést, ha $x \to x^*$, $\beta \to \beta^*$ változtatásokat egyidejűleg hajtjuk végre.

4.3. LEMMA. *Ha $\delta(x,\beta) < \frac{1}{2}$, $\vartheta = \dfrac{1}{6\sqrt{n}}$ és x^*,β^* az előbbiekben adottak, akkor*

$$\delta(x^*,\beta^*) \le \frac{1}{2}.$$

Bizonyítás. Az előző lemmákat felhasználva:

$$\delta(x^*,\beta^*) \le \frac{1}{1-\vartheta}(\delta(x^*,\beta) + \vartheta\sqrt{n}) \le \frac{1}{1-\vartheta}(\delta(x,\beta)^2 + \vartheta\sqrt{n}) \le$$

$$\le \frac{1}{1-\frac{1}{6\sqrt{n}}}\left(\frac{1}{4} + \frac{1}{6\sqrt{n}}\sqrt{n}\right) = \frac{6\sqrt{n}}{6\sqrt{n}-1}\cdot\frac{5}{12} \le \frac{1}{2}.$$

A 4.3 lemma alapján már egy iteratív algoritmus konstruálható x és β sorozatos módosításával. Mielőtt ezt leírnánk, előbb még egy becslést bizonyítunk.

4.4. LEMMA. *Ha $\delta(x,\beta) < 1$, (y,z) kielégíti az (F8.5), (F8.6) feltételeket, akkor $z \ge 0$, és*

$$\beta(n - \delta(x,\beta)\sqrt{n}) \le cx - yAx \le \beta(n + \delta(x,\beta)\sqrt{n}).$$

Bizonyítás. $\delta(x,\beta)$ definíciója szerint $\delta(x,\beta) = \|\dfrac{Dz}{\beta} - \mathbb{1}\| \le 1$, ahonnan $\dfrac{Dz}{\beta} \ge 0$, azaz $x > 0$ miatt $z > 0$. (Ekkor $A^T y < c$). Másrészt

$$cx - yAx = (c - A^T y)^T x = zx, \quad \text{és}$$

$$\delta(x,\beta)\sqrt{n} = \left\|\frac{Dz}{\beta} - \mathbf{1}\right\| \|\mathbf{1}\| \geq \left|\frac{xz}{\beta} - n\right|, \quad \text{ahonnan}$$

$$n - \delta(x,\beta)\sqrt{n} \leq \frac{xz}{\beta} \leq n + \delta(x,\beta)\sqrt{n}$$

$$\beta(n - \delta(x,\beta)\sqrt{n}) \leq cx - yAx \leq \beta(n - \delta(x,\beta)\sqrt{n}).$$

Ezek alapján az alábbi algoritmust nyerjük.

Algoritmus (R-V)

Inicializálás. Legyen (x^0, β^0) adott olyan, hogy

$$\delta(x^0, \beta^0) \leq 1/2, \quad x^0 > 0, \quad \beta^0 > 0.$$

Legyen továbbá $\vartheta = \frac{1}{6\sqrt{n}}$.

Általános lépés.

1. Ha $n\beta^k < \varepsilon$ akkor STOP.
2. Legyen $\beta^{k+1} = \beta^k(1 - \vartheta)$

   $x^{k+1} = 2x^k - \dfrac{D_k^2 z^k}{\beta^{k+1}}$, ahol y^k, z^k az (F8.5), (F8.6) képletekkel adott.
3. k:=k+1

4.1. TÉTEL. *Legyen ε adott. Akkor $6\sqrt{n}(\log\beta^0 - \log\varepsilon)$ lépés után megáll a (R–V) algoritmus, az utolsó megoldásban y duál és x primál megengedett, továbbá*

$$cx - by \leq \frac{3}{2}\varepsilon.$$

Bizonyítás. A fentiekben, illetve az F.8. függelékben leírtak szerint az iteráció minden lépésében $x > 0$, $Ax = b$, $\beta > 0$ és $\delta(x,\beta) \leq \frac{1}{2}$. A k-adik iteráció után $\beta = (1 - \vartheta)^k\beta^0$, így

$$(1 - \vartheta)^k\beta^0 < \varepsilon$$

$$k\log(1 - \vartheta) + \log\beta^0 < \log\varepsilon$$

$$k > \frac{\log\varepsilon - \log\beta^0}{\log(1 - \vartheta)}$$

A $\vartheta = \dfrac{1}{6\sqrt{n}}$ értéket behelyettesítve, és mivel ekkor $\log(1 - \vartheta) < -\vartheta$ így

$$k > \frac{\log\varepsilon - \log\beta^0}{-\vartheta} = \frac{\log\varepsilon - \log\beta^0}{-\frac{1}{6\sqrt{n}}} = 6\sqrt{n}(\log\beta^0 - \log\varepsilon),$$

ahogy állítottuk.

A dualitási résre vonatkozó becslés 4.4. lemma alapján:

$$cx - by \leq \beta(n + \delta(x,\beta)\sqrt{n})$$
$$cx - by \leq n\beta\left(1 + \frac{\delta(x,\beta)}{\sqrt{n}}\right).$$

A megállási szabály szerint $n\beta < \varepsilon$, feltétel szerint $\delta(x,\beta) \leq \frac{1}{2}$, így $\frac{\delta(x,\beta)}{\sqrt{n}} \leq \frac{1}{2}$;

$$cx - by \leq \varepsilon\left(1 + \frac{1}{2}\right) = \frac{3}{2}\varepsilon.$$

Függelék

F1. Belső pont előállítása (első fázis)

Tekintsük a $P = \{x \mid Ax = b,\ x \geq 0\}$ feladatot. Legyen $\bar{x} > 0$ tetszőleges vektor. Az alábbi feladat tekinthető a belső pontos algoritmusok első fázis feladatának. Oldjuk meg az

$$\begin{aligned} Ax + \lambda(b - A\bar{x}) &= b \\ x &\geq 0 \\ \min \lambda& \end{aligned} \tag{F1.1}$$

feladatot egy (tetszőleges) belső pontos algoritmussal (amely belső pontokon haladva oldja meg a feladatot).

1. Észrevétel. Vegyük észre, hogy $x = \bar{x}$ és $\lambda = 1$ pozitív induló megoldást biztosít az (F1.1) feladathoz.

2. Észrevétel. $\lambda = 0$ esetén az $Ax = b$ feladat egy megoldását kapjuk.

F1. TÉTEL.

a. *Ha az (F1.1) feladat optimális megoldása pozitív, akkor $P = \emptyset$.*
b. *Ha az (F1.1) feladat optimális megoldása nulla, akkor $P \neq \emptyset$, de $P \cap \mathbf{R}^n_+ = 0$.*
c. *Ha az (F1.1) feladat optimális megoldása negatív, akkor $P \cap \mathbf{R}^n_+ \neq \emptyset$, azaz feladatunkhoz pozitív megoldást kapunk.*

Bizonyítás.

a. Mint a szimplex módszer első fázisa esetén nyilvánvaló.
b. Tegyük fel indirekt, hogy van $x > 0$ megoldása az $Ax = b$ feladatnak. $\mathrm{Rang}(A) = m$ miatt van olyan $\hat{x} \in \mathbf{R}^n$, amelyre

$$A\hat{x} = (b - A\bar{x}).$$

Legyen $\varepsilon = \frac{1}{2}\min\left\{\frac{\xi_j}{-\hat{\xi}_j} \mid \hat{\xi}_j < 0\right\}$. Ekkor $x + \varepsilon\hat{x} > 0$, $\lambda = -\varepsilon < 0$, valamint $A(x + \varepsilon\hat{x}) + \lambda(b - A\bar{x}) = Ax + \varepsilon A\hat{x} - \varepsilon(b - A\bar{x}) = b$, azaz feltevésünkkel ellentétben $\min\lambda \leq -\varepsilon < 0$, ami ellentmondás.

c. Ekkor van olyan (x^1, λ^1) pont (melyre $x^1 > 0$, $\lambda^1 \geq 0$), és (x^2, λ^2) pont (melyre $x^2 > 0$, $\lambda^2 < 0$) az algoritmus által generált sorozatban, ahol a célfüggvény érték előjelet vált. Ekkor a

$$\frac{\lambda^1}{\lambda^1 - \lambda^2}x^2 - \frac{\lambda^2}{\lambda^1 - \lambda^2}x^1$$

pont megoldás és szigorúan pozitív.

Így tetszőleges polinomiális belső pontos algoritmus alkalmazásával polinomiális lépésszámban eldönthető, hogy egy lineáris egyenlőtlenségrendszernek van-e pozitív (belő pont) megoldása vagy nincs, és az algoritmus egyben megoldást is ad.

F2. Bázismegoldás értékek korlátai

Az alábbiakban bizonyítás nélkül röviden összefoglaljuk a bázismegoldás értékekre vonatkozó [10,20] dolgozatokban található eredményeket.

Tekintsük az $Ax = b$, $x \geq 0$, illetve az $A^T y \leq c$ lineáris egyenlőtlenségrendszereket. Legyen

$$\sigma = \|a^{(1)}\| \ldots \|a^{(m)}\| \, \|c\| \, \|b\|$$

az adatok „terjedelme". ($\|\cdot\|$ az *Euklideszi normát* jelöli $a^{(i)}$ pedig az A mátrix i-ik sorát.)

F2. TÉTEL. *Legyen B az A mátrix tetszőleges bázisa.*

a. *Ha x a B-hez tartozó bázismegoldás, akkor $\xi_j = 0$, vagy $|\xi_j| \geq \frac{1}{\sigma}$.*

b. *Ha x a B-hez tartozó bázismegoldás, akkor $|\xi_j| \leq \sigma$.*

c. *Ha y a B-hez tartozó bázismegoldás, akkor $\gamma_j = a_j y$, vagy $|\gamma_j - a_j y| \geq \frac{1}{\sigma}$.*

d. *Ha y a B-hez tartozó bázismegoldás, akkor $|\gamma_j - a_j y| \leq \sigma$.*

e. *Ha x és x' két különböző nemdegenerált bázismegoldás, akkor $\|x - x'\| \geq \frac{\sqrt{2}}{\sigma}$.*

f. *Ha y és y' két különböző nemdegenerált bázismegoldás, akkor $\|y - y'\| \geq \frac{\sqrt{2}}{\sigma^2}$.*

Bizonyítás. Az *a.*, *b.*, *c.*, *d.* állítások jól ismertek, [10] és [20]-ban is megtalálhatók. Bizonyításuk a *Cramer szabály* alkalmazásával a *Hadamard egyenlőtlenség*ből adódik.

e. $x \neq x'$-ből adódik, hogy van olyan koordináta (feltehető, hogy ξ_1), mely x-ben nem zérus és x'-ben zérus, és van ilyen koordináta (ξ_2'), ha x és x' szerepét felcseréljük. Ekkor

$$\|x - x'\| \geq \sqrt{\xi_1^2 + \xi_2'^2} \geq \sqrt{\frac{1}{\sigma^2} + \frac{1}{\sigma^2}} = \sqrt{\frac{2}{\sigma^2}} = \frac{\sqrt{2}}{\sigma}.$$

f. Hasonlóan a fentiekhez

$$\|(c - A^T y) - (c - A^T y')\| \geq \frac{\sqrt{2}}{\sigma},$$

amiből

$$\frac{\sqrt{2}}{\sigma} \leq \|(c - A^T y) - (c - A^T y')\| = \|A^T(y - y')\| \leq \|A^T\| \cdot \|y - y'\| \leq \sigma\|y - y'\|,$$

melyből a kívánt $\|y - y'\| \geq \frac{\sqrt{2}}{\sigma^2}$ összefüggés adódik.

A fenti összefüggések garantálják, hogy ha egy iteratív algoritmussal „elég közel" jutunk egy bázismegoldáshoz (nemdegeneráltsági feltétel mellett az optimális megoldás egyértelmű, és bázismegoldás), akkor azonosíthatók a nemnulla koordináták, és így a bázis, melynek ismeretében *Gauss eliminációval* a bázismegoldást is kiszámíthatjuk.

F3. Lineáris célfüggvény minimalizálása ellipszoid és affin tér metszetén

Tekintsük a

$$\text{(F3.1)} \qquad \begin{gathered} \min c^T x \\ Ax = b \\ (x - a)^T P^{-1}(x - a) \leq \rho^2 \end{gathered}$$

feladatot, ahol $a > 0$ az a vektor, melyre $Aa = b$ (l. F1.1), és P pozitív definit szimmetrikus mátrix. Jól ismert, hogy az utolsó feltétel ellipszoidot határoz meg, és az első feltétel egy affin teret. Ismert, hogy a az ellipszoid középpontja, és a P mátrix felbontható $P = DD^T$ (és így $P^{-1} = D^{T^{-1}} D^{-1}$) alakba, ahol D is teljes rangú mátrix.

Az (F3.1) feladat az alábbi alakba írható:

$$\text{(F3.2)} \qquad \begin{gathered} \max (-c)^T x \\ Ax = b \\ (D^{-1}x - D^{-1}a)^2 - \rho^2 \leq 0. \end{gathered}$$

Ennek duálja pedig a következő:

$$\text{(F3.3)}\qquad \begin{aligned} &\min\left\{z^TD^{-1}a+\zeta\frac{\rho^2}{2}+u^Tb+\frac{1}{2}\sum_{j=1}^{n}\frac{\zeta_j^2}{\zeta}\right\}\\ &D^{T-1}z+A^Tu=-c\\ &\zeta\geq 0,\ \zeta=0\Rightarrow z=0 \end{aligned}$$

Az (F3.2) és az (F3.3) feladat dualitását igazolja a következő lemma.

F3. LEMMA. *Ha x megengedett megoldása az (F3.2) feladatnak és (ζ, z, u) megengedett megoldása az (F3.3) feladatnak, akkor*

$$(-c)x\leq z^TD^{-1}a+\zeta\frac{\rho^2}{2}+u^Tb+\frac{1}{2}\sum_{j=1}^{n}\frac{\zeta_j^2}{\zeta}$$

egyenlőtlenség teljesül akkor és csak akkor, ha

$$\text{(F3.4)}\qquad \zeta[D^{-1}(x-a)]^2=\zeta\rho^2 \quad \textit{és}$$

$$\text{(F3.5)}\qquad \zeta_j=\zeta(d_jx-d_ja)$$

(ahol d_j a D^{-1} mátrix j-edik sorát jelöli).

Bizonyítás. A duál majd a primál feltételt felhasználva

$$-cx=z^TD^{-1}x+u^TAx=u^Tb+z^TD^{-1}x=u^Tb+\sum_{j=1}^{n}\zeta_jd_jx,$$

amelyből az

$$\frac{1}{2\zeta}\sum_{j=1}^{n}[\zeta_j-\zeta(d_jx-d_ja)]^2\geq 0$$

(egyenlőtlenség akkor és csak akkor, ha (F3.5) fennáll) egyenlőtlenséget felhasználva a

$$-cx\leq u^Tb+\frac{1}{2}\sum_{j=1}^{n}\frac{\zeta_j^2}{\zeta}+\sum_{j=1}^{n}\zeta_jd_ja+\frac{\zeta}{2}\sum_{j=1}^{n}(d_jx-d_ja)^2$$

egyenlőtlenséget nyerjük. Ebből az ellipszoid feltétel felhasználásával (egyenlőség akkor és csak akkor ha (F3.4) fennáll) a kívánt egyenlőtlenséget nyerjük.

A lemma felcsillantja az (F3.2) feladat „könnyű" megoldhatóságának lehetőségét.

KÖVETKEZMÉNY. *Ha x olyan primál, és (ζ, z, u) olyan duál megengedett megoldások, melyek az (F3.4) és (F3.5) feltételeket is kielégítik, akkor x primál és (ζ, z, u) duál optimális megoldás.*

Bizonyítás. Nyilvánvaló.

Így a feladat megoldása egy (nemlineáris) egyenletrendszer megoldására redukálódik, melyet az alábbiak szerint oldhatunk meg.

Megoldások előállítása

F3. TÉTEL. *Az (F3.3) és (F3.2) optimális megoldásai az alábbi módon állíthatók elő:*

$$u = -(APA^T)^{-1}APc$$

$$\zeta = \frac{1}{\rho}\|D^T(c + A^T u)\|$$

$$x = a - \frac{P(c + A^T u)}{\|D^T(c + A^T u)\|}\rho = a - \frac{1}{\zeta}P(c + A^T u)$$

$$z = \zeta D^{-1}(x - a)$$

Bizonyítás. A duál feltételből

$$APA^T u = -APc - APD^{T^{-1}} z$$

$$u = -(APA^T)^{-1}APc - (APA^T)^{-1}APD^{T^{-1}} z,$$

ahonnan az (F3.5) feltételből nyert $D^{T^{-1}} z = \zeta P^{-1}(x - a)$ összefüggés, a primál feltétel és $Aa = b$ miatt

$$u = -(APA^T)^{-1}APc - (APA^T)^{-1}APP^{-1}(x - a)\zeta = -(APA^T)^{-1}APc.$$

Így az első összefüggést beláttuk. Hasonlóan a duál és (F3.5) feltételből

$$\zeta P^{-1}(x - a) + A^T u = -c, \quad \text{ahonnan} \quad x = a - \frac{1}{\zeta}P(c + A^T u),$$

ami a harmadik összefüggés. A második összefüggés belátásához tekintsük az (F3.4) és (F3.5) képleteket. Ezekből

$$z^2 = \zeta^2 \rho^2,$$

ahonnan

$$\zeta^2 \rho^2 = \zeta^2 [D^{-1}(x - a)]^2$$

melyből

$$x = a - \frac{1}{\zeta}P(c + A^T u)$$

felhasználásával

$$\zeta^2 \rho^2 = \zeta^2 \frac{1}{\zeta^2}[D^{-1}DD^T(c + A^T u)]^2 = [D^T(c + A^T u)]^2,$$

azaz

$$\zeta = \frac{1}{\rho}\|D^T(c + A^T u)\|$$

adódik.

A negyedik összefüggés valójában (F3.5) feltétel összevont alakja.

Így lehetőségünk van a kívánt feladat megoldására, a megoldásértékek explicit kiszámolására.

Megjegyezzük, hogy az (F3.2) és (F3.3) feladatok l_p programozási primál és duál feladatok, melyek dualitáselméletét [23]-t felhasználva is azonos eredményre juthattunk volna. A feladat specialitásai miatt azonban közvetlenül, elemi eszközökkel is megoldhattuk feladatunkat.

Végül egy összefüggést bizonyítunk az eredeti és az új célfüggvény értékek között.

F3. KÖVETKEZMÉNY. $cx = ca - \rho\|D^T(c + A^T u)\|$.

Bizonyítás. (F3.3) duál feltételből indulva

$$D^{T^{-1}} z = -(c + A^T u)$$

$$(x - a)^T D^{T^{-1}} z = -(x - a)^T (c + A^T u) = -(x - a)^T c,$$

ugyanis $A(x - a) = 0$, ahonnan $(x - a)^T A^T = 0$. Továbbá a tétel azonosságai szerint

$$(x - a)^T D^{T^{-1}} z = \zeta (x - a)^T P(x - a) = \zeta \rho^2 = \rho\|D^T(c + A^T u)\|$$

és így a fenti két összefüggésből

$$cx = ca - \rho\|D^T(c + A^T u)\|.$$

F4. Szimplex, beírt és körülírt gömbök

Általában az n dimenziós térben $n + 1$ általános helyzetű pont konvex burkát nevezzük szimplexnek. Cikkünkben az

$$S^n = \{x \in \mathbf{R}^n : x \geq 0, \sum_{i=1}^{n} \xi_i = n\}$$

speciális helyzetű $n - 1$ dimenziós szimplexszel foglalkozunk. Könnyen belátható, hogy a $(0, \ldots, 0, n, 0, \ldots, 0)$ pontok az S^n szimplex csúcsai és az $\mathbf{1} = (1, \ldots, 1)$ pont az S^n szimplex centruma.

Körülírt gömb. $(\mathbf{1}, R)$: A körülírt gömb centruma megegyezik a szimplex centrumával, R sugara pedig a centrum és a csúcsok távolsága, azaz

$$R = \|(1, \ldots, 1) - (n, 0, \ldots, 0)\| = \sqrt{(n-1)^2 + (n-1)1^2} = \sqrt{n(n-1)}.$$

Így az S^n köré írt gömb egyenlete:

$$E(\mathbf{1}, R) = \{x \mid (x - \mathbf{1})^2 \leq n(n-1)\}$$

Beírt gömb. $E(\mathbf{1}, r)$: A beírt gömb centruma is az $\mathbf{1}$ pont, r sugara pedig a középpont és a határoló lapok ($n-2$ dimenziós szimplexek) távolsága, azaz esetünkben a középpont és a határoló lapok $(0, \frac{n}{n-1}, \dots, \frac{n}{n-1})$ középpontjának távolsága.

$$r = \|(1, \dots, 1) - (0, \frac{n}{n-1}, \dots, \frac{n}{n-1})\| =$$
$$= \sqrt{1 + (n-1)(\frac{n}{n-1} - 1)^2} = \sqrt{1 + \frac{1}{n-1}}.$$

Így $r = \sqrt{\frac{n}{n-1}}$, és a beírt gömb az

$$E(\mathbf{1}, r) = \{x \mid (x - \mathbf{1})^2 \le \frac{n}{n-1}\}$$

képlettel definiálható.

Arány. Így a beírt és a körülírt gömb sugarának aránya

$$\frac{r}{R} = \sqrt{\frac{n}{(n-1)n(n-1)}} = \frac{1}{n-1}$$

Amennyiben a beírt gömböt $0 < \alpha < 1$ arányban összehúzzuk, akkor a két sugár aránya $\frac{\alpha r}{R} = \frac{\alpha}{n-1}$.

Célfüggvény becslés

Legyen $\Omega = \{x \mid Ax = 0, \mathbf{1}x = n\}$, e_i az i–edik egységvektor.

F4.TÉTEL. *Tegyük fel, hogy* $\min\{e_1 x \mid x \in \Omega \cap \mathbf{R}^n_\oplus\} = 0$. *Ekkor* $0 < \alpha < 1$ *esetén*

$$\min\{e_1 x \mid x \in \Omega \cap \mathbf{R}^n_\oplus \cap E(\mathbf{1}, \alpha r)\} \le 1 - \frac{\alpha}{n-1}$$

Bizonyítás. Legyen $\delta_1 = \min\{e_1 x \mid x \in \Omega \cap E(\mathbf{1}, \alpha r)\}$ és $\delta_2 = \min\{e_1 x \mid x \in \Omega \cap E(\mathbf{1}, R)\}$. Ekkor a feltételek miatt nyilván $\delta_1 > 0 \ge \delta_2$, ugyanis a megengedett halmazok a fenti sorrendben tartalmazzák egymást.

Az $\mathbf{1}$ megengedett pontban a célfüggvény érték nyilván 1. Legyen p ($\|p\| = 1$) az optimális csökkenési irány az $\Omega \cap E(\mathbf{1}, R)$ halmazon. Ekkor

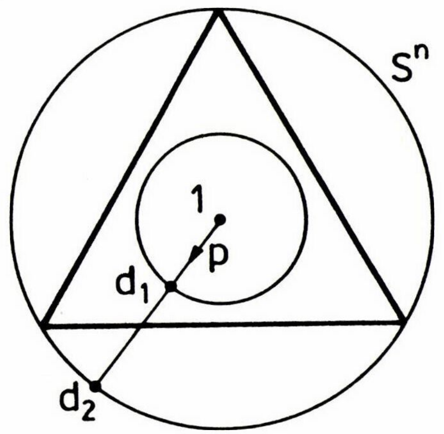
$$\delta_1 = 1 - \alpha r \Pi_1 \quad \text{és} \quad \delta_2 = 1 - R\Pi_1$$

$\delta_2 \le 0$ miatt $R\Pi_1 \ge 1$, így

$$\delta_1 = 1 - \alpha r \Pi_1 = 1 - \frac{\alpha r}{R}(R\Pi_1) \le 1 - \frac{\alpha}{n-1},$$

ahogy állítottuk. Itt felhasználtuk a beírt és körülírt gömbök sugarára vonatkozó arányt.

A fenti becslés lehetővé teszi, hogy egy (alacsonyabb dimenziós) gömbön való minimalizálások sorozatával helyettesítsük az eredeti LP feladatot, és eközben rendelkezésünkre áll egy becslés a célfüggvény aktuális értékére is.



2. ábra

F5. LP feladat zérus optimumú LP feladatba való transzformálása

Tegyük fel, hogy a standard formában adott

$$\text{(F5.1)} \qquad \text{Primál} \quad \begin{array}{r} \min \bar{c}^T\bar{x} \\ \bar{A}\bar{x} \geq \bar{b} \\ \bar{x} \geq 0 \end{array} \qquad \text{Duál} \quad \begin{array}{r} \max \bar{y}^T\bar{b} \\ \bar{A}^T\bar{y} \leq \bar{c} \\ \bar{y} \geq 0 \end{array}$$

feladatpár esetén, mind a primál, mind a duál feladatnak van megengedett megoldása. (Ez az F.1. részben leírtak szerint polinomiális időben eldönthető.) Ismert a dualitás tételből, hogy ekkor midkét feladatnak van optimális megoldása, melyekre a célfüggvényértékek megegyeznek, azaz az $(\bar{y}^T\bar{A} - \bar{c}^T)\bar{x} = 0$ és $\bar{y}^T(\bar{A}\bar{x} - \bar{b}) = 0$ komplementaritási feltételek teljesülnek. Így a fenti primál–duál feladatpár megoldása az alábbi lineáris komplementaritási feladat megoldásával ekvivalens.

$$\text{(F5.2)} \qquad LCP \quad \begin{aligned} \bar{A}\bar{x} - \bar{y}' &= \bar{b} \\ \bar{A}^T\bar{y} + \bar{x}' &= \bar{c} \\ \bar{c}^T\bar{x} - \bar{b}^T\bar{y} &= 0 \\ \bar{x}, \bar{x}', \bar{y}, \bar{y}' &\geq 0 \end{aligned}$$

F5.Lemma.

a, *Ha $P \neq \emptyset$ és $D \neq \emptyset$ akkor LCP megoldható.*
b, *LCP–nek nincs szigorúan pozitív megoldása.*

Bizonyítás.

a, A dualitás tétel közvetlen következménye.
b, A komplementaritási feltételből következik.

Jelöljük az (F5.2) feladat együtthatómátrixát B–vel, jobboldal– illetve változó vektorát d illetve u–val. Így (F5.2) az alábbi alakba írható:

$$Bu = d$$
$$u \geq 0$$

melyet az F.1. részben leírtak szerint tetszőleges $\bar{u} > 0$ vektor rögzítésével a

$$\text{(F5.3)} \qquad \begin{aligned} Bu + \lambda(d - B\bar{u}) &= d \\ u &\geq 0 \\ \min \lambda& \end{aligned}$$

feladat megoldásával oldhatunk meg. Bevezetve az $A = [d - B\bar{u}, B]$, $x = (\lambda, u)$, $b = d$ jelöléseket (F5.3) az alábbi alakot ölti:

$$\text{(F5.4)} \qquad \begin{aligned} \min \xi_1 &= \min e_1 x \\ Ax &= b \\ x &\geq 0 \end{aligned}$$

F5.TÉTEL. *Ha a standard lineáris programozási feladat primál és duál feladatának is van megengedett megoldása, akkor az (F5.1) LP feladatpár az ekvivalens (F5.4) feladatba transzformálható, ahol (F5.4) az alábbi tulajdonságokkal rendelkezik:*

a, *A teljes rangú*
b, $\min \xi_1 = 0$
c, *$Ax = b$–nek adott egy szigorú pozitív megoldása.*

Bizonyítás. Mint láttuk, (F5.1) ekvivalens transzformációkkal (F5.4) feladatba transzformálható. A nyilván teljes rangú, hiszen tartalmaz egységmátrixot (lásd (F5.2)). Nyilvánvaló az is, hogy $\min \xi_1 = 0$, ugyanis (F5.2)–nek nincs pozitív megoldása, de van megoldása, így az F.1. részben elmondottak szerint $\min \xi_1 = 0$, és jogos volt a $\xi_1 \geq 0$ feltétel bevezetése is. (F5.3) alapján ismert egy szigorúan pozitív induló megoldás is.

Így az általánosabbnak tetsző (F5.1) LP feladatok helyett, a speciálisnak tűnő, de valójában ekvivalens (F5.4) feladatokkal elegendő foglalkoznunk. Ezen feladatok egy további transzformációja található az F.6. függelékben.

F6. Projektív transzformáció

Legyen $a \in \mathbf{R}_+^{n-1}$. Legyen $D = \mathrm{diag}(\alpha_j)$ mátrix, így $D^{-1} = \mathrm{diag}(\frac{1}{\alpha_j})$.

Definíció. Legyen $T : \mathbf{R}^{n-1} \to S^n$ projektív transzformáció az alábbi:

$$T(x) = T(\xi_1, \dots, \xi_{n-1}) = \frac{n}{1 + \sum_{i=1}^{n-1} \frac{\xi_i}{\alpha_j}} \left(\frac{\xi_1}{\alpha_1}, \dots, \frac{\xi_{n-1}}{\alpha_{n-1}}, 1 \right)$$
$$= \frac{n}{1 + \mathbf{1}D^{-1}x}(D^{-1}x, 1).$$

Könnyen ellenőrizhetők a T projektív transzformáció alábbi tulajdonságai:

1. T egy–egyértelmű leképzés, és $T(\mathbf{R}^{n-1}_{\oplus}) = S^n$
2. $T^{-1}(\eta_1, \dots, \eta_{n-1}, \eta_n) = \frac{1}{\eta_n}(\alpha_1\eta_1, \dots, \alpha_{n-1}\eta_{n-1}) = \frac{1}{\eta_n} Ya$, ahol $Y = \operatorname{diag}_{0<j<n} (\eta_j)$
3. $T(\mathbf{R}^{n-1}_{\oplus} \cap \{x \mid \xi_j = 0\}) = S^n \cap \{y \mid \eta_j = 0\}$ („lapot lapra képez")
4. A „végtelen" pontok az $\eta_n = 0$ lapra képződnek le
5. $T(a) = \mathbf{1}$
6. Az $\{x \mid u^T x = \mu\}$ hipersíkot a $\{(x, \xi_n) \mid \sum_{j=1}^{n-1} \mu_j \alpha_j \xi_j - \mu\xi_n = 0\}$ altérre képezi, azaz T affin halmazt altérbe képez.
7. Így $T : \{x \in \mathbf{R}^{n-1} \mid Ax = b\} \to \{y \in \mathbf{R}^n \mid [AD, -b]y = 0\}$
8. Ha x az (F5.4) feladat egy megoldása és $y = T(x)$, akkor $e_1 x = e_1 T^{-1}(y)$ miatt $\xi_1 = e_1 x = e_1 T^{-1}(y) = \dfrac{\alpha_1 \eta_1}{\eta_n}$. Mivel $\xi_1 = e_1 x$ minimuma 0, ezért $\min \dfrac{\alpha_1 \eta_1}{\eta_n} = 0$, ami csak $\eta_1 = 0$ esetén teljesülhet, ha a primál optimális halmaz korlátos.

Így beláttuk az alábbi tételt:

F6. TÉTEL. *Az (F5.4) és így az (F5.1) feladat megoldása korlátos optimális halmaz esetén ekvivalens a*

$$\text{(F6.1)} \qquad \begin{aligned} &\min \eta_1 \\ [AD, &-b]y = 0 \\ &y \in S^n \end{aligned}$$

feladat megoldásával, és (F5.4) optimális megoldása (F6.1) optimális megoldásából $x = T^{-1}(y)$ transzformációval nyerhető.

Az $A := [AD, -b]$, $x := y$ jelölések bevezetésével az F5. és F6. tételek az alábbiakban foglalhatók össze:

F6. KÖVETKEZMÉNY. *Az (F5.1) általános lineáris programozási feladat ekvivalens a*

$$\text{(F6.2)} \qquad \begin{aligned} \min \xi_1 &= \min e_1 x \\ Ax &= 0 \\ \mathbf{1}x &= n \\ x &\geq 0 \end{aligned}$$

lineáris programozási feladat megoldásával, ahol

a) *A teljes rangú*
b) *Az* $\mathbf{1}$ *pont megoldása a feladatnak*
c) *Az optimális megoldásban* $\xi_1 = 0$, *azaz* $\min e_1 x = 0$.

Így beláttuk, hogy az általános alakú (F5.1) lineáris proramozási feladat megoldása ekvivalens azzal, hogy egyetlen változót minimalizálunk egy altér és egy szimplex metszetén.

F7. Egy szélsőérték feladat

Legyen $x \in \mathbf{R}^n$, $0 < \rho < 1$. Tekintsük az alábbi feladatot:

$$\min \prod_{j=1}^{n} \xi_j \tag{F7.1}$$

$$\sum_{j=1}^{n} \xi_j = n \tag{F7.2}$$

$$\sum_{j=1}^{n} (\xi_j - 1)^2 \leq \rho^2 \tag{F7.3}$$

$$x \geq 0 \tag{F7.4}$$

A $\rho < 1$ feltétel miatt (F7.4) feltétel elhagyható, ugyanis ez nyilván következménye az (F7.2) és (F7.3) feltételeknek. Most az (F7.1) (F7.2) (F7.3) feladat megoldásához tekintsük a *Kuhn–Tucker feltételeket* [(F7.2) *Lagrange szorzója* legyen α, (F7.3) *Lagrange szorzója* pedig $\beta \geq 0$.]

$$\prod_{j \neq k} \xi_j + \alpha + 2\beta(\xi_k - 1) = 0 \qquad \forall k \tag{F7.5}$$

$$\alpha\left(\sum_{j=1}^{n} \xi_j - n\right) + \beta\left(\sum_{j=1}^{n} (\xi_j - 1)^2 - \rho^2\right) = 0 \tag{F7.6}$$

Észrevételek

1) $\sum_{j=1}^{n} \xi_j = n$ miatt α tetszőleges.
2) $\beta > 0$, mivel $\beta = 0$ esetén (F7.5)-ből $\alpha = \prod_{j \neq k} \xi_j \ \forall k$ adódik, így $\xi_j = \xi_i \ \forall i, j$, melyből (F7.2) miatt $x = \mathbf{1}$ adódik, ami nyilván nem minimum, hanem maximumhely (pl. geometriai egyenlőtlenséggel könnyen belátható). Tehát a továbbiakban fetehető, hogy nem minden ξ_j egyenlő, mivel ilyen pont csak az $x \equiv \mathbf{1}$ pont van a megengedett tartományban.

3) (F7.5)-ből $\prod_{j=1}^{n} \xi_j = \mu$ helyettesítéssel ξ_k-val való szorzás után az alábbi egyenletet kapjuk:

$$\mu + \alpha\xi_k + 2\beta\xi_k^2 + 2\beta\xi_k = 0.$$

Mivel ez ξ_k-ra ($\forall k$ esetén) egy másodfokú egyenlet (k-tól függetlenül azonos együtthatókkal), így a ξ_k koordináták, (a két különböző gyöknek megfelelően) két különböző értéket vehetnek fel. 2) szerint viszont legalább két különböző értéket fel is vesznek. Feltehető, hogy

$$\xi_1 = \cdots = \xi_\ell > \xi_{\ell+1} = \cdots = \xi_n \qquad /\ell\xi_1 + (n-\ell)\xi_n = n/$$

4) (F7.2) és (F7.3)-ból tetszőleges $1 \le \ell < n$ esetén egyszerűen számolható ξ_1 és ξ_n értéke:

$$\xi_1 = 1 + \frac{\sqrt{n-\ell}\rho}{\sqrt{n}} \quad \text{és} \quad \xi_n = 1 - \frac{\sqrt{\ell}\rho}{\sqrt{n(n-\ell)}}.$$

Deriválással (ℓ szerint) könnyen ellenőrizhető, hogy a $\prod_{j=1}^{n} \xi_j$ szorzat értéke $\ell = 1$ esetén minimális, így

$$\xi_1 = 1 + \frac{\sqrt{n-1}\rho}{\sqrt{n}} \quad \text{és} \quad \xi_n = 1 - \frac{\rho}{\sqrt{n(n-1)}}. \tag{F7.7}$$

F7. TÉTEL. *$\rho = r\alpha = \alpha\sqrt{\frac{n}{n-1}}$ esetén az (F7.1),...,(F7.4) feladat optimális megoldása*

$$\xi_1 = 1 + \alpha \quad \xi_2 = \cdots = \xi_n = 1 - \frac{\alpha}{n-1}$$

Bizonyítás. A $\rho = \alpha\sqrt{\frac{n}{n-1}}$ értéket (F7.7)-be helyettesítve kapjuk az állítást.

F8. Parametrikus kvadratikus programozás

Tekintsük a

$$\min \left(\frac{Dz}{\beta} - 1\right)^2 = \delta^2(x,\beta) \tag{F8.1}$$
$$A^T y + z = c$$

kvadratikus programozási feladatot, ahol $x = (\xi_1, \ldots, \xi_n) > 0$ adott vektor, $\beta > 0$, $D = \operatorname{diag}(\xi_j)$, és A teljes rangú. Az (F8.1) feladat az alábbi primál–duál feladatpár megoldásával ekvivalens:

$$\textbf{Primál:}\qquad \frac{1}{2}\delta^2(x,\beta) = \min\left\{\frac{1}{2}z^T\frac{D^2}{\beta^2}z - \frac{x}{\beta}z\right\}$$

$$A^Ty + z = c$$

$$\textbf{Duál:}\qquad \frac{1}{2}\delta^2(x,\beta) = \max -\left\{\frac{1}{2}z^T\frac{D^2}{\beta^2}z + uc\right\}$$

$$Au = 0$$

$$u + \frac{D^2}{\beta^2}z = \frac{x}{\beta}$$

Mivel csak egyenlőséges feltételünk volt, így nincs komplementaritási feltétel, tehát a fenti kvadratikus programozási feladatok megoldása a primál és duál feltételek által adott egyenletrendszerek megoldásával ekvivalens, amely az alábbi

$$u = \frac{x}{\beta} - \frac{D^2}{\beta^2}z \tag{F8.2}$$

$$AD^2z = \beta Ax \tag{F8.3}$$

$$A^Ty + z = c \tag{F8.4}$$

Így tulajdonképpen az (F8.3) és (F8.4) által definált egyenletrendszereket kell csak megoldanunk. Ezek megoldása könnyen előállítható a

$$z = c - A^Ty \tag{F8.5}$$

$$y = (AD^2A^T)^{-1}AD^2c - \beta(AD^2A^T)^{-1}Ax \tag{F8.6}$$

formulákkal. (Hasonlítsuk össze az F3 tétel képleteivel.)

Így az (F8.1) feladat megoldása algebrai alapműveletekkel explicite előállítható. Megjegyezzük, hogy ezt klasszikus módon, z kifejezésével és deriválással is elvégezhettük volna.

IRODALOM

[1] ANSTREICHER, K. M., "A monotonic projective algorithm for fractional programming ", *Algoritmica* 1 (1986), 483–498.

[2] BARNES, E. R., "A variation on Karmarkar's algorithm for solving linear programming problems ", *Mathematical Programming* 36 (1986), 174-182.

[3] DANTZIG, G. B., *Linear Programming and Extensions* (Princeton University Press, Princeton, 1963).

[4] DIKIN, I. I., "Iterative solution of problems of linear and quadratic programming ", *Dokl. Akad. Nauk SSSR* 174 (1967), 747–748.

[5] GILL, P., MARXEN, A., MURRAY, W., SAUNDERS, M. and WRIGHT, M., "Newton barrier methods for LP: Some implementational aspects ", Department of Operations Research, Stanford University, ORSA/TIMS, Washington DC, April 25–27 (1988).
[6] GILL, P., MURRAY, W., SAUNDERS, M., TOMLIN J. and WRIGHT, M., "On projected Newton barrier methods for linear programming and an equivalence to Karmarkar's projective method ", *Mathematical Programming* 36 (1986), 183–209.
[7] IRI, M. and IMAI, H., "A multiplicative barrier function method for linear programming ", *Algoritmica* 1 (1986), 455–482.
[8] KARMARKAR, N., "A new polinomial-time algorithm for linear programming ", *Combinatorica* 4 (1984), 373–395.
[9] KARMARKAR, N. and RAMAKRISHNAN, K., "Implementation and computational results of the Karmarkar algorithm for linear programming, using an iterative method for computing projections ", AT&T Bell Laboratories, Murray Hill, New Jersey (1988).
[10] KLAFSZKY, E. és TERLAKY, T., "Az ellipszoid módszerről ", *Szigma* (1988).
[11] KLEE, V. and MINTY, G. J., "How good is the simplex algorithm? ", in Inequalities III, (ed. Shisha, O.) (Academic Press, Inc., New York, 1972), pp. 159–172.
[12] KOJIMA, M., "A polinomial time algorithm for a class of linear complementarity problems ", *Mathematical Programming,* (megjelenés alatt).
[13] MEGIDDO, N., "On finding primal and dual optimal bases ", IBM Almaden Research Center, RJ 6328 (61997) 7/11/88.
[14] MEGIDDO, N., "Switching from a primal-dual Newton algorithm to a primal-dual (interior) simplex algorithm ", IBM Almaden Research Center, RJ 6327 (61996) 7/11/88.
[15] MEGIDDO, N., "Boundary behavior of interior point algorithms in linear programming ", IBM Almaden Research Center, RJ 5319 (54679) 9/30/86.
[16] RENEGAR, J., "A polinomial – time algorithm, based on Newton's method for linear programming ", *Mathematical Programming* 40 (1988), 59–93.
[17] ROOS, C., "On Karmarkar's projective method for linear programming ", Delft University of Technology, Report 85–23.
[18] ROOS, C., "Linear programming along the trajectory of the problem in polynomial time ", Delft University of Technology, Report 88–17.
[19] ROOS, C. and VIAL, J-Ph., "A polynomial method of approximate centers for linear programming ", Delft University of Technology, Report 1988.
[20] SCHRIJVER, A., *Theory of Linear and Integer Programming* (John Wiley and Sons, 1986).
[21] SMALE, S., "On the average number of steps in the simplex method ", in Mathematical Programing, The State of the Art, Bonn 1982 (Bachem, A., Grötchel, M. and Korte, B. eds.) (Springer, Berlin, 1983), pp. 530–539.
[22] SONNEVEND, Gy. and STOER, J., "Global ellipsoidal approximations and Homotopy methods for solving convex analytic programs ", *Applied Mathematics and Optimization,* (megjelenés alatt).
[23] TERLAKY, T., "Az lp. programozásról ", *Alkalmazott Matematikai Lapok* 6 (1980), 27–63.
[24] TODD, M. J., "Karmarkar as Dantzig-Wolfe ", *School of O.R. and I.E., Cornell University, Technical Report, No. 782,* (1988), Itacha, New York.
[25] TOMLIN, J. A., "An experimental approach to Karmarkar's projective method for linear programming ", *Mathematical Programming Study* 31 (1987), 175–191.
[26] YE, Y., "Karmarkar's algorithm and the ellipsoid method ", *Operations Research Letters* 6 (1987), 177–182.
[27] YE, Y. and TODD, M. J., "Containing and shrinking ellipsoids in the path – following algorithm ", School of OR and IE, Cornell University, Research Report 1987.
[28] YE, Y., "Further development on the Interior Algorithm for convex quadratic programming ", Stanford University, Stanford, Research Report (1987).

(Beérkezett: 1989. május 28.)

TERLAKY TAMÁS
ELTE TTK, OPERÁCIÓKUTATÁSI TANSZÉK
1088 BUDAPEST MÚZEUM KRT. 6-8.

ON KARMARKAR TYPE ALGORITHMS

T. Terlaky

An overview of interior point methods for linear programming is presented in this paper. Since Karmarkar's 1984 paper interior point methods have been widely considered and examined both from theoretical and implementational point of view.

Three main groups of interior point methods are distinguished: Projective methods with a potencial function (Karmarkar), affine methods (Dikin and Barnes) and path following methods. A simple, easy to understand variant for all of the three groups are presented.

The solution of quadratic subproblems, projective transformations and the first phase (how to find a strictly positive starting solution) are presented as well.

Alkalmazott Matematikai Lapok **15** (1990–91) 163–179

PARCIÁLIS FÜGGŐSÉG RELÁCIÓS ADATBÁZISOKBAN*

DEMETROVICS JÁNOS, KATONA GYULA és MIKLÓS DEZSŐ

Budapest

Az *Armstrong féle funkcionális függőség* fogalmának gyengítésével kapjuk a relációs adatbázisokon értelmezett parciális függőség fogalmát. Megmutatjuk, hogy a parciális függőségek jól jellemezhetők az adatbázis attributum-halmazán definiált parciális függvények által alkotott parciálisan rendezett halmaz lezárásaival. Más oldalról, szükséges és elégséges feltételt adunk ahhoz, hogy egy ilyen lezáráshoz találjunk az adott attributum-halmazon olyan adatbázist, ami által definiált parciális függőségek pont az adott lezárást generálják. Megvizsgáljuk azt is, hogy milyen tulajdonságú függőséget definiálnak a parciális függőségek az attributumok részhalmazai által alkotott parciálisan rendezett halmazon és hogy ezek miképpen realizálhatók adatbázisokkal.

1. Bevezetés

Ebben a dolgozatban relációs adatbázisok egy újfajta függőségét, az ún. *parciális függőséget* vizsgáljuk, amely először [4]-ben került bevezetésre. Az adatbázisok modelljének itt a mátrixot tekintjük. A mátrix oszlopai, azaz az egyes lehetséges tulajdonságok nevei az adatbázis *attributumai*, míg a sorok, azaz az egyes elemek tulajdonságainak az összessége az adatbázis *rekordjai*.

Jelöljük Ω-val az attributumok (azaz a mátrix oszlopainak) halmazát. Legyen $A \subseteq \Omega$, $b \in \Omega$. A függőség hagyományos definíciója szerint (Armstrong [1], Codd [2]) akkor mondjuk, hogy b *funkcionálisan függ* A-tól, ha az A-ban álló adatok már egyértelműen meghatározzák a b-ben lévő adatot is. Sok gyakorlati esetben egy ilyen függés csak majdnem teljesül. Pl. egy nem túl nagy adathalmaz esetén (egy iskola tanulóinak adatai) a név rendszerint meghatározza a személyt, azaz minden más adatot, azaz ettől az egyelemű A adathalmaztól minden más b adat funkcionálisan függ. „Csak” néhány kivétel van. A *Kovácsok, Tóthok*, esetleg két unokatestvér azonos névvel. Tehát az A-ban lévő legtöbb adat már egyértelműen meghatározza a többi b oszlop adatait, de néhány más kivételes adat nem. Ezekhez még egy (vagy több) azonosító adat is kell. Ez motiválja a parciális függőség fogalmának bevezetését. Heurisztikusan szólva akkor mondjuk, hogy az A oszlophalmazban álló adatsorozattól b parciálisan függ, ha minden sorban (rekordban), ahol A-ban ezen értékek állnak, b-ben ugyanaz az érték van.

Két indokot tudunk felsorakoztatni e struktúra vizsgálata mellett. Az egyik elméleti. Minden funkcionális függőségi modell mögött ott van a parciális függőségek modellje is. Ez tehát egy finomabb modell, ami azonban még mindig csak modell

*Az 2575–ös Országos Tudományos Kutatási Alap által támogatva

a valódi adatbázis és a funkcionális függőségi modell között. Tehát a rá vonatkozó eredmények más megvilágításba helyezhetik a funkcionális függőségek modelljét.

A másik indok gyakorlati. Tekintsünk egy példát. Tegyük fel, hogy $\Omega = \{x_1, x_2, x_3, x_4\}$, x_2 funkcionálisan függ x_1-től, x_4 pedig x_3-tól. Az egész 4-oszlopú mátrix tárolása helyett elegendő x_1 és x_3 oszlopait tárolni, emellett két kis mátrix tárolására van szükség, amely megadja a két funkcionális függést. Jól látható, hogy ez jelentős memóriamegtakarítást eredményez. Ha x_2-nek x_1-től, illetve x_4-nek x_3-tól való függése csak néhány kivételes érték miatt nem teljesül, akkor még mindig sok tárolóhelyet takaríthatunk meg, ha különválasztjuk a kivételes x_1, x_3 párokat, és mellettük felsoroljuk a hozzájuk tartozó x_2-t és x_4-et, a többivel meg úgy járunk el, mint a „rendesen" függő esetben. Tehát ez a finomabb modell helymegtakarítást tehet lehetővé olyankor, amikor a funkcionális függőség segítségével ezt nem érhetjük el.

A dolgozat 2. fejezetében megadjuk a szükséges definíciókat, a parciális függőség definícióját és különböző modelljeit. A 3. fejezetben megvizsgáljuk a parciális függőségek realizálhatóságának kérdését, illetve azt, hogy a parciális függőségek milyen struktúrát hoznak létre az attributumok halmazán.

2. A parciális függőség definíciója és különböző modelljei

Az adatbázis modellje itt az M mátrix, amelynek a sorai számát m és az oszlopai számát n jelöli. Az oszlopokat más néven attributumoknak nevezzük, ezek halmazát Ω jelöli. Az adatbázis elemein a továbbiakban a mátrix elemeit értjük.

Az adatbázisokon definiált függőségek talán legismertebbike a *funkcionális függőség* (vagy röviden csak függőség), amelyet a következő módon definiálunk: ha $A, B \subseteq \Omega$ két részhalmaza az attributumok halmazának, akkor azt mondjuk, hogy B funkcionálisan függ A-tól, amennyiben az adatbázisnak nincs két olyan sora, melyben az elemek megegyeznek az A halmazon, de legalább egy helyen különböznek a B halmazon. Jelölése $A \longrightarrow B$. Más szavakkal ez azt jelenti, hogy az adatbázis A-beli oszlopaiban lévő adatok meghatározzák a B-beli oszlopokban levő adatokat. Ebben az esetben nem lényeges, hogy mik a konkrét elemek az egyes attributumoknál, csak az, hogy ha azok két különböző sorban megegyeznek az A-beli attributumoknál, akkor meg kell egyezniük a B-belieknél is. Az egyes adatbázisokat nyilván nem különböztetik meg egyértelműen az általuk meghatározott függőségek, de azért sok lényeges információt tartalmaznak. A funkcionális függőségek megadott halmaza felfogható, mint azon (konkrét) mátrixok modellje, amelyek kielégítik ezeket a funkcionális függőségeket.

A funkcionális függőség gyengítésével kaphatjuk a sokkal gyengébb és konkrétabb *parciális függőséget*, amely már figyelembe veszi az egyes konkrét elemeket is. Tegyük fel, hogy egy adatbázisban az a_1 és a_2 oszlopokban előforduló elemek nem határozzák meg mindig egyértelműen a b oszlopban előforduló elemet, ám ha egy sorban az a_1 attributum értéke r_1 és a_2 attributum értéke r_2, akkor a b attributum értéke r_3, azaz egyértelműen meghatározott. Ezt a tényt úgy fogjuk jelölni, hogy $(a_1, a_2; r_1, r_2) \to (b; r_3)$. Az általános definícióhoz szükségünk van a *parciális függvény* fogalmára. Ha az a_i attributum lehetséges értékeinek a halmaza az adott adatbázis esetén D_i és $r_i \in D_i$, akkor $\alpha = (a_1, \dots, a_k;\ r_1, \dots, r_k)$-t parciális függvénynek hívjuk, amennyiben az a_i elemek az Ω különböző elemei. Az $\alpha = (a_1, \dots, a_k;\ r_1, \dots, r_k)$ parciális függvény *értelmezési tartománya* az $\text{ÉT}(\alpha) = \{a_1, \dots, a_k\}$ halmaz, *értékkészlete* pedig az $\text{ÉK}(\alpha) = \{r_1, \dots, r_k\}$ halmaz. Azt mondjuk, hogy a $\beta = (b_1, \dots b_l; s_1, \dots, s_l)$ parciális függvény *függ* az α-tól (és úgy jelöljük, hogy $\alpha \to \beta$), ha minden olyan sorban, amelyben az a_i oszlopokban r_i elemek vannak $(1 \le i \le k)$ a b_j oszlopokba s_j elemek kerülnek $(1 \le j \le l)$.

Természetesen a fenti definíció nem értelmezhető az összes parciális függvényre egy adott adatbázis esetében, csak azokra, amelyek *koherensek*, azaz előfordulnak az adatbázis valamelyik sorában (vagy, kényelmi szempontokból feltehetjük, hogy az „ értelmetlen", nem koherens parciális függvényektől az összes koherens parciális függvény függ). Mindazonáltal a koherens parciális függvények egy szép struktúrát alkotnak. Ennek megadásához szükségünk van egy definícióra. Ha $\alpha = (a_1, \dots, a_k;\ r_1, \dots, r_k)$ és $\beta = (b_1, \dots, b_l;\ s_1, \dots, s_l)$ két parciális függvény, akkor $\alpha \subset \beta$, azaz α része β-nak, amennyiben $\{a_1, \dots, a_k\} \subseteq \{b_1, \dots, b_l\}$ és $a_i = b_j$-ből következik, hogy $r_i = s_j$. Azt mondjuk, hogy a parciális függvények egy P halmaza *leszálló* struktúrájú, ha az alábbi két tulajdonságot kielégíti:

(2.1) ha $\alpha \in P$ és $\beta \subset \alpha$, akkor $\beta \in P$.

(2.2) minden $\alpha \in P$-re létezik egy olyan $\gamma \in P$, amelyre $\alpha \subseteq \gamma$, $\gamma = (c_1, \dots, c_n; \dots)$ és $\{c_1, \dots, c_n\} = \Omega$.

Nyilvánvaló, hogy egy adott adatbázis esetében a koherens parciális függvények halmaza leszálló lesz.

Parciális függvények uniója és metszete is definiálható, (még ha az első csak bizonyos párokra is) habár kicsit más módon, mint az egyszerű halmazok esetében. Legyen $\alpha = (a_1, \dots, a_k;\ r_1, \dots, r_k)$ és $\beta = (b_1, \dots, b_l;\ s_1, \dots, s_l)$ két parciális függvény. Kettőjük metszete az a γ parciális függvény, amelynek az értelmezési tartománya azon c attributumok halmaza, amelyek α értelmezési tartományának is és β értelmezési tartományának is elemei (mondjuk $c = a_i = b_j$) és amelyekre $r_i = s_j$. A γ parciális függvény természetesen az $r_i = s_j$ értéket veszi fel a c helyen. Az unió értelmezése egy kicsit több figyelmet kíván. Az $\alpha = (a_1, \dots, a_k;\ r_1, \dots, r_k)$ és $\beta = (b_1, \dots, b_l;\ s_1, \dots, s_l)$ parciális függvényeknek csak akkor értelmezzük az unióját, ha a metszetük értelmezési tartománya megegyezik az értelmezési tartományaik metszetével. Nem nehéz megállapitani, hogy ez pontosan akkor áll fenn, ha a két parciális függvény a mindkettőjük értelmezési tartományában előforduló attributumokon ugyanazt az értéket veszi fel. Ebben az esetben értelmezhetjük (és értelmezzük) az uniójukként azt a γ parciális függvényt, amelynek az értelmezési tartománya a két eredeti függvény értelmezési tartományának uniója, és amely egy c attributumon a következő értéket veszi fel:

$$
\begin{aligned}
r_i - t, &\quad \text{ha } c = a_i \text{ és } c \notin \{b_1, \ldots, b_l\},\\
s_j - t, &\quad \text{ha } c = b_j \text{ és } c \notin \{a_1, \ldots, a_k\},\\
r_i = s_j - t, &\quad \text{ha } c = a_i = b_j.
\end{aligned}
$$

Az eddig megadott alapvető definíciók segítségével megadhatjuk a parciális függvények lezártját, hasonlóan, mint ahogy a funkcionális függőségek esetében az Ω egy részhalmazának a lezártja adott. A definíció világos megadása érdekében azonban először egy lemmát mondunk ki a parciális függőségek tulajdonságairól.

2.1. LEMMA. *Ha α, β, γ, δ egy adatbázishoz tartozó parciális függvények, akkor*
(2.3) $\alpha \to \alpha$;
(2.4) $\alpha \to \beta$ és $\beta \to \gamma$-ból következik, hogy $\alpha \to \gamma$;
(2.5) $\alpha \subseteq \gamma$, $\delta \subseteq \beta$ és $\alpha \to \beta$-ból következik, hogy $\gamma \to \delta$;
(2.6) ha $\alpha \to \beta$, $\gamma \to \delta$ és $\alpha \cup \gamma$ létezik és koherens, akkor $\beta \cup \delta$ is létezik és koherens, és $\alpha \cup \gamma \to \beta \cup \delta$.

Bizonyítás. (2.3) nyilvánvaló. A továbbiakban legyen $\alpha = (a_1, \ldots, a_k; r_1, \ldots, r_k)$, $\beta = (b_1, \ldots, b_l; s_1, \ldots, s_l)$, és $\gamma = (c_1, \ldots, c_p; t_1, \ldots, t_p)$. Ha $\alpha \to \beta$ és $\beta \to \gamma$, akkor az adatbázis minden olyan sorában, ahol az a_i oszlopokban r_i elemek állnak a b_j oszlopokba s_j elemek kerülnek, és így a c_h oszlopokba t_h elemek lesznek, azaz $\alpha \to \gamma$, ami bizonyítja (2.4)-et.

(2.5) bizonyításához elég belátni, hogy $\alpha \subseteq \gamma$-ból következik, hogy $\gamma \to \alpha$, ami éppen úgy nyilvánvaló a definíciókból, mint (2.3). Akkor viszont az itteni feltételekből $\gamma \to \alpha$, $\alpha \to \beta$ és $\beta \to \delta$ következik, amelyek a $\to$ reláció (2.4) szerinti tranzitív volta miatt $\gamma \to \delta$-t adják.

(2.6) bizonyításakor legyen $\delta = (d_1, \ldots, d_q; u_1, \ldots, u_q)$. Ha α és γ két parciális függvény, akkor a „$\alpha \cup \gamma$ létezik és koherens" feltétel azt jelenti, hogy léteznek sorok, amelyek az a_i oszlopokban r_i értékeket vesznek fel és a c_h oszlopokban t_h értékeket, beleértve azt is, hogy amennyiben valamely $a_i = c_h$, akkor r_i megegyezik t_h-val. Ezekben a sorokban viszont a $\alpha \to \beta$ feltételből következően a b_j oszlopokban s_j értékeknek és az d_g oszlopokban u_g értékeknek kell szerepelniük. Ez természetesen maga után vonja, hogy amennyiben valamely $b_j = d_g$, akkor s_j-nek meg kell egyeznie u_g-vel, azaz a $\beta \cup \delta$ parciális függvény létezik (definiált). Koherens is lesz, hiszen minden olyan sor, amely az $\alpha \cup \gamma$ parciális függvényt tartalmazza, tartalmazza $\beta \cup \delta$-t is. Természetesen az eddig elmondottak a $\alpha \cup \gamma \to \beta \cup \delta$ relációt is bizonyítják. □

Ezen lemma birtokában térjünk rá a parciális függvények lezárásának a definíciójára. Adott adatbázis esetén egy α tetszőleges (koherens) parciális függvényhez rendeljük hozzá $L(\alpha)$-t, azt a legnagyobb β parciális függvényt, amelyre $\alpha \to \beta$. Itt a legnagyobb alatt azt a parciális függvényt értjük, amelynek az értelmezési tartománya a legnagyobb. Ez nyilván létezik az adott esetben, mivel ha $\alpha \to \beta_1$ és

$\alpha \to \beta_2$, akkor a β_1 és β_2 parciális függvényeknek a (2.6) pont értelmében létezik az uniója és $\alpha \to \beta_1 \cup \beta_2$. Tehát $L(\alpha)$ nem más, mint

$$L(\alpha) = \bigcup_{\alpha \to \beta} \beta . \tag{2.7}$$

Természetesen a fenti unióban nem kell az összes parciális függvényt szerepeltetni, elég azokat, amelyek egyetlen attributumot tartalmaznak. $L(\alpha)$ ezek birtokában is megkapható:

$$L(\alpha) = \bigcup_{\alpha \to (b;s)} (b;s) . \tag{2.8}$$

Ezen L függvényről az alábbi tulaldonságokat állíthatjuk.

2.2. LEMMA. *Ha α és β egy adott adatbázishoz tartozó két parciális függvény, akkor*

$$\alpha \subseteq L(\alpha) ; \tag{2.9}$$

$$\alpha \subseteq \beta - \textit{ból következik, hogy } L(\alpha) \subseteq L(\beta) ; \tag{2.10}$$

$$L(L(\alpha)) = L(\alpha) . \tag{2.11}$$

Bizonyítás. $L(\alpha)$ definíciója szerint (2.9) egyszerű következménye (2.3)-nak. (2.5) szerint $\alpha \subseteq \beta$-ból következik, hogy $\alpha \to \gamma$ maga után vonja a $\beta \to \gamma$ relációt is. Így minden olyan parciális függvény, amely $L(\alpha)$ (2.7) szerinti definíciójában az unióban szerepel, szerepel $L(\beta)$ definíciójában is, azaz $L(\alpha) \subseteq L(\beta)$ fennáll.

(2.11) bizonyításához (2.9) szerint elég belátni, hogy $L(L(\alpha)) \subseteq L(\alpha)$. Tegyük fel, hogy a c attributum eleme $L(L(\alpha))$ értelmezési tartományának, azaz valamely $(c; s)$ parciális függvény szerepel $L(L(\alpha))$ (2.8) szerinti előállításában. Ekkor, definíció szerint, $L(\alpha) \to (b; s)$. Ugyanakkor, megint csak definíció szerint $\alpha \to L(\alpha)$, és így a $\to$ reláció 2.1. lemmában bizonyított tranzitivitása miatt $\alpha \to (b; s)$, tehát a $(b; s)$ parciális függvény szerepel $L(\alpha)$ (2.8) tipusú előállításában. □

Egy olyan L függvényt, amely parciális függvények egy leszálló tulajdonságú halmazán van értelmezve és kielégíti a (2.9)-(2.11) tulajdonságokat, *függvény-lezárásnak* [4] nevezünk. Bizonyítottuk tehát, hogy az a $\alpha \to L(\alpha)$ függvény, amely egy adott adatbázis esetén az α parciális függvényhez hozzárendeli annak (2.7) szerinti *lezártját*, egy függvénylezárás. Ezen fejezet további részében a funkcionális függőségből kapott szokásos matematikai lezárás és maga a függőség kapcsolatának [3, 4]-ben elvégzett ill. található vizsgálatait ismételjük meg parciális függőségekre.

Először is megjegyezzük, hogy egy (adott adatbázishoz tartozó) parciális függőségből kapott L függvény-lezárás egyértelműen meghatározza a parciális függőséget.

2.3. LEMMA. *$\alpha \to \beta$ pontosan akkor áll fenn, ha $\beta \subseteq L(\alpha)$.*

Bizonyítás. Nyilvánvaló $L(\alpha)$ (2.7) definíciójából. □

Egy α parciális függvényről azt mondjuk, hogy egy adott adatbázishoz tartozik, ha α koherens az adatbázisban. Legyen T (egy adott adatbázishoz tartozó) parciális függőségekből alkotott rendezett párok családja. Azt mondjuk, hogy a T egy *függőségi család*, ha a T-ből definiált

$$\alpha \to \beta \text{ akkor és csak akkor, ha } (\alpha, \beta) \in T$$

$\to$ reláció kielégíti a (2.3)–(2.6) tulajdonságokat.

A 2.1. lemmában láttuk, hogy ha a T családot úgy definiáljuk, hogy $(\alpha, \beta) \in T$ pontosan akkor, ha β függ α-tól, akkor így egy függőségi családot kapunk. A 2.2. lemmában a lezárás tulajdonságairól bizonyítottak abból következtek, amit a $\to$ relációról a 2.1. lemmában bebizonyítottunk, tehát egy L függvény-lezárás függvényt minden függőségi családhoz definiálhatunk.

2.4. TÉTEL. *Az ugyanazon alaphalmazon értelmezett T függőségi családok és L függvény-lezárási operációk között egy egyértelmű megfeleltetés adható meg az alábbiak szerint:*

$$T \to_1 L(\alpha) = \bigcup_{(\alpha,(b;s)) \in T} (b; s). \tag{2.12}$$

A megfeleltetés inverzét az alábbiak szerint kapjuk.

$$L \to_2 T = \{(\alpha, \beta) : \beta \subseteq L(\alpha)\}. \tag{2.13}$$

Bizonyítás. Azt, hogy az $\to_1$ megfeleltetés által adott operáció függvény-lezárás lesz, lényegében a 2.2. lemmában bizonyítottuk. Most belátjuk, hogy a (2.13) által adott családok függőségi családok lesznek, azaz kielégítik a (2.3)–(2.6) feltételeket. (2.3) abból következik, hogy L kielégíti a (2.9) feltételt. Ha $\beta \subseteq L(\alpha)$ és $\gamma \subseteq L(\beta)$, akkor (2.10) és (2.11) szerint $\gamma \subseteq L(\beta) \subseteq L(L(\alpha)) = L(\alpha)$, azaz (2.4) fennál T-re.

(2.5) a definíciókból és a (2.10) tulajdonságból következik. Ha $\alpha \subseteq \gamma$, $\delta \subseteq \beta$ és $\beta \subseteq L(\alpha)$, akkor $\delta \subseteq \beta \subseteq L(\alpha) \subseteq L(\gamma)$, azaz $(\gamma, \delta) \in T$.

Végül (2.6) bizonyítása egy kicsit bonyolultabb gondolatmenetet igényel. Mivel $\alpha \cup \gamma$ létezik és koherens, $L(\alpha \cup \gamma)$ is létezik és koherens. A parciális függvények uniója definíciójából könnyen látható, hogy ha két parciális függvénynek létezik közös majoránsa, akkor az uniójuk is létezik, sőt ha a majoráns koherens, akkor az unió is az lesz. Márpedig ebben az esetben $\beta \subseteq L(\alpha) \subseteq L(\alpha \cup \gamma)$ és $\delta \subseteq L(\gamma) \subseteq L(\alpha \cup \gamma)$, azaz $\beta \cup \delta$ létezik, koherens, és természetesen része $L(\alpha \cup \gamma)$-nak, ami éppen (2.6) konklúziója.

Azt kell még belátnunk, hogy a két megfeleltetés egymás inverze. Tekintsük először a T_1 függőségi családhoz hozzárendelt L függvény-lezárásból kapott T_2 függőségi családot. Ha egy (α, β) pár eleme T_1-nek, akkor β minden attributuma a β

által ott felvett értékkel szerepel $L(\alpha)$ (2.12) előállításában T_1 (2.5) tulajdonsága szerint, azaz $\beta \subseteq L(\alpha)$, ami maga után vonja, hogy $(\alpha,\beta) \in T_2$. Fordítva, ha egy (α,β) pár eleme a T_2 családnak, akkor $\beta \subseteq L(\alpha)$, azaz β minden b attributumára és a β által ott felvett s értékre $\alpha \to (b;s)$, azaz a T_1 család (2.6) tulajdonsága szerint $\alpha \to \beta$, $(\alpha,\beta) \in T_1$.

Végül legyen az L_1 lezárásból (2.13) szerint kapott T családhoz rendelt lezárás L_2 és α egy tetszőleges koherens parciális függvény. $L_1(\alpha)$ minden b attributumára és az ott felvett s értékére $(b;s) \subseteq L_1(\alpha)$, így $(\alpha,(b;s)) \in T$. Ebből viszont a (2.12) definíció szerint $L_2(\alpha) = L_1(\alpha)$. □

Az eddigiekben tehát láttuk, hogy a parciális függőségek egyértelműen jellemezhetők a megfelelő függvény-lezárással (2.2. lemma), vagy a (lényegében a függőséggel azonos) megfelelő függőségi családdal (2.4. tétel). A függvény-lezárás nyilván egyszerűbb struktúra, mint a függőségi család, így ennek vizsgálatával foglalkozunk a fejezet hátralevő részében. Ugyanakkor megjegyezzük, hogy nem minden, a parciális függvényeken értelmezett függvény-lezárás (és így az annak megfelelő parciális függőség) áll elő mint valamely adatbázishoz tartozó függvény-lezárás. Ennek a kérdésnek a közelebbi vizsgálatára a 3. fejezetben térünk ki.

Egy α parciális függvény (egy adott adatbázisban, ill. annak L függvény-lezárása szerint) *zárt*, ha $\alpha = L(\alpha)$. A funkcionális függőségek esetében az Ω zárt részhalmazai egyértelműen meghatározzák a lezárást ill. a függőséget [3]. Hasonló állítás igaz a parciális függőségek esetére is.

2.5. TÉTEL. *Legyen $\mathcal{G}$ ugyanazon Ω alaphalmazon értelmezett parciális függvények egy családja. Ekkor $\mathcal{G}$ pontosan akkor lesz egy valamely L függvény-lezáráshoz tartozó zárt parciális függvények családja ha*

(2.14) minden $\alpha \in \mathcal{G}$-re létezik egy olyan $\gamma \in \mathcal{G}$, amelyre $\alpha \subseteq \gamma$, $\gamma = \{c_1,\ldots,c_n;\ldots\}$ és $\{c_1,\ldots,c_n\} = \Omega$;

(2.15) $\alpha,\beta \in \mathcal{G}$-ből következik, hogy $\alpha \cap \beta \in \mathcal{G}$.

Bizonyítás. Először lássuk be, hogy az adott függvény-lezáráshoz tartozó zárt függvények családja kielégíti (2.14) és (2.15)-öt. (2.2) szerint minden α-ra, és persze így minden zárt α-ra is létezik egy (2.14)-et kielégítő γ. (2.9) szerint erre a γ-ra $\gamma \subseteq L(\gamma)$, de persze $L(\gamma) \subseteq \gamma$ is fennáll, így γ zárt.

Legyenek most α és β zárt parciális függvények. Ekkor $\alpha = L(\alpha)$ és $\beta = L(\beta)$. $\alpha\cap\beta \subseteq \alpha$, és így (2.10)-ből következik, hogy $L(\alpha\cap\beta) \subseteq L(\alpha) = \alpha$. Hasonlóan látható be, hogy $L(\alpha \cap \beta) \subseteq L(\beta) = \beta$, és így a parciális függvények metszete definíciójából könnyen láthatóan $L(\alpha \cap \beta) \subseteq \alpha \cap \beta$. Ennek a fordítottja következik (2.9)-ből, így $L(\alpha \cap \beta) = \alpha \cap \beta$, azaz $\alpha \cap \beta$ is zárt parciális függvény az adott lezárásban.

Legyen most $\mathcal{G}$ egy tetszőleges parciális függvénycsalád, amely kielégíti (2.14) és (2.15)-öt. Meg kell adnunk egy olyan L függvény-lezárást, amelyben a zárt függvények pontosan $\mathcal{G}$ elemei. Legyen a függvény-lezárás azokon a parciális függvényeken értelmezve, amelyeket $\mathcal{G}$ valamely maximális tagja (azaz olyan tagja, amelynek az értelmezési tartománya Ω) tartalmaz. Ezekre a parciális függvényekre legyen

$$L(\alpha) = \bigcap_{\alpha \subseteq \beta \in \mathcal{G}} \beta.$$

L értelmezési tartományának választása következtében a fenti metszet nem üres. (2.9) és (2.10) tulajdonságok a definícióból következnek. Ha $\alpha \in \mathcal{G}$, akkor α szerepel a metszetben, és a metszet minden más tagja nagyobb α–nál, tehát $L(\alpha) = \alpha$. (2.15) szerint minden α–ra $L(\alpha) \in \mathcal{G}$, tehát $L(L(\alpha)) = L(\alpha)$, azaz (2.11) is fennál, L egy függvény-lezárás. Azt kell még belátni, hogy egy α parciális függvény pontosan akkor zárt, ha benne van $\mathcal{G}$–ben. Mint láttuk, a $\mathcal{G}$ α elemeire $\alpha = L(\alpha)$ igaz, így ezek zártak. A zárt elemekre viszont $\alpha = L(\alpha)$, és mivel $L(\alpha)$ benne van $\mathcal{G}$–ben, így α is. □

A fenti tételben a (2.15) tulajdonságot úgy is megfogalmazhatjuk, hogy $\mathcal{G}$ zárt a metszetre. A (2.14) és (2.15) tulajdonságokat kielégítő parciális függvény családokat *parciális metszet-félhálóknak* fogjuk nevezni.

Láttuk, hogy a zárt parciális függvények családja mindig parciális metszet-félhálókat alkotnak. A függvény-lezárásból nyilván egyértelműen kapható meg a metszet-félháló. Bebizonyítjuk ennek az ellenkezőjét is.

2.6. TÉTEL. *Az ugyanazon alaphalmazon értelmezett $\mathcal{G}$ parciális metszet-félhálók és L függvény-lezárási operációk között egy egyértelmű megfeleltetés adható meg az alábbiak szerint:*

$$\mathcal{G} \to_1 L(\alpha) = \bigcap_{\alpha \subseteq \beta \in \mathcal{G}} \beta. \tag{2.16}$$

A megfeleltetés inverzét az alábbiak szerint kapjuk.

$$L \to_2 \mathcal{G} = \{\alpha : L(\alpha) = \alpha\}. \tag{2.17}$$

Bizonyítás. Már beláttuk az előző tételben, hogy a (2.17) definíció egy parciális metszet-félhálót, a (2.16) pedig egy függvény-lezárási operációt hoz létre és (2.17) megadja az összes parciális metszet-félhálót. Lássuk be, hogy $\to_2$ injektív. Legyen L_1 és L_2 két különböző függvény-lezárás ugyanazon az alaphalmazon és $\mathcal{G}_1$ és $\mathcal{G}_2$ a megfelelő parciális metszet-félhálók. Akkor létezik egy α parciális függvény, amelyre vagy $L_1(\alpha)$ és $L_2(\alpha)$ különbözők, vagy pedig csak az egyik van definiálva. Ha $L_1(\alpha)$ létezik és $L_2(\alpha)$ nem, akkor az előbbi nem lehet eleme $\mathcal{G}_2$–nek, hisz akkor L_2–ben lenne egy zárt parciális függvény, ami tartalmazza α–t, és így $L_2(\alpha)$ is értelmezve lenne. Tehát ekkor $L_1(\alpha) \in \mathcal{G}_1$ és $L_2(\alpha) \notin \mathcal{G}_2$, így $\mathcal{G}_1 \neq \mathcal{G}_2$. Ha mind $L_1(\alpha)$, mind $L_2(\alpha)$ léteznek, csak különbözőek, akkor létezik egy $(b; s)$ egyetlen helyen értelmezett parciális függvény, amelyet $L_1(\alpha)$ és $L_2(\alpha)$ pontosan egyike tartalmaz. Legyen $(b; s) \subseteq L_1(\alpha)$ és $(b; s) \not\subseteq L_2(\alpha)$. Ekkor (2.10) következményeként, mivel

$\alpha \subseteq L_2(\alpha)$, $L_1(L_2(\alpha))$ tartalmazza $L_1(\alpha)$-t, tehát $L_1(L_2(\alpha))$ tartalmazza $(b;s)$-et. Ugyanakkor $L_2(\alpha)$ létezik, így (2.11) szerint $L_2(L_2(\alpha)) = L_2(\alpha)$ is létezik, de ez már nem tartalmazza $(b;s)$-et. Tehát az $L_2(\alpha)$ parciális függvény lezártja L_1-ben nagyobb $L_2(\alpha)$-nál, azaz $\mathcal{G}_1$ nem tartalmazza $L_2(\alpha)$-t, amit viszont $\mathcal{G}_2$ nyilván tartalmaz.

Már csak azt kell belátni, hogy a (2.16) szerinti operáció a (2.17) szerintinek az inverze. Legyen az L_1 függvény-lezáráshoz rendelt $\mathcal{G}_1$ parciális metszet-félhálóból kapott lezárás L_2 és α egy tetszőleges parciális függvény az adott alaphalmazon. Ha $L_1(\alpha)$ adott, akkor $L_1(\alpha) \in \mathcal{G}$, és így (2.16) szerint $L_2(\alpha)$ is definiálva van, sőt az előző tételben bizonyítottak szerint egyenlő is $L_1(\alpha)$-val. Ha viszont $L_2(\alpha)$ adott, akkor $\mathcal{G}$ tartalmaz α-t tartalmazó parciális függvényeket, azaz lezárása definiált L_1-ben is. Viszont láttuk már, hogy $L_1(\alpha)$ létezéséből következik, hogy $L_1(\alpha) = L_2(\alpha)$. □

Az a tény, hogy a zárt halmazok rendszere egyértelműen leírja a lezárási operációt és így a függőséget, nagy mértékben lecsökkentheti az információ tárolására igénybe vett helyet és egyszerűsítheti az információtárolás struktúráját a funkcionális függőségek esetében.

Ellentétben ezzel, a parciális függőségek esetén a zárt halmazok rendszere egy adott adatbázisra nemhogy egyszerűbb (és így egy kicsit kevesebb, de sok esetben még mindig elegendő információt hordozó) struktúrát alkotnak, hanem bonyolultabbat. Gondoljunk csak arra, hogy az adatbázis minden sora szükség szerint egy zárt parciális függvényt ad (lényegében (2.14)), de ugyanakkor még rengeteg más zárt parciális függvényünk lehet. Tehát egy sokkal bővebb struktúránk van, amely természetesen nem hordozhat több információt, mint az eredeti adatbázis. Ezen redundancia feloldásához még visszatérünk a következő fejezetben is. Most itt azt próbáljuk megvizsgálni, hogy hogyan tudjuk a zárt függvények egy olyan részhalmazát kivenni, amely még mindig hordozza az összes információt, vagy legalábbis annak egy nagy részét.

A (2.15) tulajdondonságból következően a $\mathcal{G}$ parciális metszet-félhálót az elemei számánál sokkal kevesebb tagjával is meghatározhatjuk. Jelölje $\mathcal{M}(\mathcal{G})$ a $\mathcal{G}$ azon γ elemeinek a halmazát, amelyek nem állnak elő mint a $\mathcal{G}$ két másik tagjának a metszete, azaz amelyekre nem léteznek olyan α, β elemei a $\mathcal{G}$-nek amelyek mind különbözőek γ-tól és amelyekre $\gamma = \alpha \cap \beta$. Nyilvánvaló, hogy egy adott adatbázishoz tartozó parciális metszet-félháló esetén az adatbázis sorai, mint parciális függvények elemei lesznek $\mathcal{M}(\mathcal{G})$-nek, de valószínűleg egyéb parciális függvények is.

2.7. Lemma. *Egy $\mathcal{G}$ parciális metszet-félhaló minden eleme előáll mint néhány (≥ 1) $\mathcal{M}(\mathcal{G})$-beli parciális függvény metszete, de $\mathcal{M}(\mathcal{G})$-nek nincs egyetlen valódi részhalmaza sem, amelyik ezzel a tulajdonsággal bírna.*

Bizonyítás. Bizonyítsuk először az első állítást. Ha van olyan eleme $\mathcal{G}$-nek, ami nem áll elő a megkívánt módon, akkor legyen α egy maximális ilyen. Az $\mathcal{M}(\mathcal{G})$ elemei előállnak, mint egyetlen $\mathcal{M}(\mathcal{G})$-beli elem metszete, így $\alpha \notin \mathcal{M}(\mathcal{G})$. Akkor viszont $\alpha = \beta \cap \gamma$, ahol β és γ α-tól különböző elemek és így annál nagyobbak is, tehát előállnak $\mathcal{M}(\mathcal{G})$-beli elemek metszeteként. β és γ metszetként való előállítását behelyettesítve a $\beta \cap \gamma$ képletbe kapjuk α egy előállítását, mint néhány $\mathcal{M}(\mathcal{G})$-beli halmaz metszete.

Ha viszont egy α parciális függvényt kitörlünk az $\mathcal{M}(\mathcal{G})$-ből, akkor éppen ez az α parciális függvény nem lesz előállítható, mint néhány $\mathcal{M}(\mathcal{G})$-beli elem metszete. Ha ugyanis előállítható lenne, akkor vegyünk egy olyan előállítást, amely minimális számú tagot tartalmaz. A tagok száma nyilván legalább 2, és így $\alpha = \beta \cap (\cap_i \beta_i)$, ahol β és az összes β_i elemek $\mathcal{M}(\mathcal{G})$-ből valók. Tehát $\beta \neq \alpha$ és a minimalitás miatt $\cap_i \beta_i \neq \alpha$, azaz α előállt mint két tőle különböző $\mathcal{G}$-beli elem metszete. Ez viszont a definíció szerint ellentmond annak a ténynek, hogy $\alpha \in \mathcal{M}(\mathcal{G})$. □

A következő két tétel leírja az $\mathcal{M}(\mathcal{G})$ családokat és azok viszonyát a parciális metszet-félhálókhoz.

2.8. TÉTEL. *Parciális függvények egy $\mathcal{Z}$ családja pontosan akkor egyezik meg egy valamely $\mathcal{G}$ parciális metszet-félhálóhoz tartozó $\mathcal{M}(\mathcal{G})$ családdal, ha kielégíti az alábbi feltételeket:*

(2.18) minden $\alpha \in \mathcal{Z}$-re létezik egy olyan $\gamma \in \mathcal{Z}$, amelyre $\alpha \subseteq \gamma$, $\gamma = \{c_1, \ldots, c_n; \ldots\}$ és $\{c_1, \ldots, c_n\} = \Omega$;

(2.19) ha $\alpha = \bigcap_{i=1}^r \alpha_i$, $(r \geq 1)$, $\alpha, \alpha_1, \ldots \alpha_r \in \mathcal{Z}$, akkor $\alpha = \alpha_i$ valamely i-re.

Bizonyítás. Legyen először $\mathcal{G}$ egy tetszőleges metszet-félháló. Akkor $\mathcal{M}(\mathcal{G})$ (2.18) tulajdonsága következik a (2.14)-ből. (2.19)-hez legyen $\alpha = \bigcap_{i=1}^r \alpha_i$, $(r \geq 1)$, $\alpha, \alpha_1, \ldots, \alpha_r \in \mathcal{M}(\mathcal{G})$. Ekkor $\alpha = (\alpha_1 \cap \cdots \cap \alpha_{r-1}) \cap \alpha_r$, és itt $\alpha, \alpha_1, \ldots, \alpha_r$, $\alpha_1 \cap \cdots \cap \alpha_{r-1} \in \mathcal{G}$, valamint a minimalitás miatt $\alpha_1 \cap \cdots \cap \alpha_{r-1} \neq \alpha$. Tehát, mivel $\alpha \in \mathcal{M}(\mathcal{G})$, $\alpha_r = \alpha$.

Tegyük most fel, hogy parciális függvények egy családja kielégíti a (2.18) és (2.19) definíciókat. Konstruáljunk egy olyan $\mathcal{G}$ parciális metszet-félhálót, amelyre $\mathcal{M}(\mathcal{G}) = \mathcal{Z}$. Legyenek $\mathcal{G}$ elemei az összes $\mathcal{Z}$-ből formálható metszetek. Ahhoz, hogy az így kapott struktúra egy parciális metszet-félháló legyen, azt kell belátnunk 2.7 lemma szerint, hogy $\mathcal{Z}$ egyetlen eleme sem áll elő, mint tőle különböző két $\mathcal{G}$-beli elem metszete, de a $\mathcal{G} - \mathcal{Z}$-beli elemeket elő tudjuk állítani így.

Tegyük fel először, hogy $\alpha \in \mathcal{Z}$, $\alpha = \beta \cap \gamma$, $\beta, \gamma \in \mathcal{G}$. A definíció szerint β és γ előállnak, mint $\mathcal{Z}$-beli elemek metszetei, tehát α-ra is kapunk egy ilyen előállítást. Viszont minden ilyen előállításban (2.19) szerint az egyik tag α, tehát β és γ egyike is α kell hogy legyen.

Tegyük most fel, hogy $\alpha \in \mathcal{G} - \mathcal{Z}$. Akkor α-t felírhatjuk, mint $\alpha = \alpha_1 \cap \cdots \cap \alpha_s$; $s \geq 2$, $\alpha_1, \ldots, \alpha_s \neq \alpha$. Tegyük fel hogy a felírásban s-t minimálisnak választottuk. Akkor $\alpha_1 \cap \cdots \cap \alpha_{s-1} \neq \alpha$, és így $\alpha = (\alpha_1 \cap \cdots \cap \alpha_{s-1}) \cap \alpha_s$ az α két tőle különböző $\mathcal{G}$-beli elemként való előállítása. □

A (2.18) és (2.19) tulajdonságokat kielégítő parciális függvény családot *metszet-mentes* családoknak nevezzük. Az alábbiakban bebizonyítjuk, hogy a metszet-mentes családok egyértelműen meghatározzák a parciális metszet-félhálókat.

2.9. TÉTEL. *Az ugyanazon alaphalmazon értelmezett $\mathcal{G}$ parciális metszet-félhálók és $\mathcal{Z}$ metszet-mentes parciális függvény családok között egy egyértelmű megfeleltetés adható meg az alábbiak szerint:*

$$\mathcal{G} \to_1 \mathcal{M}(\mathcal{G}). \tag{2.20}$$

A megfeleltetés inverzét az alábbiak szerint kapjuk.

$$\mathcal{Z} \to_2 \mathcal{G} = \{\alpha_1 \cap \cdots \cap \alpha_r, \quad r \geq 1, \ \alpha_1, \ldots, \alpha_r \in \mathcal{Z}\}. \tag{2.21}$$

Bizonyítás. A 2.6 tétel bizonyításához hasonlóan először azt látjuk be, hogy a $\to_1$ hozzárendelés injektív (azt már láttuk az előző lemmában, hogy az értéke egy metszet-mentes család lesz). Legyenek $\mathcal{G}_1$ és $\mathcal{G}_2$ két különböző parciális metszet-félháló és tegyük fel, hogy $\alpha \in \mathcal{G}_1$ de $\alpha \notin \mathcal{G}_2$. Akkor a 2.7 lemma szerint α előáll, mint néhány $\mathcal{M}(\mathcal{G}_1)$-beli parciális függvény metszete, de nem állhat elő, mint néhány $\mathcal{M}(\mathcal{G}_2)$-beli parciális függvény metszete. Így tehát $\mathcal{M}(\mathcal{G}_1) \neq \mathcal{M}(\mathcal{G}_2)$.

Azt is láttuk már a 2.8 tételben, hogy a $\to_1$ operáció értékkészlete az összes metszet-mentes család. Már csak azt kell belátni, hogy (2.20) inverze (2.21). Rendelje hozzá $\mathcal{M}(\mathcal{G}_1)$-hez $\to_2$ $\mathcal{G}_2$-t. Ha egy α parciális függvény eleme $\mathcal{G}_1$-nek, akkor az $\mathcal{M}(\mathcal{G})$ definíciója és a 2.7 lemma szerint előáll, mint néhány $\mathcal{M}(\mathcal{G}_1)$-beli parciális függvény metszete, így $\alpha \in \mathcal{G}_2$ is fennáll. Ha viszont $\alpha \in \mathcal{G}_2$, akkor lényegében a 2.8. tétel bizonyításának a második felében leirtak szerint $\alpha \in \mathcal{G}_1$ is, azaz $\mathcal{G}_1 = \mathcal{G}_2$.

□

3. A parciális függőségek realizálása

Az eddigiek szerint a parciális függvények metszet-mentes családjai egyértelműen leírják a parciális függvény-lezárásokat ill. a parciális függőségeket. Mást nem is várhatunk, hiszen (2.18) és a 2.9. tétel azt is jelentik, többek között, hogy egy adatbázis minden sora szerepel mint parciális függvény az adatbázis által meghatározott metszet-mentes családban. Ugyanez az állítás viszont már nem igaz az adatbázisok és metszet-mentes családok viszonyára. Noha a 2. fejezetben bizonyítottuk, hogy egy adatbázis egyértelműen meghatározza a metszet-mentes családot, ennek fordítottjából még csak annyit láttunk, hogy a metszet-mentes családhoz egyértelműen tartozik egy parciális függőségi család, vagy még pontosabban egy függvény-lezárás. Azt nem, hogy találunk egy megfelelő adatbázist is. Persze ha találunk egyet, akkor az egyértelmű, hiszen sorai a fentiek szerint adottak. (A továbbiakban is feltesszük, hogy a parciális függvényeink értelmezési tartománya mindig része ugyanannak az Ω alaphalmaznak.)

Azonban nem mindig létezik ilyen adatbázis. Legyenek a parciális függvények metszet-mentes halmazának elemei $(a, b, c; p, q, r)$ és $(a, b; p, q)$. Az ehhez tartozó adatbázis csak az (a, b, c) attributum-halmazú és az egyetlen (p, q, r) sorból álló adatbázis lehetne. Ehhez viszont egy olyan függvény lezárás tartozik, aminek a parciális függvényekből álló metszet-mentes halmaza csak az egyetlen $(a, b, c; p, q, r)$ függvényből áll.

A fentebbi példában az okozta a bajt, hogy volt olyan nem Ω–n értelmezett elemünk, amely nem állt elő metszetként. Belátjuk, hogy ez soha nem fordulhat elő.

3.1. TÉTEL. *Legyen $\mathcal{G}$ egy adatbázis zárt parciális függvényei által alkotott parciális metszet-félháló. Akkor $\mathcal{G}$ minden olyan eleme, amely nem az Ω–n van értelmezve, előáll, mint két, tőle különböző $\mathcal{G}$–beli elem metszete.*

Bizonyítás. Legyen $\{a_1, \ldots, a_k\}$ egy valódi részhalmaza Ω–nak és $\alpha = (a_1, \ldots, a_k; r_1, \ldots, r_k)$ egy zárt parciális függvény, azaz $\mathcal{G}$ egy eleme. Először belátjuk, hogy α két tőle különböző $\mathcal{G}$–beli parciális függvénynek is a (valódi) része. Legyen $c \in \Omega$, de $c \notin \{a_1, \ldots, a_k\}$. Az, hogy az $\alpha = (a_1, \ldots, a_k; r_1, \ldots, r_k)$ parciális függvény zárt azt jelenti, hogy létezik két olyan s, t adat és az adatbázisnak két olyan sora, amelyek a $(a_1, \ldots, a_k, c; r_1, \ldots, r_k, s)$ ill. a $(a_1, \ldots, a_k, c; r_1, \ldots, r_k, t)$ parciális függvényeket tartalmazzák. Legyen $(a_1, \ldots, a_k, c; r_1, \ldots, r_k, s)$ lezártja β és $(a_1, \ldots, a_k, c; r_1, \ldots, r_k, t)$ lezártja γ. Ekkor β és γ valóban tartalmazza α–t, tehát $\beta \cap \gamma \supseteq \alpha$.

Legyen most β és γ két olyan, az α–tól különböző zárt (azaz $\mathcal{G}$–beli) parciális függvény, amelyre $\beta \cap \gamma \supseteq \alpha$ és az $\text{ÉT}(\beta \cap \gamma) - \text{ÉT}(\alpha)$ halmaz nagysága minimális. Ekkor persze a (2.15) tulajdonság szerint $\beta \cap \gamma$ is $\mathcal{G}$–beli. Tegyük fel, hogy a $\text{ÉT}(\beta \cap \gamma) - \text{ÉT}(\alpha)$ halmaznak van legalább egy, mondjuk c eleme, és a $\beta \cap \gamma$ parciális függvény s értéket vesz fel a c helyen. Az $\alpha = (a_1, \ldots, a_k; r_1, \ldots, r_k)$ parciális függvény zárt, így az adatbázisnak létezik egy olyan sora, ami tartalmazza α–t, de a c helyen egy más, mondjuk t értéket vesz fel. Legyen δ az $(a_1, \ldots, a_k, c; r_1, \ldots, r_k, t)$ parciális függvény lezártja. Ekkor $\delta \supseteq \alpha$, tehát $(\beta \cap \gamma) \cap \delta \supseteq \alpha$. Ugyanakkor a $(\beta \cap \gamma)$ és a δ parciális függvények különböző értékeket vesznek fel a c helyen, így c már nem eleme $(\beta \cap \gamma) \cap \delta$ értékkészletének, azaz $\text{ÉT}(\beta \cap \gamma) - \text{ÉT}(\alpha)$ szigorúan tartalmazza $\text{ÉT}\{(\beta \cap \gamma) \cap \delta\} - \text{ÉT}(\alpha)$–t ami ellentmondás. Így az $\text{ÉT}(\beta \cap \gamma) - \text{ÉT}(\alpha)$ halmaz üres kell hogy legyen, azaz $\beta \cap \gamma = \alpha$. Tehát α–t valóban előállítottuk, mint két, tőle különböző $\mathcal{G}$–beli parciális függvény metszetét. □

Tehát egy adott adatbázis esetén az $\mathcal{M}(\mathcal{G})$ halmaz minden eleme az Ω–n kell, hogy értelmezett legyen, azaz $\mathcal{M}(\mathcal{G})$ kielégíti az alábbi feltételt:

(3.1) minden $\alpha \in \mathcal{M}(\mathcal{G})$–re $\alpha = \{c_1, \ldots, c_n; \ldots\}$ és $\{c_1, \ldots, c_n\} = \Omega$.

Parciális függvények ilyen tulajdonságú halmazát a továbbiakban parciális függvények *nagy* halmazának fogjuk nevezni. A nagy halmazok már kielégítik azt az elvárást, hogy realizálhatók lesznek (természetesen egyértelműen) adatbázisok parciális függőségi relációival.

3.2. LEMMA. *Legyen M a parciális függvények egy adott $\mathcal{Z}$ nagy halmazához tartozó azon adatbázis (mátrix) amelynek attributumai (oszlopai) az Ω elemei és amelynek sorai a $\mathcal{Z}$–beli (és így tehát a teljes Ω halmazon értelmezett) parciális függvények. Akkor az M adatbázishoz a 2. fejezetben leírtak szerint hozzárendelt* *metszet-mentes család éppen $\mathcal{Z}$; vagy ami ezzel ekvivalens, az M által meghatározott parciális függőségek rendszere (függőségi család) megegyezik a $\mathcal{Z}$-hez egyértelműen tartozó függőségi családdal.*

Bizonyítás. A bizonyításhoz meg kell gondolni, hogy mit jelent az $\alpha \rightarrow \beta$ reláció M-ben, ill. a parciális függvények $\mathcal{Z}$ nagy halmaza szerint. A parciális függőség definíciójából jól látható, hogy M-ben $\alpha \rightarrow \beta$ pontosan akkor áll fenn, ha M minden olyan sora, amely tartalmazza α-t, tartalmazza β-t is.

Tegyük most fel, hogy valamely lezárási operáció (és így a hozzá tartozó parciális metszet-félháló és a parciális halmazok nagy halmaza — mivel már csak olyan lezárási operációkat tekintünk, amelyekhez tartozó metszet-mentes halmazok „nagyok" is egyben) szerint $\alpha \rightarrow \beta$, azaz $\beta \subseteq L(\alpha)$. Ez a félhálóban éppen azt jelenti, hogy minden olyan félhálóbeli (azaz zárt) parciális függvény, amely tartalmazza α-t, tartalmazza β-t is. A parciális függvények $\mathcal{Z}$ nagy halmazára pedig ez éppen azt jelenti, hogy minden $\mathcal{Z}$-beli parciális függvény, ami tartalmazza α-t, tartalmazza β-t is. Viszont a $\mathcal{Z}$ parciális függvényei éppen az adott adatbázis sorai.

A fentebb leírtak tehát azt jelentik, hogy az így definiált M mátrixból kapott parciális függőség megegyezik a $\mathcal{Z}$ nagy halmazhoz tartozó parciális függőséggel. □

Az eddigiekben a parciális függőségek által meghatározott olyan következményekkel ill. struktúrákkal foglalkoztunk, amelyek maguk a parciális függőséggel ill. az azt megadó adatbázissal ekvivalensek voltak. A továbbiakban azt vizsgáljuk meg néhány esetben, hogy mi történik akkor, ha csak gyengébb struktúrákat veszünk figyelembe.

A 2. fejezet végén vagy az ezen fejezet elején említett redundanciát elkerülhetnénk, ha $\mathcal{M}(\mathcal{G})$-ből kitörölnénk a maximális elemeket (azaz azokat, amelyek az Ω alaphalmazon vannak értelmezve). Ekkor azonban egy üres parciális függvény halmazt kapnánk.

Ezen probléma kiküszöbölésére egy $\mathcal{G}$ parciális metszet-félhálóhoz megadhatunk egy másik struktúrát is oly módon, hogy először töröljük el $\mathcal{G}$-nek az Ω-n értelmezett függvényeit (nevezzük az így kapott parciális függvény halmazt $\mathcal{G}'$-nek), és utána a maradékból vesszük azokat, amelyek nem állnak elő mint két másik, tőle különböző $\mathcal{G}'$-beli függvény metszete. Nevezzük az így kapott parciális függvény családot $\mathcal{M}_1(\mathcal{G})$-nek. Könnyen látható, hogy $\mathcal{M}_1(\mathcal{G})$-ből egyértelműen megkapható az összes, nem Ω-n értelmezett $\mathcal{G}$-beli parciális függvény. Sőt, nem nehéz meggondolni, hogy a $\mathcal{G}'$ és az $\mathcal{M}_1(\mathcal{G})$ viszonyára szinte szóról szóra megismételhetők a 2.8. és 2.9. tételek, azzal a különbséggel, hogy 2.8.-ból törölni kell a (2.18) feltételt.

Tehát egy adott adatbázishoz újabb struktúrát rendeltünk, amely esetenként jól használható az adatbázis leírásához. Láttuk, hogy adott $\mathcal{G}$ metszet-félháló esetén az $\mathcal{M}_1(\mathcal{G})$ parciális függvény halmaz kielégíti a (2.19) feltételt. Azonban az $\mathcal{M}_1(\mathcal{G})$ parciális függvény halmazra már nem lesz igaz az a tulajdonság, hogy $\mathcal{M}_1(\mathcal{G})$ minden (a tartalmazásra nézve) nem maximális halmaza előáll, mint két tőle különböző

$\mathcal{M}_1(\mathcal{G})$-beli parciális függvény metszete. Ezt a következő adatbázis mutatja:

attributumok:	a	b	c
első sor	p	r	s
második sor	p	q	t
harmadik sor	p	q	u

Itt a zárt parciális függvények a sorokon kivül $(a, b; p, q)$ és $(a; p)$, ahol az utóbbi nem maximális, de mégsem metszete két másiknak.

Persze az $\mathcal{M}_1(\mathcal{G})$ halmaz kielégít még egy triviális feltételt, nevezetesen a (3.1) feltétel ellentetjét, azaz egyetlen $\mathcal{M}_1(\mathcal{G})$-beli parciális függvény sem lehet az Ω halmazon értelmezve:

(3.2) minden $(a_1, \ldots, a_k; \ldots) \in \mathcal{M}_1(\mathcal{G})$-re $\{a_1, \ldots, a_k\}$ valódi része Ω-nak.

Kérdés, hogy a (2.19) és (3.2) feltételek jól jellemzik-e ezeket a halmazokat?

3.3. TÉTEL. *Minden olyan $\mathcal{Z}$ parciális függvény halmazhoz, amely kielégíti a (2.19) és (3.2) feltételeket, létezik egy adatbázis, amelyhez tartozó $\mathcal{G}$ parciális metszet-félhálóra $\mathcal{M}_1(\mathcal{G}) = \mathcal{Z}$.*

Bizonyítás. Azt bizonyítjuk be, hogy létezik parciális függvények egy olyan $\mathcal{G}$ metszet-félhálója, amelyre $\mathcal{M}(\mathcal{G})$ nemcsak metszet-mentes, hanem „nagy" is, és $\mathcal{M}_1(\mathcal{G}) = \mathcal{Z}$. Ehhez akkor 3.2. lemma szerint létezik egy őt realizáló adatbázis, amely az itteni céloknak is meg fog felelni.

A $\mathcal{Z}$ minden egyes α parciális függvényéhez felveszünk a leendő adatbázisban két, addig még egyáltalán nem használt elemet, mondjuk a-t és b-t. Ezek után α-hoz definiálunk két, a teljes Ω halmazon értelmezett parciális függvényt: mindkettő tartalmazza α-t, és az α-n kivüli helyeken az egyik a, a másik pedig b értékeket vesz fel. Tekintsük azt az $\mathcal{N}$ „nagy" halmazt, amelyet a $\mathcal{Z}$ összes eleméhez ily módon felvett parciális függvények (és csak azok) alkotnak. Az α-hoz felvett párok metszete így α lesz, viszont könnyen láthatóan akármely más, legalább kéttagú $\mathcal{N}$-beli metszet előáll, mint néhány $\mathcal{Z}$-beli parciális függvény metszete. Így az $\mathcal{N}$-ből előállított $\mathcal{G}$ metszet-félhálóhoz tartozó $\mathcal{M}_1(\mathcal{G})$ halmaz valóban $\mathcal{Z}$ lesz. □

A dolgozat hátralévő részében még egyetlen problémát fogunk megvizsgálni. Számos dolgozatban (pl. [3,4]) megtalálható a 2.1. lemma változata funkcionális függőségekre:

3.4 LEMMA. *Ha egy Ω attributum-halmazon értelmezett adatbázisban $A, B, C, D \subseteq \Omega$ és $\to$ jelöli a funkcionális függőséget, akkor*

(3.3) $$A \to A;$$

(3.4) $A \to B$ és $B \to C$-ből következik, hogy $A \to C$;

(3.5) $A \subseteq C$, $D \subseteq B$ és $A \to B$-ből következik, hogy $C \to D$;

(3.6) $A \to B$ és $C \to D$-ből következik, hogy $A \cup C \to B \cup D$. □

Ellentétben a parciális függőségek esetével, ennek a lemmának a fordítottja is igaz. Az Ω részhalmazaiból képzett párok egy halmazát *teljes családnak* nevezzük, ha az kielégíti a (3.3)–(3.6) feltételeket.

3.5. TÉTEL. [3] *Legyen $\mathcal{T}$ egy az Ω-n értelmezett teljes család. Akkor létezik egy adatbázis, amelynek az Ω az attributum halmaza és amelyben a (funkcionális) függősége rendszere megegyezik $\mathcal{T}$-val.* □

Mi most itt azt vizsgáljuk meg, hogy a parciális függőségek rendszere milyen struktúrát generál az attributumok halmazán. A továbbiakban egy az Ω attributum halmazon adott adatbázis esetén ha $A, B \subseteq \Omega$, akkor $A \to B$ alatt azt értjük, hogy létezik az adatbázishoz tartozó két olyan α és β parciális függvény, hogy α értelmezési tartománya A, β értelmezési tartománya B és $\alpha \to \beta$. $A \to B$-t úgy mondjuk, hogy B parciálisan függ A-tól.

3.6. LEMMA. *Ha egy adott adatbázis esetén A, B, C, D az Ω négy részhalmaza, akkor (3.3) és (3.5) fennáll rájuk.*

Bizonyítás. (3.3) nyilvánvalóan következik (2.3)-ból. (3.5) pedig (2.5)-nek és annak a ténynek a következménye, hogy ha α, β két parciális függvény és $\alpha \subseteq \beta$, akkor $\text{ÉT}(\alpha) \subseteq \text{ÉT}(\beta)$. □

Az Ω részhalmazaiból képzett párok egy halmazát *parciálisan teljes családnak* nevezzük, ha az kielégíti a (3.3) és (3.5) feltételeket. A dolgozat utolsó eredményeként azt látjuk be, hogy a parciálisan teljes családok reprezentálhatók adatbázisokkal.

3.7. TÉTEL. *Minden, az Ω-n értelmezett parciálisan teljes családhoz létezik egy olyan adatbázis, amely a részhalmazain adott parciális függőségként az adott parciálisan teljes családot adja.*

Bizonyítás. Tekintsük Ω minden A részhalmazára azokat a tartalmazásra nézve maximális B halmazokat, amelyekre $A \to B$. Minden ilyen A, B párra vezessünk be három új értéket (mondjuk a, b, c–t) a konstruálandó adatbázisban, és konstruáljunk néhány új parciális függvényt. Legyen α értelmezési tartománya A és vegyen fel minden A-beli helyen a–t. Ha A és B különböznek, akkor legyen β értelmezési tartománya B és ugyancsak vegyen fel minden helyen a–t. Legyen végül γ és δ értelmezési tartománya Ω, vegyen fel mindkettő a B–beli helyeken a–t, a B–n kivüli helyeken pedig γ b–t és δ c–t. Ekkor $\alpha \to \beta(\alpha)$, $\gamma \to \beta$ és $\delta \to \beta$. Természetesen ezekkel a parciális függvényekkel együtt fel kell venni az általuk tartalmazottakat is. Ha ezt a műveletet elvégezzük Ω minden részhalmazára, és egy adott részhalmaz esetén az összes őt tartalmazó olyan maximális részhalmazra, amely parciálisan függ tőle, kapunk egy parciális függvényekből álló halmazt. Nem nehéz ellenőrizni, hogy ez a halmaz kielégíti a (2.1) és (2.2) feltételeket, azaz leszálló lesz. Definiáljunk ezen parciális függvények között parciális függőségeket. Ha $\alpha, \beta, \gamma, \delta$ egy a fentebb értelmezett $A \rightarrow B$ párhoz értelmezett parciális függvények, akkor legyen $\alpha \rightarrow \beta$ (ill. α, ha $A = B$), $\gamma \rightarrow \beta(\alpha)$ és $\delta \rightarrow \beta(\alpha)$. Definiáljuk még az összes olyan parciális függőséget is, amelyet definiálni kell a (2.5) szerint.

Nem nehéz látni, hogy a parciális függvényeken a fentebb értelmezett parciális függőségek által az Ω részhalmazain definiált parciális függőségek éppen az adott parciálisan teljes családnak megfelelőek lesznek. Ugyancsak rutin, habár hosszadalmas feladat annak ellenőrzése, hogy a parciális függőségek itt adott rendszere kielégíti a (2.3)–(2.6) feltételeket. Itt megjegyezzük, hogy a (2.4) és (2.6) feltételek lényegében azért lesznek igazak, mert a feltételeik az adott esetben üres állítások.

Végül az eddigiek fényében azt kell csak belátnunk, hogy az így adott parciális függőség szerint zárt parciális függvények közül azok, amelyek nem az Ω–n vannak értelmezve, előállnak, mint két másik zárt metszete. Ezt viszont a konstrukció biztosítja. Az Ω–nál kisebb halmazokon értelmezett zárt parciális függvények éppen a fenti konsrtukció β–i lesznek (ill. α–i), amik viszont minden egyes esetben előállnak, mint $\gamma \cap \delta$. □

A fentiekben megadott eredményeken kivül még számos dolgot kérdezhetük a parciális függőségről. Néhány ezek közül:

— Az előzőekben több „realizálhatósági" eredményt láttunk. Tudjuk-e ezekben valahogy a realizáló adatbázist minimalizálni?

— Mit tudunk azokról a struktúrákról mondani, amelyeket úgy kapunk, hogy a parciális függőség függvény-lezárási operációjának, vagy a zárt halmazok metszet-félhálójának ill. egy $\mathcal{M}(\mathcal{G})$ halmaznak a „vázát" vesszük, azaz tekintjük a bennük található parciális függvények értékkészleteinek a halmazát?

— Hogyan változnak a parciális függőséghez tartozó struktúrák, ha az adatbázist változtatjuk, azaz pl. eltöröljük egy sorát?

IRODALOM

[1] ARMSTRONG, W. W., "Dependency structures of data base relationships ", *Information Processing 74'* (North Holland, Amsterdam, 1974), 580–583.

[2] CODD, E. F., "A relational model of data for large shared data banks ", *Com. ACM* **13** (1970), 377–387.

[3] DEMETROVICS, J and KATONA, G. O. H., "Combinatorial problems of database models ", Coll. Math. Soc. János Bolyai, 42. Algebra, Combinatorics and Logic in Computer Science, Győr (Hungary), 1983, pp. 331–353.

[4] DEMETROVICS, J. and KATONA, G. O. H., "Extremal Combinatorial problems of database models ", Kézirat.

(Beérkezett: 1988. november 28.)
(Végleges változat: 1991. szeptember 15.)

DEMETROVICS JÁNOS
MTA SZÁMÍTÁSTECHNIKAI ÉS AUTOMATIZÁLÁSI KUTATÓ INTÉZET
1132 BUDAPEST, VICTOR HUGO U. 18

KATONA GYULA ÉS MIKLÓS DEZSŐ
MTA MATEMATIKAI KUTATÓ INTÉZET
1053 BUDAPEST, REÁLTANODA U. 13-15

PARTIAL DEPENDENCIES IN RELATIONAL DATABASES

J. DEMETROVICS, G. O. H. KATONA, D. MIKLÓS

Weakening the functional dependencies defined by AMSTRONG we get the notion of partial dependencies defined the relational databases. We show that partial dependencies can be characterized by closure operations of the poset formed by the partial functions on the attributes of the databases. On the other hand, we give necessary and sufficient conditions so that for such a closure operation one can find on the given set of attributes a database whose partial dependencies generate the given closure operation. We characterize the dependencies defined by the set of partial dependencies on the poset of the subsets of the attributes and describe how we can realize such a dependency.

Alkalmazott Matematikai Lapok 15 (1990–91) 181–196

ELÁGAZÓ FÜGGŐSÉG RELÁCIÓS ADATBÁZISOKBAN*

DEMETROVICS JÁNOS, KATONA GYULA ÉS SALI ATTILA

Budapest

Egy relációs adatmodell úgy tekinthető, mint egy mátrix, melynek sorai az individuumok, az oszlopai az attributumok. Ezen elméletnek egy jól ismert fogalma a funkcionális függőségek fogalma. Ha A oszlopoknak egy halmaza, b pedig egy oszlop, b–ről azt mondják, hogy függ A–tól, ha nincs két olyan sor, amely b–ben különböző, de A–ban azonos. Ezt a fogalmat általánosítjuk itt. Legyenek $p \leq q$ pozitív egészek. Azt mondjuk, hogy b (p,q)–függ A–tól, ha nincs $q+1$ olyan sor, amelyek b–ben teljesen különböznek, de A minden oszlopában legfeljebb p különféle érték áll. A hagyományos függőségekre ismert több tétel van a cikkben általánosítva.

1. Bevezetés

Egy adatbázis felfogható, mint egy mátrix, amelynek oszlopai az *adatfajták, attributumok* (pl. név, születési hely, idő, stb.), míg egy-egy sor tartalmazza egy-egy *individuum* adatait. Jelölje X az attributumok (azaz a mátrix oszlopainak) halmazát. Legyen $A \subseteq X$, $b \in X$. Azt mondjuk (ARMSTRONG [1], CODD [2]), hogy b *funkcionálisan függ* A-tól (jelölésben: $A \longrightarrow b$), ha az A–ban álló adatok már meghatározzák a b–ben álló adatot is, azaz nincs két olyan sor, amely A–ban egyezik és b–ben különbözik.

A funkcionális függőségek nagyon hasznosnak bizonyultak a gyakorlatban. A létező adatbáziskezelő rendszerek mind ezen a fogalmon alapszanak. Tekintsünk egy példát. Tegyük fel, hogy $X = \{x_1, x_2, x_3, x_4\}$, $x_1 \longrightarrow x_2$ és $x_3 \longrightarrow x_4$. Ha az egész mátrixot tároljuk a memóriában, az a legrosszabb esetben $4N_1N_3$ helyet jelent, ahol N_1 ill. N_3 jelöli az x_1 ill. x_3 által felvehető különböző értékek számát. Hiszen x_1 és x_3 egymástól függetlenül vehetik fel értékeiket (viszont x_2–t és x_4–et már meghatározzák), így a sorok száma maximálisan N_1N_3. Azonban — kihasználva az adott két funkcionális függőséget — rengeteg helyet megspórolhatunk. Ugyanis elegendő az x_1 és x_3 oszlopából álló mátrixon kívül (ami $2N_1N_3$ értéket tartalmaz), két kis mátrixot tárolnunk. Az egyik mátrix két oszlopa x_1 és x_2 értékeit tartalmazza. Az első oszlopban állnak x_1 lehetséges értékei, a másodikban az $x_1 \longrightarrow x_2$ funkcionális függőség által meghatározott értékek. A másik mátrixot is hasonlóan építjük fel x_3–ból és x_4–ből. A tárolt értékek száma legfeljebb $2N_1N_3 + 2(N_1 + N_3)$, ami általában sokkal kisebb $4N_1N_3$–nál.

Jelen cikkben a funkcionális függőségnél egy általánosabb (gyengébb) fogalmat fogunk bevezetni. Először csak egy nagyon speciális esetben, majd illusztráljuk a

*Az 2575–ös számú Országos Tudományos Kutatási Alap által támogatva.

fogalom hasznosságát. Ha $A \subseteq X$, $b \in X$, akkor azt mondjuk, hogy b (1,2)-függ A-tól, ha az A-ban álló értékek „kétértelműen" meghatározzák a b-ben álló értékeket, azaz nincs 3 olyan sor amelyek mind megegyeznek A-ban, de 3 különböző értéket tartalmaznak b-ben. Jelölése $A\text{—}(1,2) \rightarrow b$. Hasonlóan, $A\text{—}(1,q) \rightarrow b$, ha nincs $q+1$ olyan sor, amelyek A-ban megegyeznek, de b-ben $q+1$ különböző értéket tartalmaznak.

Tegyük fel, hogy az adatbázis egy kamion útjait tartalmazza, pontosabban csak az általa érintett országokat. Az egyszerűség kedvéért tegyük fel, hogy egy útja során egy kamion pontosan 4 országot érint (beleértve az indulási és érkezési országot is) és nem megy vissza olyan országba, ahol már volt. Tegyük fel továbbá, hogy 30 ország jön szóba, egy országnak legfeljebb 5 szomszédja van. Jelölje x_1, x_2, x_3, x_4 az első, második, harmadik ill. negyedik országot mint attributumot. Könnyen látható, hogy $x_1\text{—}(1,5) \rightarrow x_2$, $\{x_1, x_2\}\text{—}(1,4) \rightarrow x_3$ és $\{x_2, x_3\}\text{—}(1,4) \rightarrow x_4$. Most nem tudjuk a tárolt mátrix nagyságát csökkenteni, mint a funkcionális függőség ((1,1)-függőség) esetén, de tudjuk a mátrix elemeinek értékkészletét csökkenteni. Az eredeti mátrix minden eleme 30 féle értéket vehet fel, az országnevet, vagy azok valamilyen kódjait (5 bit). Tároljunk egy kis (legfeljebb $30 \times 5 \times 5 = 750$–ös) táblázatot, amely minden ország szomszédainak egy számozását tartalmazza, azaz hozzájuk rendeli a 0,1,2,3,4 számokat. Ekkor az x_2 attributumot helyettesíthetjük ezen számokkal, hiszen x_1 már megadja az induló országot, az $x_2{}^*$ attributum értéke a kis táblázatból megállapíthatóan meghatározza a második országot. Ugyanez érvényes x_3–ra is, de itt még csökkenthetjük a lehetséges értékek számát, ha a lehetséges x_1, x_2 párokhoz harmadikként csatlakozó országokat megszámozó táblázatot is megadunk. Ekkor az így kapott $x_3{}^*$ már csak 4 különböző értéket vehet fel. Ugyanez érvényes x_4–re is. Tehát, míg az eredeti mátrixban minden elem 5 bittel volt leírható, most 2 kis póttáblázat árán a második oszlopban állókat 3 bitre, a harmadik és negyedik oszlopban állókat 2 bitre csökkentettük.

Könnyen látható, hogy ez a gondolat teljesen hasonlóan alkalmazható minden olyan esetben, amikor valamilyen gráf útjait kell tárolni, ahol a gráfban a maximális fok sokkal kisebb a szögpontok számánál. Vagy másképpen fogalmazva, ha valamilyen folyamat állapotsorozatait kell tárolni, ahol az állapotok száma sokkal nagyobb az állapotok rákövetkező állapotainak maximális számánál. De más esetekben is, ha sok $(1,q)$–függés van, ahol q kicsi.

Az általános fogalom, amit tárgyalni fogunk, a (p,q)–függés ($1 \leq p \leq q$, egészek). Akkor mondjuk, hogy b az A–tól (p,q)–függ, ha nincs $q+1$ olyan sor, amely A minden oszlopában legfeljebb p különböző értéket tartalmaz, b–ben pedig $q+1$ csupa különbözőt. Ez nyilván általánosítása a funkcionális függőség fogalmának. A cikk fő célja bizonyos, a funkcionális függőségre vonatkozó tételek általánosítása a (p,q)–függőségekre.

2. A (p,q)–függőségek jellemzése

Ha adott az M adatbázis (vagy mátrix), definiáljuk a $\mathcal{J}_{Mpq}(A)$, $A \subseteq X$

függvényt a következőképpen:

$$\mathcal{J}_{Mpq}(A) = \{b : A -\!\!\!(p,q) \rightarrow b\}.$$

Ha egyébként nem okoz félreértést, az M, p, q indexeket elhagyjuk.

Könnyen látható, hogy ha $b \in A$, akkor $A -\!\!\!(p,q) \rightarrow b$, tehát

$$A \subseteq \mathcal{J}_{Mpq}(A). \tag{2.1}$$

Másrészt, ha $A \subseteq B$ és $A -\!\!\!(p,q) \rightarrow b$, akkor $B -\!\!\!(p,q) \rightarrow b$. Tehát

$$\mathcal{J}_{Mpq}(A) \subseteq \mathcal{J}_{Mpq}(B), \text{ ha } A \subseteq B. \tag{2.2}$$

A (2.1) és (2.2) feltételeket kielégítő halmaz-halmaz függvényeket *növelő-monoton függvényeknek* nevezzük. Ha egy $\mathcal{N}$ növelő-monoton függvényhez található egy olyan M mátrix, hogy $\mathcal{J}_{Mpq} = \mathcal{N}$, akkor azt mondjuk, hogy (p,q)–reprezentálható.

2.1. TÉTEL. *Ha X–en adott egy $\mathcal{N}$ növelő-monoton függvény, melyre $\mathcal{N}(\emptyset) = \emptyset$ teljesül, és $1 < q$, akkor $\mathcal{N}$ $(1,q)$–reprezentálható.*

Bizonyítás. Nevezzük *láncnak* X részhalmazainak egy $\emptyset \neq A_1 \subset \dots \subset A_k$ sorozatát, amelyekre teljesülnek a következő feltételek:

$$\mathcal{N}(A_i) = A_{i+1} \quad (1 \le i \le k), \tag{2.3 }$$

$$\mathcal{N}(A_k) = A_k. \tag{2.4}$$

Egy ilyen adott L lánchoz készítsük el a következő $M(L)$ mátrixot. (Lásd a táblázatot.)

A_1	$A_2 - A_1$	$A_3 - A_2$	$\cdots$	$A_k - A_{k-1}$	$X - A_k$
1	1	1		1	1
1	1	1		1	2
1	·	·		·	3
·	·	·		·	4
·	·	·		·	·
·	1	$r+1$		$(k-2)r+1$	$kr+1$
1	2	$r+2$		$(k-2)r+2$	$kr+2$
1	3	$r+3$		$(k-2)r+3$	$kr+3$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
1	$r+1$	$2r+1$		$(k-1)r+1$	$(k+1)r+1$,

ahol $r = \lceil \frac{q}{2} \rceil$.

A mátrix minden oszlopa néhány egyessel kezdődik, majd egy helytől kezdve a természetes számok következnek növekvő sorrendben: $1, \dots, 1, 1, 2, 3, 4, \dots$. Az

$A_i - A_{i-1}$-be $(1 \leq i \leq k,\ A_0 = \emptyset)$ tartozó oszlopok mind egyformák, ugyanez érvényes $X - A_k$-ra is. $X - A_k$ oszlopai legyenek ilyenek: $1, 2, 3, 4, \ldots$. Másrészt A_1 oszlopai álljanak csupa 1-ből. $A_2 - A_1$ oszlopai A_1 oszlopaihoz képest r–rel vannak eltolva, tehát az elejükön r–rel kevesebb az egyes, az utolsó elemük viszont $r+1$. Általában, $A_i - A_{i-1}$ oszlopai az $A_{i-1} - A_{i-2}$ $(1 < i \leq k)$ oszlopaihoz képest r–rel vannak eltolva. Ennyivel kevesebb az elejükön az egyes, és ennyivel nagyobb az utolsó elemük. $X - A_k$ oszlopai viszont már $2r$–rel vannak eltolva $A_k - A_{k-1}$ oszlopaihoz képest. $A_i - A_{i-1}$ $(1 \leq i \leq k)$ a lánc definíciója miatt nem lehet üres. $X - A_k$ viszont lehet üres. Ekkor a mátrix ilyen oszlopokat nem tartalmaz.

A könnyebb érthetőség kedvéért bemutatjuk a $k = 4$, $q = 4$, $r = 2$ esetet:

A_1	$A_2 - A_1$	$A_3 - A_2$	$A_4 - A_3$	$X - A_4$
1	1	1	1	1
1	1	1	1	2
1	1	1	1	3
1	1	1	1	4
1	1	1	1	5
1	1	1	2	6
1	1	1	3	7
1	1	2	4	8
1	1	3	5	9
1	2	4	6	10
1	3	5	7	11

E mátrixnak a következő, könnyen ellenőrízhető tulajdonságait fogjuk felhasználni:

(i) A sorok száma $(k+1)\lceil q/2 \rceil + 1$.

(ii) Ha $A_i - A_{i-1}$ $(1 < i \leq k)$ egy oszlopában két pozíció azonos számokat tartalmaz (ami csak egy lehet), és $j < i$ akkor $A_j - A_{j-1}$ megfelelő pozícióiban is azonos számok állnak (amik szintén egyesek).

(iii) Kiválasztva egy 1-est $A_i - A_{i-1}$ egy oszlopában, $A_{i+1} - A_i$ egy oszlopának megfelelő pozíciójában nem állhat más, csak 1 vagy 2 vagy 3 vagy $\ldots r+1$. Ha viszont egy 1–től különböző s számot választunk ki $A_i - A_{i-1}$ egy oszlopában, akkor $A_{i+1} - A_i$ egy oszlopának megfelelő pozíciójában csak $s + r$ állhat $(1 \leq i < k)$.

(iv) Ha $k \geq j > i + 1 \geq 2$, akkor $A_j - A_{j-1}$ egy oszlopában kiválasztható $2r + 1$ különböző szám (mégpedig az $1, 2, \ldots, 2r + 1$) úgy, hogy $A_i - A_{i-1}$ minden oszlopában a megfelelő helyeken csupa 1–es álljon. Jegyezzük meg, hogy $2r + 1 \geq q + 1$.

(v) $X - A_k$ egy oszlopában kiválasztható $2r + 1$ $(\geq q + 1)$ különböző szám (mégpedig az $1, 2, \ldots, 2r + 1$) úgy, hogy $A_i - A_{i-1}$ $(1 \leq i \leq k)$ minden oszlopában a megfelelő helyeken csupa 1–es álljon.

Vegyünk láncoknak egy olyan $L_1, \ldots, L_m$ halmazát, melyben minden olyan A, b párra, amelyre $A \neq \emptyset$, $b \notin \mathcal{N}(A)$ teljesül, található egy olyan lánc és benne egy olyan A_i halmaz, melyre

$$A \subseteq A_i \text{ és } b \notin \mathcal{N}(A_i) \tag{2.5}$$

fennálnak. Láncoknak ilyen halmazát kapjuk például, ha A_1-ként minden lehetséges nem üres A-t kiválasztunk. Ekkor $m = 2^{|X|} - 1$, de általában kevesebb lánc is elegendő.

Az $M(L_1), \ldots, M(L_m)$ mátrixokat írjuk egyszerűen egymás alá, de a bennük szereplő számokat úgy módosítsuk, hogy két mátrixban ne szerepeljen ugyanaz a szám. Ha történetesen valamely oszlopban kevesebb, mint $q+1$ különböző érték szerepel, akkor $M(L_1)$-ből vegyünk annyi különböző példányt, teljesen különböző számjegyekkel, hogy már minden oszlopban legyen $q+1$ különböző érték. Az így kapott M mátrix által definiált $\mathcal{J} = \mathcal{J}_{M1q}$ — mint látni fogjuk — éppen $\mathcal{N}$ lesz.

Ehhez két dolgot kell igazolnunk; 1) Ha $b \notin \mathcal{N}(A)$, akkor $b \notin \mathcal{J}(A)$ és 2) ha $b \in \mathcal{N}(A)$, akkor $b \in \mathcal{J}(A)$.

1) Tegyük fel tehát, hogy $b \notin \mathcal{N}(A)$. Ha $A = \emptyset$, akkor a $b \notin \mathcal{J}(\emptyset)$ állítás abból következik, hogy M minden oszlopában legalább $q+1$ különböző érték van. Ha viszont $A \neq \emptyset$, válasszunk egy L_v $(1 \leq v \leq m)$ láncot és benne egy A_i-t (2.5) szerint. Mivel $b \notin \mathcal{N}(A_i) = A_{i+1}$ (lásd (2.3)), így $b \in A_j - A_{j-1}$, $k \geq j > i+1$, vagy $b \in X - A_k$ teljesül. Az első esetben használjuk (iv)-et. Eszerint $M(L_v)$-ben kiválasztható $q+1$ sor, ami $A_i - A_{i-1}$ oszlopaiban csupa egyest tartalmaz, míg b-ben mind a $q+1$ érték különböző. (2.5) és (ii) miatt A minden oszlopában is csupa egyes áll. Tehát valóban $b \notin \mathcal{J}(A)$. Ha $b \in X - A_k$, akkor (v)-öt használva ugyanúgy járunk el.

2) Most tegyük fel, hogy $b \in \mathcal{N}(A)$. Feltehető, hogy $A \neq \emptyset$, mert $\mathcal{N}(\emptyset)$ üres. Tekintsünk egy tetszőleges L_v láncot: $A_1, \ldots, A_k$. Legyen $i = i(L_v) = k+1$, ha $A \cap (X - A_k)$ nem üres, ellenkező esetben viszont i legyen a legnagyobb index, amire $A \cap (A_i - A_{i-1})$ nem üres. $A \subseteq A_i$-ből (2.2) miatt $b \in \mathcal{N}(A) \subseteq \mathcal{N}(A_i) = A_{i+1}$ következik, ha $i < k$. Ellenkező esetben $b \in \mathcal{N}(A_i) = A_i$ (ahol $A_{k+1} = X$).

Vegyünk ki $q+1$ sort M-ből úgy, hogy ezekben a sorokban A minden oszlopa csupa azonos számból álljon. M konstrukciójából következik, hogy a soroknak ugyanazon $M(L_v)$-ben (vagy $M(L_1)$-nek ugyanabban a példányában) kell lenniük. Tehát $A \cap (A_i - A_{i-1})$ az $M(L_v)$ mátrixnak e $q+1$ sorában csupa egyes. Ha $b \in A_{i+1}$, akkor (iii)-ből következik, hogy b-ben e sorokban legfeljebb $r+1 = \lceil q/2 \rceil + 1 \leq q$ különböző érték állhat. Ha $b \in A_i$, akkor b-ben e sorokban ráadásul csupa egyes áll. Tehát $b \in \mathcal{J}(A)$, amint azt bizonyítani kívántuk. □

2.1. *Probléma.* Igaz-e a 2.1 tétel állítása tetszőleges (p, q) $(p < q)$ függőségekre? Elhagyható-e az $\mathcal{N}(\emptyset) = \emptyset$ feltétel?

A választ jelenleg még a (2,3) esetre sem tudjuk.

Ha $p = q$, a helyzet jelentősen megváltozik. Mint azt [6]-ban már megmutattuk, $\mathcal{J} = \mathcal{J}_{Mpp}$ (2.1) és (2.2) mellett még egy fontos feltételt köteles kielégíteni:

2.2. ÁLLÍTÁS.

$$\mathcal{J}_{Mpp}(\mathcal{J}_{Mpp}(A)) = \mathcal{J}_{Mpp}(A). \tag{2.6}$$

Bizonyítás. Az egyik irányú tartalmazás (2.1)–ből és (2.2)–ből következik. Tehát csak azt kell igazolni, hogy $b \in \mathcal{J}(\mathcal{J}(A))$-ból következik $b \in \mathcal{J}(A)$. Vegyük a mátrix sorainak egy olyan halmazát, melyekben A összes oszlopa legfeljebb p különböző értéket tartalmaz. $\mathcal{J}$ definíciója szerint ekkor $\mathcal{J}(A)$ minden oszlopára is teljesül ez. A $b \in \mathcal{J}(\mathcal{J}(A))$ feltevés szerint b–ben is legfeljebb p különböző érték állhat e sorokban, s ez bizonyítja $b \in \mathcal{J}(A)$–t. □

A (2.1), (2.2) és (2.6) feltételeket kielégítő halmaz-halmaz függvényeket *lezárásoknak* nevezi az irodalom. Jól ismert (más formában lásd ARMSTRONG [1], ilyen alakban lásd pl. [5]), hogy $p = 1$–re az állítás megfordítható, azaz minden lezáráshoz található olyan M mátrix, melyre $\mathcal{J}_{M11}$ éppen ez az adott lezárás, vagyis, hogy minden lezárás (1,1)–reprezentálható. A következőkben belátjuk, hogy ugyanez igaz $p = 2$–re is, de $p > 3$–ra már általában nem.

A bizonyításhoz szükségünk van néhány, a lezárásokra vonatkozó, egyszerű fogalomra és állításra, amelyeket itt bizonyítás nélkül közlünk, megtalálhatók pl. [5]–ben. *Zártnak* nevezünk egy A halmazt az $\mathcal{L}$ lezárás szerint, ha $\mathcal{L}(A) = (A)$. A zárt halmazok összességét jelölje $\mathcal{Z} = \mathcal{Z}(\mathcal{L})$. Két zárt halmaz metszete zárt. $\mathcal{L}(A)$ egyenlő az A–t tartalmazó zárt halmazok metszetével. Jelölje továbbá $\mathcal{M} = \mathcal{M}(\mathcal{Z}(\mathcal{L}))$ azon zárt halmazok összességét, amelyek nem kaphatók meg tőlük különböző két zárt halmaz metszeteként. Bármely zárt halmaz előállítható $\mathcal{M}$–beliek metszeteként, következésképpen $\mathcal{L}(A)$ egyenlő az A–t tartalmazó $\mathcal{M}$–beli halmazok metszetével.

2.3. TÉTEL. *Minden lezárás (2,2)–reprezentálható.*

Bizonyítás. Legyen a lezárás $\mathcal{L}$, és $\mathcal{M} = \mathcal{M}(\mathcal{Z}(\mathcal{L})) = \{X, G_1, \dots, G_m\}$. Egyelőre tegyük fel, hogy $m > 1$. Könnyen látható, hogy $G = \mathcal{L}(\emptyset)$ minden zárt halmaznak, s így minden G_i–nek részhalmaza.

Minden G_i–hez készítsünk az M mátrixban két sort, a $2i-1$–ediket és a $2i$–ediket. A G–nek megfelelő helyeken álljon 0. A $G_i - G$–nek megfelelő helyeken mindkét sorban álljon i. A többi helyeken a j–edik sorban $j + 2m$ álljon.

Be kell látnunk, hogy erre az M mátrixra $\mathcal{J}_{M22}(A) = \mathcal{L}(A)$ teljesül minden $A \subseteq X$–re. Azaz $b \in \mathcal{J}_{M22}(A)$ akkor és csak akkor, ha $b \in \mathcal{L}(A)$.

Tegyük fel, hogy $b \notin \mathcal{L}(A)$. Ha $A = \emptyset$, akkor felhasználva azt az állítást, hogy $\mathcal{L}(\emptyset)$ egyenlő az összes $\mathcal{M}$–beli metszetével, található egy G_i, amelyre $b \notin G_i$ teljesül. Ez ad két különböző értéket b oszlopában. X, ami mindig egy G, ad egy újabbat. Egy tetszőleges harmadik sorban vagy 0 áll, vagy újabb érték. Tehát b oszlopában van legalább 3 különböző érték, $b \notin \mathcal{J}(\emptyset)$. Ha $A \neq \emptyset$, akkor $\mathcal{L}(A)$ ismét egyenlő az A–t tartalmazó $\mathcal{M}$–beliek metszetével, tehát van egy olyan G_i, amelyre $A \subseteq G_i$ és $b \notin G_i$ fennállnak. De ekkor a G_i–nek megfelelő két sor A–ban megegyezik, b–ben viszont különbözik. E két sort kiegészítve bármely további sorral, találtunk 3 sort, ami bizonyítja, hogy $b \notin \mathcal{J}_{M22}(A)$ valóban igaz.

Tegyük most fel, megfordítva, hogy $b \in \mathcal{L}(A)$. Két esetet különböztetünk meg. Ha $A \subseteq \mathcal{L}(\emptyset)$, akkor (2.2) miatt $\mathcal{L}(A) \subseteq \mathcal{L}(\emptyset)$. Ekkor $b \in G$, tehát b oszlopában csupa 0 áll. $b \in \mathcal{J}_{M22}(\emptyset)$ valóban következik. Ha viszont $A \not\subseteq \mathcal{L}(\emptyset)$, akkor A-nak van egy $\mathcal{L}(\emptyset)$-n kivüli a eleme. Vegyünk ki 3 sort a mátrixból, amelyek A oszlopaiban legfeljebb két különböző értéket tartalmaznak. Az a oszlopot tekintve, azt nyerjük, hogy a három sor közül kettőnek ugyanahhoz a G_i-hez kell tartoznia. Továbbá fenn kell állnia $A \subseteq G_i$-nek is. G_i zártsága és (2.2) miatt $\mathcal{J}(A) \subseteq G_i$, tehát ez a két sor b-ben is egyforma. $b \in \mathcal{J}_{M22}(A)$ következik.

Az $m = 1$ esetben egy csupa új elemből álló harmadik sor hozzáadása segít. □

Vezessük be a következő jelölést:

$$\mathcal{L}^k{}_n(A) = \begin{cases} A, & \text{ha } |A| < k, \\ X, & \text{ha } |A| \geq k, \end{cases}$$

ahol k egy egész, $1 \leq k \leq |X|$, és $A \subseteq X$. $\mathcal{L}^k{}_n$ nyilván lezárás.

2.4. TÉTEL. *Ha $p > 2$ és $n > 6$, akkor $\mathcal{L}^2{}_n$ nem (p,p)-reprezentálható.*

Bizonyítás. Tegyük fel indirekte, hogy van egy olyan n oszlopú M mátrix, amelyik $\mathcal{L}^2{}_n$-t (p,p)-reprezentálja. Legyen M sorainak száma minimális. Mivel $\mathcal{L}^2{}_n$-ben az üres halmaz lezárása önmaga, ezért M minden oszlopában van legalább $p+1$ különböző érték. Ha az a oszlopban minden érték különböző lenne, akkor $b \in \mathcal{L}^2{}_n(\{a\})$ teljesülne egy tetszőleges $b \in X$-re, ellentmondásban $\mathcal{L}^2{}_n$ definíciójával.

Tegyük fel most, hogy az r és s sorok egyenlőek az a és a b oszlopokban. Definíció szerint, $c \in \mathcal{L}^2{}_n(\{a,b\})$ fennáll tetszőleges $c \in X$-re. Válasszunk ki r-hez és s-hez még $p-1$ egy sort úgy, hogy értékeik c-ben különbözzenek mind r-től, mind s-től, és egymástól is. (Ez megtehető, mert van a c oszlopban is $p+1$ különböző érték.) E $p+1$ sorban a és b egyaránt legfeljebb p különböző értéket vesz fel, tehát c is. Ez csak úgy lehet, ha r és s megegyeznek c-ben is. Vagyis az r és s sorok végig megegyeznek. Ez ellentmondásban van a minimalitással, tehát nem egyezhet meg két sor két különböző helyen.

Tegyük fel, hogy az első oszlopban a t-edik és az u-adik sorok egyeznek meg, míg a második oszlopban az r-edik és s-edik sorok ($t \neq u$, $r \neq s$). Az előző bekezdés szerint $\{t,u\} \neq \{r,s\}$, így csak a következő két esetet különböztetjük meg: (i) $t = r, u \neq s$, (ii) mind a négy sor különböző.

(i) Az első és második oszlop legfeljebb 2 különböző értéket vesz fel e három sorban. $\mathcal{L}^2{}_n$ definíciója miatt minden oszlop is csak legfeljebb két értéket vehet fel ezekben a sorokban. Ha az oszlopok száma több mint 3, akkor kell lenni olyan két oszlopnak, amelyek a fenti három sor valamelyik kettőjében megegyeznek, ellentmondásban a korábban igazolt állítással.

(ii) Az előző esethez hasonlóan belátható, hogy minden oszlop legfeljebb 3 értéket tartalmazhat a t, u, r, s sorokban. Mivel az egyezésekre csak 6-féle lehetőség van, $n > 6$ miatt ismét lesz két olyan oszlop, amely két adott sorban megegyezik. Az ellentmondás bizonyítja állításunkat. □

A most következőkben egy karakterizációt adunk arra, hogy mikor lehet egy lezárást (p,p)–reprezentálni. Egy ehhez szükséges definíció:

Legyen $\mathcal{B} = \{A_{i,j}\}$ az n–elemű X halmaz egy két indexes halmazrendszere, ahol $1 \le i < j \le m$. Akkor mondjuk, hogy teljesíti a háromszögfeltételt, ha minden $i < j < k$–ra $A_{i,j}$, $A_{j,k}$ és $A_{i,k}$ közül bármely kettő metszetét tartalmazza a harmadik.

2.5. LEMMA. *Akkor és csak akkor található olyan n oszlopú és m sorú mátrix melynek i–edik és j–edik sorai éppen az $A_{i,j}$–nek megfelelő helyeken egyeznek meg, ha $\{A_{i,j}\}$ kielégíti a háromszögfeltételt.*

Bizonyítás. A csak akkor rész triviális. Tegyük tehát fel, hogy a háromszögfeltétel teljesül és konstruáljunk egy megfelelő mátrixot. Méghozzá mohó módon, soronként. Tegyük fel, hogy az első k sor már megvan, és kielégíti a feltételt. A $k+1$–edik sor konstruálásánál csak az lehet a probléma, hogy egy helyen két különböző értékkel kell megegyezzen, mondjuk a g–edik és h–adik $(g < h)$ sorokban felvett értékekkel. Ez viszont ellentmond annak, hogy $A_{g,h}$, $A_{g,k+1}$ és $A_{h,k+1}$ kielégítik a háromszögfeltételt. □

Most már meg tudjuk adni a karakterizációt.

2.6. TÉTEL. *Az $\mathcal{L}$ lezárás akkor és csak akkor (p,p)–reprezentálható, ha létezik egy $\mathcal{B} = \{A_{i,j}\}$ $(1 \le i < j \le m)$ halmazrendszer úgy, hogy kielégíti a háromszögfeltételt, a következő alakú halmazok mind zártak $\mathcal{L}$ szerint:*

$$\bigcup_{0 \le r < s \le p} A_{j_r,j_s} \tag{2.7}$$

$(1 \le j_0, j_1, \dots, j_p \le m$ tetszőlegesen rögzített különböző egészek), és minden zárt halmaz előállítható (2.7) alakú halmazok metszeteként.

Bizonyítás. Bizonyítsuk először a következő lemmát.

2.7. LEMMA. *Tegyük fel, hogy az m sorú M mátrix (p,p)–reprezentálja az $\mathcal{L}$ lezárást és M i–edik és j–edik sorai az $A_{i,j}$ halmazban egyeznek meg. $A \subseteq X$ akkor és csak akkor zárt, ha (2.7) alakú halmazok metszete.*

Bizonyítás. A (p,p)–függés definíciója szerint $A \not\to (p,p) \to b$ akkor és csak akkor teljesül egy M mátrixban, ha található $p+1$ sor, amely A oszlopaiban legfeljebb p értéket vesz fel, viszont a b oszlopban mind a $p+1$ érték különböző. Ezt úgy is lehet fogalmazni, hogy valamilyen $1 \le j_0, j_1, \dots, j_p \le m$ különböző egészekre

$$A \subseteq \bigcup_{0 \le r < s \le p} A_{j_r,j_s} \not\ni b$$

teljesül. A tehát zárt akkor és csak akkor, ha ilyen található minden $b \notin A$–hoz. Ez utóbbi viszont akkor és csak akkor teljesül, ha A (2.7) alakú halmazok metszete.

□

Térjünk vissza a tétel bizonyításához. Ha (p,p)–reprezentálható, akkor a reprezentáló M mátrix definiálja az $A_{i,j}$ soregyenlőséghalmazokat. Ezek a 2.5 lemma miatt kielégítik a háromszögfeltételt. Egy (2.7) alakú halmaz ilyen alakú halmazoknak (egynek) a metszete, tehát a 2.7 lemma miatt zárt. Ugyancsak a 2.7 lemmából következik, hogy minden zárt halmaz (2.7) alakúak metszete.

Megfordítva, ha találhatók a tétel feltételeit kielégítő $A_{i,j}$ halmazok, a háromszögfeltétel miatt létezik egy M mátrix, amelyben a soregyenlőséghalmazok éppen $A_{i,j}$–k. Az $\mathcal{L}$ szerint zárt halmazok a tétel feltételei miatt előállnak (2.7) alakú halmazok metszeteként. Megfordítva, a nem zártak nem állíthatók elő, mert a (2.7) alakúak szerint zártak, s tudjuk, hogy zárt halmazok metszetei is zártak. Így a 2.7 lemma befejezi a bizonyítást. □

Sajnos a tétel feltétele algoritmikusan nem hatékony. A következő következmény mutatja azonban, hogy mégse teljesen haszontalan.

2.8. ÁLLÍTÁS. *Tegyük fel, hogy az $\mathcal{L}$ lezárás olyan, hogy $\mathcal{M}(\mathcal{Z}(\mathcal{L})) = \{G_1, \ldots, G_t\}$ zárt az unió képzésére. Ekkor (p,p)–reprezentálható minden p–re.*

Bizonyítás. Tegyük fel először, hogy $t \geq p$. Legyen a 2.6 tételbeli $m = 2t$, $A_{2i-1,2i} = G_i$ $(1 \leq i \leq t)$, míg a többi A legyen az üres halmaz. A rendszer kielégíti a háromszögfeltételt. A (2.7) alakú halmazok a G–k uniói lesznek, a $t \geq p$ feltétel miatt a $p+1$ darab index választható úgy, hogy (2.7) éppen egy (bármelyik) G legyen. A feltétel szerint a G–k uniói mind zártak, másrészt minden zárt halmaz előállítható $\mathcal{M}$–beliek (azaz G–k) metszeteként. Így a 2.6 tételből következik az állítás.

Ha $t < p$, akkor G–k valamelyikét ismételjük meg annyiszor, hogy m elérje a szükséges méretet. □

A következő egyszerű állítást bizonyítás nélkül közöljük.

2.9. ÁLLÍTÁS. *Legyen az $\mathcal{L}$ lezárás egy n–elemű halmazon megadva. Ha $\mathcal{L}$ $(2n,2n)$–reprezentálható, akkor (k,k)–reprezentálható is minden $k > 2n$ esetén.*

Összefoglalva, a kérdés nagyrészt még nyitott maradt:

2.2. *Probléma.* Keressünk egy algoritmikusan hatékony jellemzését azon lezárásoknak, amelyek (p,p)–reprezentálhatóak.

A következő probléma azonban könnyebbnek látszik. Legyen $\mathcal{N}$ egy növelő-monoton függvény az X halmazon. A K halmazt kulcsnak mondjuk, ha $\mathcal{N}(K) = X$. K *minimális kulcs,* ha kulcs, és nincs olyan valódi része, ami kulcs. Könnyen látható, hogy a minimális kulcsok nem tartalmazhatják egymást, így a minimális kulcsok $\mathcal{K}$ rendszere kielégíti a *Sperner-feltételt:* Ha $K_1, K_2 \in \mathcal{K}$, $K_1 \neq K_2$, akkor $K_1 \not\subseteq K_2$. Ezt úgy is mondhatjuk, hogy $\mathcal{K}$ *Sperner-rendszer.* Az X–en adott *Sperner-rendszerről* azt mondjuk, hogy (p,q)–reprezentálható $(p < q)$, ha X–en van egy növelő-monoton függvény, ami (p,q)–reprezentálható és amiben a minimális kulcsok rendszere éppen $\mathcal{K}$.

A (p,p)–reprezentálhatóság definíciója analóg, csak egy lezárás keresendő.

2.3. *Probléma.* Igaz-e hogy minden nem-üres *Sperner-rendszer* (p,q)–reprezentálható bármely $p < q$ mellett?

3. A (p,q) függőség egymáshoz való viszonyai

Jelölje $(p,q) \Rightarrow (p',q')$ azt, hogy ha $b \in \mathcal{J}_{Mpq}(A)$–ból minden M mátrixra következik $b \in \mathcal{J}_{Mp'q'}(A)$. A következő állítások triviálisak.

3.1. LEMMA.

$$(p,q) \Rightarrow (p,q+1)$$
$$(p,q) \Rightarrow (p-1,q).$$

□

Több állítható, ha feltesszük, hogy az M mátrix minden oszlopában legalább m különböző érték áll. Erre az esetre használjuk a $(p,q) \underset{m}{\Rightarrow} (p',q')$ jelölést.

3.2. LEMMA.

$$(p,q) \underset{q+1}{\Rightarrow} (p-1,q-1), \quad de \quad (p,q) \underset{q}{\not\Rightarrow} (p-1,q-1).$$

Bizonyítás. Az első állítás bizonyításához tegyük fel, hogy valamely M mátrixban $b \in \mathcal{J}_{Mpq}(A)$. Bizonyítani kívánjuk, hogy $b \in \mathcal{J}_{M,p-1,q-1}(A)$. Ha ez utóbbi nem teljesülne, akkor a mátrixnak volna q olyan sora, amely A–ban legfeljebb $p-1$ különböző értéket vesz fel, míg b–ben q különböző értéket. A feltevés szerint a b oszlopban van legalább $q+1$ különböző érték. Az előző q sort egészítsük ki egy újabb sorral úgy, hogy b–ben $q+1$ különböző érték legyen. A semelyik oszlopában se lehet több, mint p érték, s ez ellentmondásban van $b \in \mathcal{J}_{Mpq}(A)$–val.

A második állítást egy olyan mátrix bizonyítja, amelyben b oszlopa pontosan q különböző értéket tartalmaz, A minden oszlopa pedig legfeljebb $p-1$–et. □

3.3. LEMMA. $(p,q) \underset{q}{\not\Rightarrow} (1,q-1)$.

Bizonyítás. A következő ellenpélda adja a bizonyítást, ahol az első oszlop A oszlopait jelképezi, a második pedig b–ét.

$$\begin{array}{cc}
1 & 1 \\
1 & 2 \\
\vdots & \vdots \\
1 & q \\
2 & 1 \\
2 & 2 \\
\vdots & \vdots \\
2 & q \\
\vdots & \vdots \\
q & 1 \\
q & 2 \\
\vdots & \vdots \\
q & q
\end{array}$$

□

3.4. LEMMA. $(p,q) \underset{m}{\not\Rightarrow} (1, q-p), \quad (p < q)$.

Bizonyítás. A következő ellenpélda adja a bizonyítást, ahol az első oszlop A oszlopait jelképezi, a második pedig b-ét.

$$\begin{array}{cc}
1 & 1 \\
1 & 2 \\
\vdots & \vdots \\
1 & q-p+1 \\
2 & q-p+2 \\
\vdots & \vdots \\
m & q-p+m
\end{array}$$

□

3.5. LEMMA. $(p,q) \underset{m}{\not\Rightarrow} (p+1, N), \quad (p+1 \leq N)$.

Bizonyítás. Először egy olyan konstrukciót adunk, ami $(p,q) \not\Rightarrow (p+1, N)$-et bizonyítja. A mátrixnak $N+1$ sora van, b-ben sorban az $1, 2, \ldots, N+1$ számok állnak. A oszlopaiban az $1, 2, \ldots, p+1$ számok állhatnak. Az oszlopokat úgy válasszuk meg, hogy bármely $p+1$ sorhoz legyen egy olyan oszlop, amelyikben $p+1$ különböző érték áll. Ha A oszlopainak száma elég nagy, ez elérhető. Könnyen látható, hogy ezen M mátrixban $b \in \mathcal{J}_{Mpp}(A)$ azaz a 3.1 lemma miatt $b \in \mathcal{J}_{Mpq}(A)$. Másrészt azonban $b \notin \mathcal{J}_{M,p+1,N}(A)$.

Már csak úgy kell módosítanunk M-et, hogy minden oszlopban legalább m érték legyen. Írjunk az $N+1+i$-edik sorba csupa $N+1+i$-t $(1 \leq i \leq m-p-1)$. Ez a módosítás nem rontja el a fenti tulajdonságot. □

3.6. TÉTEL. *Legyen* $m > q$.

$$(p, q) \underset{m}{\Rightarrow} (p', q'), \tag{3.1}$$

akkor és csak akkor teljesül, ha $1 \leq p' \leq p$, *és* $q - p \leq q' - p'$. *Ha viszont* $m \leq q$, *akkor* (3.1) *szükséges és elégséges feltétele* $1 \leq p' \leq p$ *és* $q \leq q'$.

Bizonyítás. A 3.1–3.5 lemmákból az állítás könnyen következik. □

Vegyük észre, hogy a 3.5 lemma bizonyításában A-nak elég nagynak kell lennie. Ez azt jelenti, hogy kis méretű A-ra a $b \in \mathcal{J}_{Mpq}(A)$ feltevésből még következik $b \in \mathcal{J}_{M,p+1,N}(A)$. Ha például $|X| = n$ is kisebb, mint ez a határ, akkor itt minden A-ra is teljesül ez a következtetés. Ez veti fel a következő problémát.

3.1. *Probléma.* Milyen méretű A-kra teljesül, hogy $b \in \mathcal{J}_{Mpq}(A)$-ból következik $b \in \mathcal{J}_{M,p+1,N}(A)$ minden M-re (p, q és N rögzített)?

Két speciális esetet bizonyítás nélkül közlünk:

3.7. ÁLLÍTÁS. *Ha* $|A| < \lceil \log(N+1) \rceil$, *akkor* $b \in \mathcal{J}_{M11}(A)$*-ból következik* $b \in \mathcal{J}_{M2N}(A)$ *minden* M *mátrixra, de ha* $|A| \geq \lceil \log(N+1) \rceil$, *akkor nem.* □

3.8. ÁLLÍTÁS. *Ha*

$$|A| < \left\lceil \frac{q+2}{2(q-p+1)} \right\rceil,$$

akkor $b \in \mathcal{J}_{Mpp}(A)$*-ból következik* $b \in \mathcal{J}_{M,p+1,q+1}(A)$ *minden* M *mátrixra, de ha* A *ennél nagyobb, akkor nem.* □

4. Minimális reprezentáció

E fejezetben azzal a kérdéssel foglalkozunk, hogy ha létezik reprezentáció, akkor mi a reprezentáló M mátrix minimális sorszáma. Ha $\mathcal{N}$ egy növelő-monoton függvény, jelölje $s_{pq}(\mathcal{N})$ ezt a minimális sorszámot. A 2.1 tétel biztosítja, hogy $s_{1q}(\mathcal{N})$ létezik. Az ott szereplő konstrukcióban a táblázatban szereplő mátrix sorainak száma (lásd (i)) $= (k+1)\lceil q/2 \rceil + 1 \leq 2nq$. Másrészt annyi ilyen mátrixot veszünk, amennyi a szükséges láncok száma. Ez legfeljebb 2^n. Tehát a következő igaz.

4.1. TÉTEL. *Ha* $q > 1$ *és* $\mathcal{N}$ *egy tetszőleges növelő-monoton függvény, akkor*

$$s_{1q}(\mathcal{N}) < 2qn2^n.$$

□

Ez a becslés lényegesen javítható volna, ha be tudnánk látni, hogy viszonylag kevés lánc elég a (2.5) feltétel kielégítéséhez. A másik javítási lehetőség a láncok tényleges hosszának figyelembevétele. Általában a növelő-monoton függvények szerkezetének tanulmányozása szükséges. Mindenesetre a 4.1 tétel becslése sokkal rosszabb, mint amit $s_{11}(\mathcal{L})$-re ismerünk (n alatt $\lfloor n/2 \rfloor$, lásd [3]).

A továbbiakban csak a $p = q = 2$ esettel foglakozunk. $s_{pp}(\mathcal{K}) = s_p(\mathcal{K})$ ill. $s_{pp}(\mathcal{L}) = s_p(\mathcal{L})$ jelölje a $\mathcal{K}$ *Sperner-rendszer* ill. az $\mathcal{L}$ lezárás (p,p)-reprezentálásához szükséges minimális sorszámot.

4.2. TÉTEL.

$$s_2(\mathcal{K}) < 2\binom{n}{\lfloor \frac{n}{2} \rfloor}.$$

Bizonyítás. A 2.3 tétel bizonyításából kiolvasható. □

Ez az eredmény már csak egy kettes faktorral rosszabb az (1,1) eset becslésénél (lásd [3]).

4.3. TÉTEL. *Van olyan Sperner-rendszer, melyre*

$$s_2(\mathcal{K}) \geq \frac{1}{n^2}\binom{n}{\lfloor \frac{n}{2} \rfloor}.$$

Bizonyítás. A bizonyítás szó szerint ugyanaz, mint [4] bizonyítása. □

4.4. LEMMA.

$$\binom{s_2(\mathcal{L}^k{}_n)}{3} \geq \binom{n}{k-1}.$$

Bizonyítás. Legyen M egy olyan $s_2(\mathcal{L}^k{}_n)$ sorú mátrix, amely $(2,2)$–reprezentálja $\mathcal{L}^k{}_n$–t. Ha B egy $k-1$–elemű részhalmaza X–nek, akkor van három olyan sora M–nek melyek legfeljebb két különböző értéket tartalmaznak B–beli oszlopokban, és van olyan oszlop, ahol három különbözőt. Ha B-hez a (b,c,d) sorok, C–hez pedig az (e,f,g) sorok tartoznak, akkor $(b,c,d) \neq (e,f,g)$, mert különben $B \cup C$ nem lenne kulcs. Azaz minden $k-1$ elemű halmazhoz különböző sorhármas tartozik, vagyis a lehetséges sorhármasok száma legalább akkora mint a $k-1$–elemű oszlophalmazok száma. □

Bizonyos speciális k értékekre $s_2(\mathcal{L}^k{}_n)$–t pontosan is meg tudjuk határozni, másokra pedig jó felső becslést tudunk adni:

4.5. TÉTEL.
(i) $s_2(\mathcal{L}^1{}_n) = 3$,
(ii) $s_2(\mathcal{L}^2{}_n) = 2n$, *ha* $n > 3$,
(iii) $s_2(\mathcal{L}^n{}_n) \leq n+2$.

Bizonyítás. (i) Az $s_2(\mathcal{L}^1{}_n) \geq 3$ egyenlőtlenség a 4.4 lemmából következik. A másik irányt a következő konstrukció adja:

$$\begin{array}{cccc} 0 & 0 & \dots & 0 \\ 1 & 1 & \dots & 1 \\ 2 & 2 & \dots & 2 \end{array}$$

(ii) Az $s_2(\mathcal{L}^2{}_n) \leq 2n$ egyenlőtlenség a 2.3 tétel bizonyításából kiolvasható. A másik irányú egyenlőtlenséget direkt bizonyítjuk. A 4.4 lemma szerint minden oszlophoz van három sor, hogy azok legfeljebb két különböző értéket tartalmaznak ezen oszlopban. Azaz minden oszlophoz van két sor, melyek ott megegyeznek. Belátjuk, hogy a különböző oszlopokhoz tartozó sorpárok diszjunktak.

a) Tegyük fel, hogy az i és j oszlopokban az r és s sorok megegyeznek. Mivel $\{i,j\} \dashv (2,2) \to X$ szükséges, ezért r és s az összes többi oszlopban is meg kell egyezzen. Ekkor r–et elhagyva, a kapott mátrix még mindig reprezentálja $\mathcal{L}^2{}_n$–t.

b) Ha van 3 olyan sor, hogy az i és j oszlopokban is legfeljebb két értéket vesznek fel, akkor az összes többi oszlopban is legfeljebb két különböző értéket vesznek fel. Ekkor viszont, ha $n > 3$, akkor az a) esetre jutunk vissza.

Azt kaptuk tehát, hogy az egyes oszlopokhoz tartozó párok diszjunktak. Jegyezzük meg, hogy $s_2(\mathcal{L}^2{}_2) = 4$, $s_2(\mathcal{L}^2{}_3) = 5$.

(iii) Az $s_2(\mathcal{L}^n{}_n) \leq n + 2$ felső korlátot a következő konstrukció bizonyítja:

$$\begin{matrix} 0 & 0 & \dots & 0 \\ 1 & 1 & \dots & 1 \\ 2 & 0 & \dots & 0 \\ 0 & 2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 2. \end{matrix}$$

□

4.1. *Probléma.* Határozzuk meg $s_2(\mathcal{L}^n{}_n)$–et, $s_2(\mathcal{L}^3{}_n)$–at és $s_2(\mathcal{L}^4{}_n)$–et.

4.2. *Probléma.* Határozzuk meg $s_2(\mathcal{L}^k{}_n)$ nagyságrendjét rögzített k és nagy n esetére.

Az adatbázisok elméletében fontos szerepe van a lezárások direkt szorzatának. Ha $\mathcal{L}$ egy lezárás az U alaphalmazon, $\mathcal{N}$ pedig a V halmazon ($U \cap V = \emptyset$), akkor $\mathcal{L}$ és $\mathcal{N}$ direkt szorzata az $U \cup V$ alaphalmazon a következőképpen definiált függvény:

$$(\mathcal{L} \times \mathcal{N})(A) = \mathcal{L}(A \cap U) \cup \mathcal{N}(A \cap V).$$

Könnyen látható, hogy ez is egy lezárás.

4.6. TÉTEL. *Legyenek $\mathcal{L}$ és $\mathcal{N}$ lezárások az U ill. V alaphalmazokon. Ekkor*

$$s_p(\mathcal{L} \times \mathcal{N}) \leq s_p(\mathcal{L}) + s_p(\mathcal{N}) - p.$$

Bizonyítás. Ha $\mathcal{L}$ vagy $\mathcal{N}$ valamelyike nem reprezentálható, akkor értelmezzük az s_p–t végtelennek. Ekkor az egyenlőtlenség triviális. Legyen tehát M_1 egy minimális reprezentáló mátrixa $\mathcal{L}$–nek, M_2 pedig $\mathcal{N}$–nek. Képezzük a következő mátrixot:

$$M = \begin{pmatrix} Q & W \\ R & T \\ Y & P \end{pmatrix},$$

ahol Q az M_1–ből az utolsó p sor levágásával keletkezik. W az M_2 első sora annyiszor véve, ahány sora van Q–nak, R az M_1, T pedig az M_2 első p sora, P úgy keletkezik M_2–ből, hogy első p sorát levágjuk, végül Y pedig M_1 utolsó sora annyiszor véve, ahány sora van P–nek.

Tegyük fel, hogy $y \in U$ és $y \notin (\mathcal{L} \times \mathcal{N})(A)$ valamely $A \subseteq U \cap V$–re. Ez pontosan akkor van, ha $y \notin \mathcal{L}(A \cap U)$, azaz M_1–nek van $p+1$ sora, melyek A–n legfeljebb p értéket vesznek fel, de y–n mind különböznek. Ez a $p+1$ sor fellép az M mátrixban is, mégpedig úgy, hogy a V–re való kiterjesztésükben legfeljebb p különböző érték lép fel, azaz $y \notin \mathcal{J}_{M_{pp}}(A)$.

Legyen most $y \in (\mathcal{L} \times \mathcal{N})(A)$, és legyenek $r_0, \ldots, r_p$ olyan különböző sorok, melyek A–ban legfeljebb p értéket vesznek fel. Ha az $r_0, \ldots, r_p$ közül legalább kettő esik az Y részre, illetve még R utolsó sora valamelyik, akkor y–ban az a két sor megegyezik, tehát $r_0, \ldots, r_p$ legfeljebb p értéket vesz fel y–on. Ha viszont csak legfeljebb egy sor esik az r_j–k közül a mátrix „alsó" felébe, akkor az A–n és y–on felvett értékeik ugyanazok, mint az M_1 megfelelő soráé, tehát $y \in \mathcal{J}_{M_{pp}}$.

4.3. *Probléma.* Milyen feltételekkel áll a 4.6 tételben egyenlőség?

IRODALOM

[1] ARMSTRONG, W. W., "Dependency structures of data base relationships ", *Information Processing* **74** (North Holland, Amsterdam, 1974), 580–583.

[2] CODD, E. F., "A relational model of data for large shared data banks ", *Comm. ACM* **13** (1970), 377–387.

[3] DEMETROVICS, J., FÜREDI, Z. and KATONA, G. O. H., "Minimum matrix representation of closure operations ", *Discrete Appl. Math.* **11** (1985), 115–128.

[4] DEMETROVICS, J. and GYEPESI, GY., "A note on minimal matrix representation of closure operations ", *Combinatorica* **3** (1983), 177–180.

[5] DEMETROVICS, J. and KATONA , G. O. H., "Combinatorial problems of database models ", *Coll. Math. Soc. János Bolyai* **42**, Algebra, Combinatorics and Logic in Computer Science, Győr (Hungary), 1983, 331–353.

[6] DEMETROVICS, J. and KATONA, G. O. H., "Extremal combinatorial problems of database models ", MFDBS' 87, 1st Symposium on Mathematical Fundamentals of Database Systems, Dresden, GDR, January, 1987, Springer Verlag, Lecture Notes in Computer Science, 1987.

(Beérkezett: 1988. november 28.)
(Végleges változat: 1991. szeptember 5.)

DEMETROVICS JÁNOS
MTA SZÁMÍTÁSTECHNIKAI ÉS AUTOMATIZÁLÁSI KUTATÓ INTÉZET
1132 BUDAPEST, VICTOR HUGO U. 18

KATONA GYULA ÉS SALI ATTILA
MTA MATEMATIKAI KUTATÓ INTÉZET
1053 BUDAPEST, REÁLTANODA U. 13–15

BRANCHING DEPENDENCY IN RELATIONAL DATABASES

J. DEMETROVICS, G. O. H. KATONA and A. SALI

A relational database model can be considered as a matrix, where the individuums are the rows and the attributes are the columns. A well-known notion in this theory is the notion of

functional dependencies. If A is a set of columns and b is a column then b is said to depend on A if there are no two rows different in b but identical in A. This concept is generalized here. Let $p \leq q$ be positive integers. We say that b (p,q)–depend on A if there are no $q+1$ rows having all different entries in b but having at most p different entries in the columns of A. Several results known for the traditional dependencies are generalized.

Alkalmazott Matematikai Lapok 15 (1990–91) 197–211

A TÖBBDIMENZIÓS DIRICHLET ELOSZLÁS MOMENTUM ÉS LIKELIHOOD ILLESZTÉSÉNEK NUMERIKUS MEGOLDÁSÁRÓL*

KLAFSZKY EMIL — GRUBERT LÁSZLÓ
Budapest Miskolc

Műszaki berendezések teljesítmény vizsgálatánál gyakran felmerülő feladat a mérési eredményekre *Dirichlet eloszlás* illesztése. A dolgozat célja a likelihood illesztés numerikus megoldása. Minthogy a likelihood illesztés iteratív megoldásához „jó" induló megoldás kívánatos, ezért induló megoldásnak a „kevert momentum" megegyezési elven alapuló momentum becslést használjuk. Numerikus példák tapasztalatai azt mutatják, hogy számos esetben a momentumbecslés elég jól megközelíti a likelihood becslést, azonban szélsőséges esetben nagyon el is térhet.

A dolgozatot egyúttal a *Dirichlet eloszlás* áttekintésének is szántuk, ezért a főbb eredményeket elöljáróban összefoglaljuk, így alakult ki a dolgozat alábbi 6 fejezetre tagolása:

1. Dirichlet eloszlás definíciója, peremeloszlások.

Az n dimenziós tér standard szimplexét jelöljük S_n-nel, azaz

$$S_n := \left\{(x_1, \dots, x_n) \mid x_i \geq 0,\ i = 1, \dots, n;\ \sum_{i=1}^{n} x_i \leq 1\right\}.$$

A $(\xi_1, \dots, \xi_n)$ véletlen vektort $(p_1, \dots, p_n; p_{n+1})$ paraméterű *Dirichlet eloszlásnak* nevezzük, ha $f(x_1 \dots, x_n)$ sűrűségfüggvénye az S szimplexen kívül zérus, az S szimplexen pedig

(1.1) $$f(x_1, \dots, x_n) = \\ = \frac{\Gamma(p_1 + \dots + p_n + p_{n+1})}{\Gamma(p_1) \dots \Gamma(p_n)\Gamma(p_{n+1})} x_1^{p_1 - 1} \dots x_n^{p_n - 1}(1 - x_1 - \dots - x_n)^{p_{n+1} - 1},$$

ahol $p_1 > 0, \dots, p_n > 0, p_{n+1} > 0$.
Jól ismert az alábbi lemma, amely a *többdimenziós Dirichlet eloszlás* definíciójának jogosultságát igazolja.

*A dolgozat az EGPO-59/86 kutatási pályázat keretében készült

1.1 LEMMA. *Ha* $(\xi_1,\dots,\xi_n)$ *egy* $(p_1,\dots,p_n;p_{n+1})$ *paraméterű Dirichlet eloszlású n-dimenziós véletlen vektor, akkor a* $(\xi_1,\dots,\xi_k)$ $(k\le n)$ *k dimenziós véletlen vektor* $(p_1,\dots,p_k;p_{k+1}+\dots+p_n+p_{n+1})$ *paraméterű Dirichlet eloszlású.*

2. Dirichlet eloszlás momentumai

Az alábbiakban az elsőrendű egyszerű és kevert momentumokat számítjuk ki. A számolás egyszerűsítése érdekében vezessük be a

$$\xi_{n+1} := 1-\xi_1-\dots-\xi_n$$

véletlen számot, valamint a

$$p := p_1+\dots+p_n+p_{n+1}$$

számot.

Az alábbi lemma elemi számolással nyerhető.

2.1 LEMMA. *a.) A szimpla momentum értéke:*

$$M(\xi_j)=\frac{p_j}{p},\ (j=1,\dots,n,n+1). \tag{2.1}$$

b.) A kevert momentum értéke:

$$M(\xi_1\dots\xi_n\cdot\xi_{n+1})=\frac{p_1\cdots p_n\cdot p_{n+1}}{p(p+1)\dots(p+n)}. \tag{2.2}$$

Bizonyítás. A bizonyításban felhasználjuk a Γ függvény alábbi tulajdonságát:

$$\Gamma(x+1)=x\Gamma(x),$$

valamint az ebből rekurzívan nyerhető

$$\Gamma(x+n+1)=(x+n)\dots(x+1)x\Gamma(x)$$

tulajdonságot.
a., A szimpla momentum esete:

$$M(\xi_j)=\int\limits_{R^n} x_j\frac{\Gamma(p)}{\Gamma(p_1)\dots\Gamma(p_{n+1})}x_1^{p_1-1}\dots x_n^{p_n-1}(1-x_1-\dots-x_n)^{p_{n+1}-1}\times$$
$$\times\, dx_1\dots dx_n=\frac{p_j}{p}\int\limits_{R^n}\frac{p\Gamma(p)}{p_j\Gamma(p_1)\dots\Gamma(p_j)\Gamma(p_{n+1})}\times$$
$$\times\, x_1^{p_1-1}\dots x_j^{(p_j+1)-1}\dots x_n^{p_n-1}(1-x_1-\dots-x_n)^{p_{n+1}-1}dx_1\dots dx_n=$$
$$=\frac{p_j}{p}.$$

b., A kevert momentum esete:

$$M(\xi_1 \dots \xi_n \xi_{n+1}) = \int\limits_{R^n} x_1 \dots x_n (1 - x_1 - \dots - x_n) \frac{\Gamma(p)}{\Gamma(p_1) \dots \Gamma(p_{n+1})} \times$$
$$\times x_1^{p_1-1} \dots x_n^{p_n-1} (1 - x_1 - \dots - x_n)^{p_{n+1}-1} dx_1 \dots dx_n =$$
$$= \frac{p}{p(p+1)\dots(p+n)} \quad \square$$

3. Dirichlet eloszlás entrópiája

A következőkben a tömörebb írásmód érdekében szükségünk lesz a gamma függvény logaritmusának derivált függvényére, az úgynevezett ψ vagy digamma függvényre:

$$\psi(x) := \frac{\partial}{\partial x} \log \Gamma(x).$$

A definícióból egyszerű számolással adódik, hogy

$$(3.1) \quad \frac{\partial}{\partial p_j} \left[\frac{\Gamma(p_1 + \dots + p_{n+1})}{\Gamma(p_1) \dots \Gamma(p_{n+1})} \right] =$$
$$= \left[\frac{\Gamma(p_1 + \dots + p_{n+1})}{\Gamma(p_1) \dots \Gamma(p_{n+1})} \right] [\psi(p_1 + \dots + p_{n+1}) - \psi(p_j)].$$

A következő formula a későbbiekben többször felhasználásra kerül, ezért azt lemmaként mondjuk ki.

3.1 LEMMA. *Legyen a* $\xi = (\xi_1, \dots, \xi_n)$ *véletlen vektor* $(p_1, \dots, p_n; p_{n+1})$ *paraméterű Dirichlet eloszlású, akkor*

$$(3.2) \qquad M(\log \xi_j) = \psi(p_j) - \psi(p), \qquad (j = 1, \dots n, n+1),$$

ahol $\xi_{n+1} := 1 - \xi_1 - \dots - \xi_n$ *és* $p := p_1 + \dots + p_n + p_{n+1}$.

Bizonyítás. Induljunk ki a sűrűségfüggvényből:

$$\int\limits_{R^n} \frac{\Gamma(p)}{\Gamma(p_1) \dots \Gamma(p_{n+1})} x_1^{p_1-1} \dots x_n^{p_n-1} (1 - x_1 - \dots - x_n)^{p_{n+1}-1} dx_1 \dots dx_n = 1.$$

Deriváljuk p_j szerint, és használjuk a (3.1) formulát. Így kapjuk, hogy

$$\int\limits_{R^n} \frac{\Gamma(p)}{\Gamma(p_1) \dots \Gamma(p_{n+1})} [\psi(p) - \psi(p_j)] x_1^{p_1-1} \dots x_n^{p_n-1} \cdot$$
$$\cdot (1 - x_1 - \dots - x_n)^{p_{n+1}-1} dx_1 \dots dx_n +$$
$$+ \int\limits_{R^n} \frac{\Gamma(p)}{\Gamma(p_1) \dots \Gamma(p_{n+1})} \log x_j \cdot x_1^{p_1-1} \dots x_n^{p_n-1} \cdot$$
$$\cdot (1 - x_1 - \dots - x_n)^{p_{n+1}-1} dx_1 \dots dx_n = 0$$

Ebből $M(\log \xi_j) = \psi(p_j) - \psi(p)$ □
Emlékeztetünk az entrópia definíciójára: Ha a ξ véletlen vektor sűrűségfüggvénye $f(x)$, akkor a

$$H(\xi) = \int\limits_{R^n} f(x) \log \frac{1}{f(x)} dx$$

értéket (ha létezik) nevezzük entrópiának.

3.2 TÉTEL. *Legyen a ξ véletlen vektor $(p_1, \dots, p_n;\ p_{n+1})$ paraméterű Dirichlet eloszlású, ekkor*

$$\begin{aligned} H(\xi) = \log \frac{\Gamma(p_1)\dots\Gamma(p_{n+1})}{\Gamma(p)} &+ \\ &+ (p_1 - 1)(\psi(p) - \psi(p_1)) + \\ &+ (p_2 - 1)(\psi(p) - \psi(p_2)) + \\ &\quad \vdots \\ &+ (p_n - 1)(\psi(p) - \psi(p_n)) + \\ &+ (p_{n+1} - 1)(\psi(p) - \psi(p_{n+1})). \end{aligned}$$

Bizonyítás. Az entrópia definícióját a *Dirichlet eloszlásra* felírva kapjuk, hogy

$$\begin{aligned} H(\xi) = -\int \left\{ \frac{\Gamma(p)}{\Gamma(p_1)\dots\Gamma(p_{n+1})} x_1^{p_1-1} \dots x_n^{p_n-1}(1 - x_1 - \dots - x_n)^{p_{n+1}-1} \right\} \times \\ \times \left\{ \log \frac{\Gamma(p)}{\Gamma(p_1)\dots\Gamma(p_{n+1})} + (p_1 - 1)\log x_1 + \dots + (p_{n+1} - 1) \times \right. \\ \left. \times \log(1 - x_1 - \dots - x_n) \right\} dx_1 \dots dx_n = \\ = -\log \frac{\Gamma(p)}{\Gamma(p_1)\dots\Gamma(p_{n+1})} - (p_1 - 1)(\psi(p_1) - \psi(p)) - \dots - \\ - (p_{n+1} - 1)(\psi(p_{n+1} - \psi(p)) \ \square \end{aligned}$$

KÖVETKEZMÉNY. *Egy és kétdimenziós peremeloszlások entrópiájára kapjuk, hogy*

$$\begin{aligned} H(\xi_j) = \log \frac{\Gamma(p_j)\Gamma(p - p_j)}{\Gamma(p)} &+ \\ &+ (p_j - 1)(\psi(p) - \psi(p_j)) + \\ &(p - p_j - 1)(\psi(p) - \psi(p - p_j)). \end{aligned}$$

$$\begin{aligned} H(\xi_i, \xi_j) = \log \frac{\Gamma(p_i)\Gamma(p_j)\Gamma(p - p_i - p_j)}{\Gamma(p)} &+ \\ &+ (p_j - 1)(\psi(p) - \psi(p_j)) + \\ &+ (p_i - 1)(\psi(p) - \psi(p_i)) + \\ &+ (p - p_i - p_j - 1)(\psi(p) - \psi(p - p_i - p_j)). \end{aligned}$$

Bizonyítás. A peremeloszlásoknál (1. fejezet) láttuk, hogy ξ_j *Dirichlet eloszlású* $(p_j;\ p-p_j)$ paraméterrel, (ξ_i,ξ_j) *Dirichlet eloszlású* $(p_i,p_j;\ p-p_i-p_j)$ paraméterekkel. Ezekre a tételt alkalmazva kapjuk a következményt □

4. A Dirichlet eloszlás kölcsönös információja.

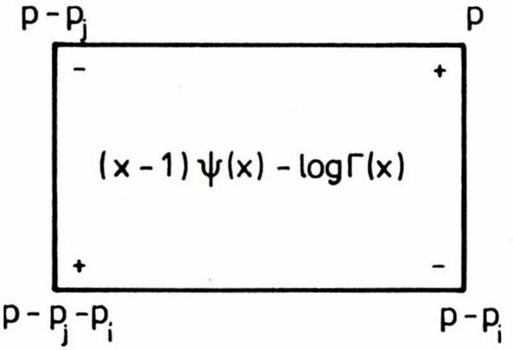
Legyen a $(\xi_1,\dots,\xi_n)$ véletlen vektor $(p_1,\dots p_n;\ p_{n+1})$ paraméterű *Dirichlet eloszlású*. Először a két koordináta közötti kölcsönös információt határozzuk meg.

4.1. LEMMA. *Két koordináta közötti kölcsönös információ:*

$$
\begin{aligned}
I(\xi_i\wedge\xi_j) =&(p-1)\psi(p)-\log\Gamma(p)-\\
&-(p-p_i-1)\psi(p-p_i)+\log\Gamma(p-p_i)-\\
&-(p-p_j-1)\psi(p-p_j)+\log\Gamma(p-p_j)+\\
&+(p-p_i-p_j-1)\psi(p-p_i-p_j)-\log\Gamma(p-p_i-p_j)
\end{aligned}
\tag{4.1}
$$

(*ahol* $p:=p_1+\dots+p_n+p_{n+1}$), $i\neq j$.

Az eredményt sémán szemléltetve:



1. ábra

Bizonyítás. Tudjuk, hogy

$$I(\xi_i\wedge\xi_j)=H(\xi_i)+H(\xi_j)-H(\xi_i,\xi_j).$$

Ebbe az entrópia értékeket behelyettesítve és rendezve kapjuk a lemma állítását □

Megjegyzés. A lemma eredménye átvihető koordináta csoportok közötti kölcsönös információ meghatározására is: Jelöljük $J:=\{1,\dots,n\}$ és legyen $J_1\subset J$, $J_2\subset J$ úgy, hogy $J_1\cap J_2=\emptyset$. Legyenek $\xi^{(1)}:=(\xi_j)$, $j\in J_1$ és $\xi^{(2)}:=(\xi_j)$, $j\in J_2$; valamint

$$q_1:=\sum_{j\in J_1}p_j \qquad \text{és} \qquad q_2:=\sum_{j\in J_2}p_j.$$

Ezen jelöléssel a $\xi^{(1)}$ és $\xi^{(2)}$ véletlen vektorok közötti kölcsönös információ:

$$
\begin{aligned}
I(\xi^{(1)} \wedge \xi^{(2)}) = &+ (p-1)\psi(p) - \log\Gamma(p) - \\
&- (p-q_1-1)\psi(p-q_1) + \log\Gamma(p-q_1) - \\
&- (p-q_2-1)\psi(p-q_2) + \log\Gamma(p-q_2) + \\
&+ (p-q_1-q_2-1)\psi(p-q_1-q_2) - \log\Gamma(p-q_1-q_2).
\end{aligned} \tag{4.2}
$$

Belátása a peremeloszlások entrópiájából azonnal adódik.

Az összes koordináta közötti kölcsönös információra az alábbi formulát kapjuk:

4.2. TÉTEL. *A koordináták közötti kölcsönös információ:*

$$
\begin{aligned}
I(\xi_1 \wedge \xi_2 \wedge \ldots \wedge \xi_n) = &\log\Gamma(p) + \\
&+ n\{(p-1)\psi(p) - \log\Gamma(p)\} - \\
&- \sum_{j=1}^{n}\{(p-p_j-1)\psi(p-p_j) - \log\Gamma(p-p_j)\} + \\
&+ (p_{n+1}-1)\psi(p_{n+1}) - \log\Gamma(p_{n+1}).
\end{aligned} \tag{4.3}
$$

Bizonyítás. Tudjuk, hogy

$$I(\xi_1 \wedge \xi_2 \wedge \ldots \wedge \xi_n) = H(\xi_1) + H(\xi_2) + \ldots + H(\xi_n) - H(\xi_1, \xi_2, \ldots, \xi_n).$$

Ebbe az entrópia értéket behelyettesítve és rendezve kapjuk a lemma állítását □

5. A Dirichlet eloszlás likelihood illesztése.

A feladat: Adott $(\eta_1, \ldots, \eta_n)$ véletlen vektorhoz keressük azt a *Dirichlet eloszlású* $(\xi_1, \ldots, \xi_n)$ véletlen vektort, amely $(\eta_1, \ldots, \eta_n)$-hez legjobban hasonló. Emlékeztetünk arra, hogy a *Gauss eloszlásnál* az alábbi eredményt ismerjük: Adott $(\eta_1, \ldots, \eta_n)$ véletlen vektorhoz a legjobb likelihood illesztésű *Gauss eloszlású* $(\xi_1, \ldots, \xi_n)$ véletlen vektor az, melyre

$$
\begin{aligned}
M(\xi_j) &= M(\eta_j), \qquad \forall j, \\
M(\xi_i \cdot \xi_j) &= M(\eta_i \cdot \eta_j), \qquad \forall i, j.
\end{aligned}
$$

Ezzel rokonságot mutató tételt tudunk kimondani *Dirichlet eloszlásra* is.

5.1. TÉTEL. *Adott $(\eta_1, \ldots, \eta_n)$ véletlen vektorhoz a legjobb likelihood illesztésű Dirichlet eloszlású $(\xi_1, \ldots, \xi_n)$ véletlen vektor az, melyre*

$$
\begin{aligned}
M(\log\xi_j) &= M(\log\eta_j), \qquad \forall_j, \\
M(\log(1 - \sum_{j=1}^{n}\xi_j)) &= M(\log(1 - \sum_{j=1}^{n}\eta_j)).
\end{aligned} \tag{5.1}
$$

Bizonyítás. Legyen $(\eta_1, \dots, \eta_n)$ eloszlásfüggvénye $G(x)$, $(\xi_1, \dots, \xi_n)$ sűrűségfüggvénye pedig $f(x)$, ekkor a hasonlóság mértéke

$$L := \int_{R^n} \log f(x) dG.$$

Felhasználva, hogy f *Dirichlet sűrűségfüggvény*, kapjuk, hogy

$$L = \int_{R^n} \Big\{ \log \frac{\Gamma(p_1 + \dots + p_{n+1})}{\Gamma(p_1) \dots \Gamma(p_{n+1})} + \\ + (p_1 - 1) \log x_1 + \dots + (p_{n+1} - 1) log(1 - x_1 - \dots - x_n) \Big\} dG,$$

vagy tömörebb formában:

$$L = \log \frac{\Gamma(p_1 + \dots + p_{n+1})}{\Gamma(p_1) \dots \Gamma(p_{n+1})} + (p_1 - 1) M(\log \eta_1) + \dots + \\ + (p_n - 1) M(\log \eta_n) + (p_{n+1} - 1) M \left(\log \left(1 - \sum_{j=1}^{n} \eta_j \right) \right).$$

A feladat az $L(p_1, \dots, p_{n+1})$ függvény maximalizálása.
Az L függvénynek lokális maximuma ott lehet, ahol a változók szerinti parciális deriváltja zérus, azaz

$$\psi(p_1 + \dots + p_{n+1}) - \psi(p_j) + M(\log \eta_j) = 0, \qquad \forall j,$$

$$\psi(p_1 + \dots + p_{n+1}) - \psi(p_{n+1}) + M \left(\log \left(1 - \sum_{j=1}^{n} \eta_j \right) \right) = 0.$$

Tudjuk, hogy (használva a 3. fejezet lemmáját azaz (3.2)-t) $M(\log \xi_j) = \psi(p_j) - \psi(p)$, ezt behelyettesítve kapjuk a tétel állítását □

6. Algoritmus a momentum és a likelihood illesztésre.

Ebben a fejezetben célunk a likelihood illesztés numerikus kivitele, azonban ezt megelőzően megadjuk a momentum illesztés numerikus kivitelét.

A momentumillesztést két dolog is indokolja. Egyrészt, mint később numerikus példákon látni fogjuk, „elég jó" becslést ad, kivitele lényegesen egyszerűbb mint a likelihood illesztésé, másrészt alkalmas kiindulópont a likelihood illesztés iteratív eljárásához.

Az előzőekben (2. és 5. fejezet) kiszámoltuk a momentum és likelihood illesztés feltételeit. Ezeket, mint végeredményeket, a könnyebb egybevethetőség érdekében egymás mellé helyezzük:

Momentum illesztés hipotézise:

$$M(\xi_j) = M(\eta_j), \quad (j = 1, \dots, n),$$
$$M\left(\xi_1\xi_2\dots\xi_n\left(1-\sum_{j=1}^{n}\xi_j\right)\right) = M\left(\eta_1\eta_2\dots\eta_n\left(1-\sum_{j=1}^{n}\eta_j\right)\right).$$

Likelihood illesztés hipotézise:

$$M(\log\xi_j) = M(\log\eta_j), \quad (j = 1, \dots, n),$$
$$M\left(\log\left(1-\sum_{j=1}^{n}\xi_j\right)\right) = M\left(\log\left(1-\sum_{j=1}^{n}\eta_j\right)\right).$$

Megjegyezzük, hogy a hipotézisekből automatikusan kapjuk:

Momentum illesztésnél:

$$M\left(1-\sum_{j=1}^{n}\xi_j\right) = M\left(1-\sum_{j=1}^{n}\eta_j\right).$$

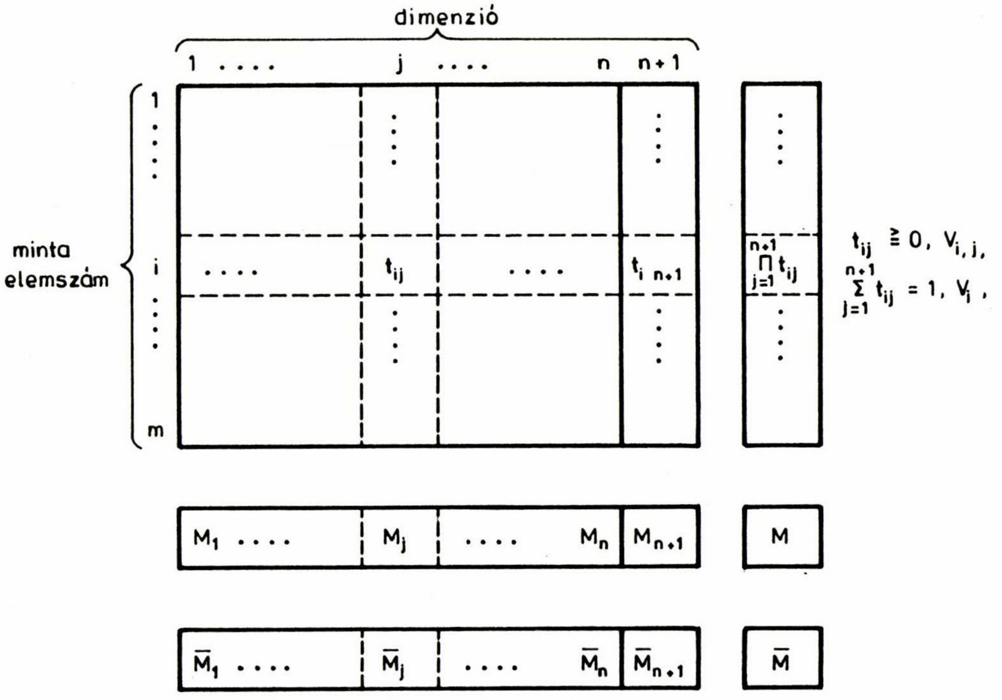
Likelihood illesztésnél:

$$M\left(\log\xi_1\xi_2\dots\xi_n\left(1-\sum_{j=1}^{n}\xi_j\right)\right) = M\left(\log\eta_1\eta_2\dots\eta_n\left(1-\sum_{j=1}^{n}\eta_j\right)\right).$$

Amennyiben ezen automatikusan adódó összefüggéseket a hipotézisek mellé tennénk, akkor a formulákban erős szimmetriát tapasztalnánk.

6.1. A minta aggregálása

Legyen $(\eta_1, \dots, \eta_n)$ egy empirikus eloszlás, azaz egy minta, amelyet célszerűen táblázatosan helyezünk el. A táblázatot kiegészítjük egy $(n+1)$-edik „dimenzió" oszloppal és egy a szorzatokat tartalmazó oszloppal, valamint a táblázat alján helyezzük el az illesztésekhez szükséges momentumokat, azaz a tábla azon aggregált értékeit, amelyekre szükségünk lesz:



2. ábra

ahol M_j, M a momentum, $\overline{M}_j, \overline{M}$ a likelihood illesztéshez szükséges aggregált értékek:

$$M_j := M(\eta_j) = \frac{1}{m}\sum_{i=1}^{m} t_{ij}, \qquad (j = 1, \dots, n+1),$$

$$M := M(\eta_1 \dots \eta_n \eta_{n+1}) = \frac{1}{m}\sum_{i=1}^{m}\prod_{j=1}^{n+1} t_{ij},$$

és

$$\overline{M}_j := M(\log \eta_j) = \frac{1}{m}\sum_{i=1}^{m} \log t_{ij}, \qquad (j = 1, \dots, n+1),$$

$$\overline{M} := M(\log(\eta_1 \dots \eta_n \eta_{n+1})) = \frac{1}{m}\sum_{i=1}^{m} \log \prod_{j=1}^{n+1} t_{ij},$$

Megjegyezzük, hogy

$$\overline{M} = \sum_{j=1}^{n+1} \overline{M}_j,$$

így ez a tábla aggregálásakor ellenőrzésként szolgálhat.

6.2 Momentum illesztés numerikus megoldása

A 2. fejezet eredményeit felhasználva, illesztési feladatunk olyan $p_1, \ldots, p_n, p_{n+1}$ (legyen mint előzőkben is $p := p_1 + \ldots + p_n + p_{n+1}$) pozitív számok meghatározása, melyekre

$$\frac{p_j}{p} = M_j, \quad (j = 1, \ldots, n+1),$$

$$\frac{p_1 \cdots p_n p_{n+1}}{p(p+1)\ldots(p+n)} = M$$

teljesül.
Amennyiben p már ismert, akkor p_j értékek azonnal adódnak. Feladatunk tehát csak a p meghatározása. Az elsőt a másodikba helyettesítve p-re az alábbi egyenletet kapjuk:

$$\left(1 + \frac{1}{p}\right)\left(1 + \frac{2}{p}\right)\ldots\left(1 + \frac{n}{p}\right) = \frac{\prod_{j=1}^{n+1} M_j}{M},$$

amelyet iterációval egyszerűen tudunk megoldani.

6.3 Likelihood illesztés numerikus megoldása

A 3. fejezet eredményeit felhasználva illesztési feladatunk olyan $p_1, \ldots p_n, p_{n+1}$ pozitív számok meghatározása, melyekre:

$$\psi(p_j) - \psi(p) = \overline{M}_j, \quad (j = 1, \ldots, n+1). \tag{6.1}$$

A (6.1) egyenletrendszer megoldására az alábbi iteratív eljárást javasoljuk.
A ψ függvény alakja a logaritmus függvény alakjával erős hasonlóságot mutat, ez azt sugallja, hogy a ψ függvényből válasszuk le a logaritmus függvényt.
Legyen

$$\delta(x) := \log x - \psi(x). \tag{6.2}$$

Mint ismeretes [1] a δ függvény végtelen sor előállítása:

$$\delta(x) = \sum_{k=0}^{\infty} \left[\frac{1}{x+k} - \log\left(1 + \frac{1}{x+k}\right)\right] \tag{6.3}$$

Az (6.1) egyenletrendszerbe (6.2)-t behelyettesítve kapjuk, hogy

$$\log(p_j) - \delta(p_j) - \log(p) + \delta(p) = \overline{M}_j. \tag{6.4}$$

A (6.4)-ből az alábbi iterációs formula rendezhető ki:

$$p_j^{l+1} := K_j \sqrt{\frac{\exp \delta(p_j^{(l)})}{\exp \delta(p^{(l)})} p^{(l)} p_j^{(l)}}, \tag{6.5}$$

ahol

$$K_j = \exp(\frac{\overline{M}_j}{2}), \quad (j = 1, \dots, n+1).$$

A (6.5) iterációs formula konvergenciáját egzaktan nem tudtuk belátni, azonban számos numerikus példán kipróbálva konvergenciát tapasztaltunk.

Numerikus számolásnál a $\delta(x)$ függvényt a $\delta_N(x)$ véges tagú sorával helyettesítettük (az N tagszám a kompúter kódban input paraméter; tapasztalataink szerint N=100 illetve N=1000 választásnál a p_j értékek első három jegyében nincs eltérés).

Az iteráció kiindulásául szolgáló p_j^0 $(j = 1. \dots, n+1)$ értékeknek célszerű a momentum illesztéssel kapott p_j értékeket venni.

6.4 Az illesztés pontossága

Az illesztés eltérésére a likelihood függvény negatívjának exponensét, az úgynevezett exponenciális inaccuranciát használjuk:

$$E = \exp(-L) \tag{6.6}$$

A loggamma függvényt N tagú sorával közelítjük, azaz

$$\log \Gamma(x) \approx x \log N - \log x + \sum_{k=1}^{N} \log \frac{k}{x+k}.$$

Ezen közelítéssel egyszerű számolással adódik az 5. fejezetből, hogy

$$\begin{aligned} L_N = {} & \sum_{j=1}^{n+1} \log p_j - \log p - n \sum_{k=1}^{N} \log k + \sum_{j=1}^{n+1} \sum_{k=1}^{N} \log(p_j + k) - \\ & - \sum_{k=1}^{N} \log \left(k + \sum_{j=1}^{n+1} p_j \right) + \sum_{j=1}^{n+1} p_j \overline{M}_j - \overline{M}. \end{aligned} \tag{6.7}$$

Ezt (6.6)-ba helyettesítve megkapjuk az N sortagszámtól függő E_N eltérést.

6.5 Néhány numerikus feladat

1. feladat:

A MINTA (7 DB 4 DIMENZIÓS):	1	2	3	4	5	
1	0,0300	0,7000	0,1000	0,0200	0,1500	6,300E-06
2	0,5000	0,0500	0,0300	0,2000	0,2200	3,300E-05
3	0,0080	0,3000	0,0300	0,6000	0,0620	2,678E-06
4	0,0700	0,7000	0,2000	0,0080	0,0220	1,725E-06
5	0,8000	0,0800	0,0700	0,0030	0,0470	6,317E-07
6	0,0500	0,1000	0,0300	0,0300	0,7900	3,555E-06
7	0,8000	0,0030	0,0040	0,0060	0,1870	1,077E-08
	0,3226	0,2761	0,0663	0,1239	0,2111	6,843E-06
	-2,1613	-2,2215	-3,2303	-3,6132	-2,1398	-13,3662

AZ ITERÁLT PARAMÉTEREK: (pontosság = 0,00001 , sor − tagszám = 400)

lépés	p1	p2	p3	p4	p5	Eltérés
0	0,6443	0,5515	0,1324	0,2474	0,4217	0,02453
1	0,5404	0,5323	1,7146	0,4247	0,5960	0,11598
.	.	.	.	.	.	.
17	0,6608	0,6419	0,4225	0,3708	0,6678	0,01367

2. feladat:

A MINTA (4 DB 4 DIMENZIÓS):	1	2	3	4	5	
1	0,0300	0,7000	0,1000	0,0200	0,1500	6,300E–06
2	0,5000	0,0500	0,0300	0,2000	0,2200	3,300E–05
3	0,0080	0,3000	0,0300	0,6000	0,0620	2,678E–06
4	0,0700	0,7000	0,2000	0,0080	0,0220	1,725E–06
	0,1520	0,4375	0,0900	0,2070	0,1135	1,093E–05
	–2,9218	–1,2283	–2,7313	–2,7152	–2,5021	–12,0987

AZ ITERÁLT PARAMÉTEREK: (pontosság = 0,00001 , sor – tagszám = 400)

lépés	p1	p2	p3	p4	p5	Eltérés
0	0,4038	1,1624	0,2391	0,5500	0,3016	0,03453
1	0,4906	1,0984	0,8216	0,4972	0,7266	0,02323
.	.	.	.	.	.	.
7	0,5608	1,4584	0,6110	0,6156	0,6818	0,01965

3. feladat:

A MINTA (3 DB 4 DIMENZIÓS):	1	2	3	4	5	
1	0,0100	0,2000	0,1500	0,3000	0,3400	3,060E–05
2	0,0080	0,1800	0,0200	0,7000	0,0920	1,855E–06
3	0,0500	0,2200	0,0210	0,5000	0,2090	2,414E–05
	0,0227	0,2000	0,0637	0,5000	0,2137	1,886E–05
	–4,1431	–1,6128	–3,2241	–0,7513	–1,6767	–11,4080

AZ ITERÁLT PARAMÉTEREK: (pontosság = 0,00001 , sor – tagszám = 400)

lépés	p1	p2	p3	p4	p5	Eltérés
0	0,4285	3,7809	1,2036	9,4523	4,0393	0,00243
1	0,7561	3,9906	1,1875	9,3069	3,9766	0,00213
2	0,7039	4,1189	1,1937	9,3172	3,9833	0,00211

4. feladat:

A MINTA (30 DB 4 DIMENZIÓS):

	1	2	3	4	5	
1	0,1761	0,3462	0,2360	0,0506	0,1911	1,390E–04
2	0,2444	0,1543	0,1956	0,2200	0,1857	3,014E–04
3	0,0068	0,2069	0,3315	0,1338	0,3208	2,015E–05
4	0,1742	0,1298	0,1601	0,2351	0,3008	2,560E–04
5	0,2370	0,3244	0,2220	0,0175	0,1992	5,936E–05
6	0,2011	0,2761	0,1793	0,0673	0,2763	1,850E–04
7	0,2832	0,1586	0,0406	0,3224	0,1951	1,148E–04
8	0,1735	0,1184	0,3162	0,0387	0,3531	8,880E–05
9	0,0637	0,2542	0,2354	0,3517	0,0949	1,273E–04
10	0,0447	0,4425	0,3864	0,0648	0,0616	3,053E–05
11	0,1578	0,0807	0,3221	0,3788	0,0606	9,417E–05
12	0,2021	0,2202	0,0406	0,2778	0,2592	1,303E–04
13	0,0380	0,2492	0,3990	0,2320	0,0817	7,168E–05
14	0,2359	0,0122	0,5058	0,0885	0,1575	2,034E–05
15	0,0582	0,3789	0,1664	0,3385	0,0581	7,206E–05
16	0,3806	0,1834	0,3487	0,0407	0,0467	4,624E–05
17	0,4805	0,0874	0,1132	0,0579	0,2610	7,182E–05
18	0,3590	0,3014	0,1754	0,0509	0,1134	1,094E–04
19	0,1014	0,3111	0,3021	0,1831	0,1024	1,785E–04
20	0,1253	0,3690	0,0221	0,2230	0,2605	5,940E–05
21	0,0813	0,1485	0,2616	0,3325	0,1761	1,850E–04
22	0,1196	0,4276	0,2261	0,0422	0,1846	8,996E–05
23	0,0332	0,2395	0,2599	0,2680	0,1994	1,104E–04
24	0,3420	0,0209	0,1131	0,2026	0,3213	5,270E–05
25	0,0307	0,1618	0,2991	0,1149	0,3935	6,713E–05
26	0,3432	0,0396	0,2919	0,0476	0,2779	5,234E–05
27	0,2777	0,2061	0,1763	0,0774	0,2626	2,050E–04
28	0,2965	0,1621	0,0211	0,2825	0,2379	6,803E–05
29	0,1371	0,3717	0,0038	0,1922	0,2952	1,099E–05
30	0,2648	0,1221	0,1901	0,1514	0,2715	2,528E–04
	0,1890	0,2168	0,2180	0,1695	0,2067	1,090E–04
	–1,9789	–1,7669	–1,8440	–2,0689	–1,7212	–9,3800

AZ ITERÁLT PARAMÉTEREK:	(pontosság = 0,00001	, sor – tagszám = 400)				
lépés	p1	p2	p3	p4	p5	Eltérés
0	1,5291	1,7544	1,7643	1,3713	1,6722	0,03230
1	1,5180	1,7631	1,6996	1,4060	1,7757	0,03192
2	1,5217	1,7743	1,6867	1,4225	1,8196	0,03189

IRODALOM

[1] KORN, G.A. és KORN, T.M., *Matamatikai Kézikönyv Műszakiaknak* (Műszaki Könyvkiadó, 1975).

[2] MARDIA, K.V. and KENT, J.T. and BIBBY, J.M., *Multivariate Analysis* (Academic Press, 1979).

[3] PRÉKOPA, A., *Valószínűségelmélet műszaki alkalmazásokkal* (Műszaki Könyvkiadó, 1972).

(Beérkezett: 1988. október 3.)

KLAFSZKY EMIL
NME MATEMATIKAI INTÉZET
MISKOLC, EGYETEMVÁROS 3515

GRUBERT LÁSZLÓ
LKM SZÁMÍTÁSTECHNIKA
MISKOLC, DIÓSGYŐR 3540

ON NUMERICAL SOLUTION OF THE LIKELIHOOD AND MOMENTS FITTING OF MULTIVARIATE DIRICHLET DISTRIBUTIONS

E. KLAFSZKY— L. GRUBERT

The paper deals with finding both the likelihood and the mixed moments fitting of *multivariate Dirichlet distributions*. More over the different estimates are compared. In the first section of the paper crucial results for the general fitting problem are reviewed.

KÖNYVISMERTETÉS

COMPUTER ALGEBRA /Lecture Notes in Pure and Applied Mathematics Series /113/
Szerkesztette: DAVID V. CHUDNOVSKY és RICHARD D. JENKS
Marcel Decker, New York and Basel, 1989. 256 oldal.

A kötet az 1984 április 5-6 között *New Yorkban* megrendezett, több mint 400 résztvevős, „*Computer algebra as a Tool for Research in Mathematics and Phisics*" című nemzetközi konferencia meghívott előadóinak az előadásanyagát tartalmazza. A konferencia célja a már meglévő szimbolikus algebrai programrendszereknek és alkalmazásainak a bemutatása, valamint továbbfejlesztési lehetőségeinek vizsgálata volt. A konferenciával párhuzamosan a *New York University*-n kiállítás nyílt, melyen interaktív módon bemutatták a legfontosabb szimbolikus algebrai rendszereket. (MACSYMA, MAPLE, mu MATH, REDUCE, SCRATCHPAD, SMP, CAYLEY stb.)
Az előadások általában nem tartalmazzák az egyes programrendszerek részletes ismertetését, helyette a fejezetek végén gazdag irodalomjegyzék található.
A továbbiakban a kötetet fejezetenként ismertetjük.

1. „*Use of Computer Algebra for Diophantine and Differential Equations*" D. V. CHUDNOVSKY, G. V. CHUDNOVSKY
   Ez a fejezet differenciál algebrai módszereket mutat be (pld. a *Padé approximációs technikát*), melyeket diofantoszi approximációval és diofantoszi egyenletekkel kapcsolatos problémák megoldására alkalmaz. A szerzők olyan eredményeket ismertetnek, amelyek bár könnyen megfogalmazhatók, bizonyításuk számítógép használata nélkül igen nehézkes. (pld. *Ramanujan egyenlőség*) A bizonyításoknál az IBM SCRATCHPAD régi és új változatát, illetve az SMP rendszert használják.

2. „*Computer Algebra and Hilbert Modular Equation*" HARVEY COHN
   A szerző a legegyszerűbb *Hilbert-féle moduláris egyenletet* konstruálja meg számítógép segítségével. Nem használja az eddigi formális algebrai programrendszereket, bár utal arra, hogy korábbi vizsgálatoknál más szerzők sikerrel alkalmazták a MACSYMA-t és a REDUCE-t.

3. „*Some CAYLEY Examples*" MICHAEL C. SLATTERY
   Ezt a formális algebrai programnyelvet az algoritmikus csoportelméleti eredmények alkalmazása és hozzáférhetősége érdekében fejlesztették ki. A CAYLEY lehetővé teszi a felhasználó számára, hogy definiáljon egy véges vagy végesen prezentált csoportot (generátorokkal és relációkkal, permutációkkal vagy véges test feletti mátrixokkal). Ha megadtuk a csoportot, a CAYLEY egy sor általános csoportelméleti számítást tud végezni (pld. p-Sylowok, centralizátorok meghatározása). Hasznos oktatási és kutatási segédeszköz, elsősorban kis példák konstruálásánál.
   Jelen fejezetben a szerző által írt példaprogramok láthatók egy H csoport x automorfizmussal való bővítésére.

4. „*Phisics, Ramanujan, and Computer Algebra*" GEORGE E. ANDREWS
   Ez a fejezet kétféle módszert mutat a *Rogers-Ramanujan egyenlőségek* bizonyítására. Ezen módszereknek van statisztikus fizikai alkalmazásuk is. A fejezetben új, egyszerű bizonyítást

láthatunk GÖLLNITZ „mod 6" partíciós tételére is. A bizonyításoknál az IBM SCRATCHPAD programrendszert használják.

5 *„POLYPAK: An Algebraic Processor for Computations in Celestial Mechanics"* DIETER S. SCHIMDT
Az égi mechanikai vizsgálatokhoz gyakran van szükség hosszú algebrai számításokra. (Pld. DELAUNAY 20 évet töltött a hold helyzetének számításával.) Így ezen a területen is hasznos a számítógép alkalmazása. Elsősorban különféle sorokkal végzett műveletek fordulnak elő igen gyakran.
Ez a fejezet ismerteti a POLYPAK programrendszert, amely valós és komplex többváltozós sorokkal képes többféle műveletet végezni nagy hatékonysággal.

6 *„Computer Algebra and Definite Integrals"* RICHARD ASKEY
A fejezet többdimenziós beta-integrálokkal foglalkozik. (*Selberg beta-integrál, Mehta-Dyson sejtés, Andrews sejtés, MacDonald-Morris sejtés.*) Vázolja ezen integrálok kapcsolatát és rámutat a MACSYMA rendszer hasznosságára ezen integrálok vizsgálatánál illetve bizonyításánál.

7 *„Algebraic Computation and Structures"* JAMES H. DAVENPORT
A fejezet azzal a kérdéssel foglalkozik, hogy milyen típusú formális algebrai számításokat lehet számítógéppel elvégezni. Kapcsolódik olyan általános problémákhoz is, mint a kanonikus reprezentáció kérdése és a ritkaság. Érinti a konstruktív algebra kérdéskörét is: foglalkozik a konstruktív gcd tartomány és a konstruktív faktorizációs tartomány kapcsolatával, bebizonyítja, hogy konstruktív gcd tartomány feletti polinomgyűrű konstruktív gcd, míg konstruktív faktorizációs tartományra az analóg állítás nem igaz. A bizonyítás során algoritmust ad többhatározatlanú polinomgyűrűben legnagyobb közös osztó meghatározására.

8 *„Computation of Galois Groups from Polynomials over the Rationals"* DAVID J. FORD, JOHN MCKAY
Egy racionális együtthatós polinom *Galois csoportjának* meghatározása matematikai értelemben eddig is megoldott volt. Csak számítási szempontból okoz nehézséget $n!$ fokú $n+1$ változós polinom faktorizációja. A szerzők hatékony módszert dolgoztak ki a *Galois csoport* meghatározására és a jövőben szeretnék több algebrai programrendszeren tesztelni. Eddig az IBM PC-n futó ALGEB nyelven és a MAPLE rendszeren próbálták ki.
A fejezet kitér a fordított problémára is: mely csoportok állnak elő *Galois csoportként*. Ez a probléma megoldatlan, csak részeredmények vannak.

9 *„On the Use SCRATCHPAD in the Construction of Convolution Algorithms"* LOUIS AUSLANDER, ALLAN J. SILBERGER
A szerzők a digitális jelfeldolgozásban alkalmazható algoritmust dolgoztak ki a racionális számtest feletti végesen generált algebrák egy osztályában műveletek végzésére. Az algoritmus használja a racionális számtest algebrai bővítései feletti kínai maradéktételt és eleget tesz bizonyos stabilitási feltételeknek. Az algoritmust SCRATCHPAD és SCRATCHPAD-típusú rendszerekben lehet megvalósítani.

10 *„Automated Generation of Optimized Convolution Algorithms"* JAMES W. COOLEY
A szerző a SCRATCHPAD programrendszert használja többhatározatlanú polinomgyűrű faktorgyűrűiben algebrai műveletek végzésére. A kapott output a szerző által írt programok inputja, amely optimalizálja az elvégzendő összeadások sorrendjét.

11 *„Manual for the System PNCRE"* ROBERT RILEY
A PNCRE rendszer 44 Fortran nyelvű szubrutinból áll, amelyet egy főprogram hív meg. Eredetileg csomóelméleti problémák vizsgálatához készült. Az $SL(2,C)$ egy megadott részcsoportját tudja különböző szempontokból megvizsgálni.

A kötet végén található az április 6-i vita jegyzőkönyve, amelyben a következő kérdésekkel foglalkoztak a konferencia résztvevői:
- milyen lesz a technológiai fejlődés hatása a programrendszerekre
- milyen lesz az algoritmusok fejlődésének hatása
- milyen fejlesztések lennének hasznosak a jövőben
- milyen új alkalmazási területek vannak
- mit lehet várni a mesterséges intelligencia kutatásoktól
- milyen korlátjai vannak a szimbolikus számítási rendszereknek

Összefoglalva: a kötet hasznos lehet mind a matematikai algoritmusokkal foglalkozó kutatók számára, mind olyan kutatók, elsősorban matematikusok és fizikusok számára, akik kutatásaikhoz számítógépet kívánnak használni.

Horváth Erzsébet

HÍREK és KÖZLEMÉNYEK

Megalakult a Magyar Operációkutatási Társaság!

Örömmel tudatjuk, hogy a Magyar Operációkutatási Társaság (MOT) hivatalosan is megalakult, mivel a Fővárosi Bíróság nyilvántartásba vette.

Az alakuló közgyűlés az MOT örökös tiszteletbeli elnökévé Prékopa Andrást, majd titkos szavazással – két évi időtartamra – Ziermann Margitot az öttagú vezetőség elnökévé, Rapcsák Tamást alelnökévé, Csendes Tibort, Komlósi Sándort és Szántai Tamást tagjaivá választotta. A háromtagú Ellenőrző Bizottság elnöke Molnár Sándor, tagjai Varga László és Tétényi Tamás lettek. A vezetőség titkárnak Szántai Tamást választotta.

Új tagok jelentkezését várjuk!

A társaság fontosabb postai adatai:

Levelezési cím:	1111 Budapest, Kende u. 13–17.
Telefonszám:	149–7531
Fax:	129–7866
Telex:	22–4694
Ella postafiók:	2780
Bankszámlaszám:	529–043419–5

Ziermann Margit
az MOT elnöke

A kiadásért felelős az ELTE TTK dékánja
Szedte a KLTE Matematikai Intézete és Nyomta az ELTE Sokszorosító Üzeme
Felelős vezető: Arató Tamás
Budapest, 1992. - ELTE 92091
Megjelent: 19,25 (A/5) ív terjedelemben
500 példányban
HU ISSN 0133-3395

ÚTMUTATÁS A SZERZŐKNEK

Az Alkalmazott Matematikai Lapok csak magyar nyelvű dolgozatokat közöl. A kéziratok gépelését, olyan formában kérjük, hogy minden gépelt oldal 25, egyenként átlag 50 betűhelyes sort tartalmazzon. A közlésre szánt dolgozatokat három példányban kell beküldeni. Előnyben részesülnek a TEX-ben elkészített dolgozatok. Ezeket két kinyomtatott példány kíséretében diszketten kérjük beadni.

A kéziratok szerkezeti felépítésének a következő követelményeket kell kielégíteni. A fejlécnek tartalmaznia kell a dolgozat címét, a szerző teljes nevét, valamint annak a városnak a nevét, ahol a szerző dolgozik. A fejléc után egy, képletet nem tartalmazó, legfeljebb 200 szóból álló kivonatot kell minden esetben megadni. A dolgozatot címmel ellátott szakaszokra kell bontani, és az egyes szakaszokat arab sorszámmal kell ellátni. Az esetleges bevezetésnek mindig az első szakaszt kell alkotnia. Az irodalomjegyzék mindig az utolsó szakasz kell, hogy legyen, és azt nem kell sorszámmal ellátni. Az irodalomjegyzék után, a kézirat befejezéseképpen fel kell tüntetni a szerző teljes nevét és a munkahelye (illetve lakása) pontos postai címét. A dolgozatban előforduló képleteket szakaszonként újrakezdődően, a képlet előtt két zárójel közé írt kettős számozással kell azonosítani. Természetesen nem szükséges minden képletet számozással ellátni. Az esetleges definíciókat és tételeket (segédtételeket és lemmákat) ugyancsak szakaszonként újrakezdődő, kettős számozással kell ellátni. Kérjük a szerzőket, hogy ezeket, valamint a tételek bizonyítását a szövegben kellő módon emeljék ki. Minden dolgozathoz csatolni kell egy angol, német, francia vagy orosz nyelvű, külön oldalra gépelt összefoglalót. Amennyiben lehetésges, kérjük a nyomtatás számára különösen nehézkes matematikai jelölések használatának az elkerülését.

A dolgozatok ábráit és az esetleges lábjegyzeteket a dolgozat végén, különálló lapokon kérjük beküldeni. Mind az ábrákat, mind a lábjegyzeteket a dolgozat szakaszokra bontásától független, folytatólagos arab sorszámozással kell ellátni. Az ábrák elhelyezését a dolgozat megfelelő helyén, széljegyzetként feltüntetett, ábraazonosító sorszámokkal kell megadni. A lábjegyzetekre a dolgozaton belül az azonosító sorszám felső indexkénti használatával lehet hivatkozni.

Az irodalmi hivatkozások formája a következő. Minden hivatkozást fel kell sorolni a dolgozat végén található irodalomjegyzékben, a szerzők, illetve társszerzők esetén az első szerző neve szerint alfabetikus sorrendben úgy, hogy a cirill betűs szerzők nevét a Mathematical Reviews átírási szabályai szerint latin betűsre kell átírni. A folyóiratban megjelent cikkekre [1], a könyvekre [5], a kötetben megjelent dolgozatokra [4], a disszertációkra [3] és a gépi program leírásokra [2] a következő minta szerint kell hivatkozni:

[1] Farkas, J., Über die Theorie der einfachen Ungleichungen, *Journal für die reine und angewandte Mathematik* **124** (1902) 1-27.

[2] Kéri, G., „DUALSIMP", rutin a CDC 3300-as gépekre (Magyar Tudományos Akadémia Számítástechnikai és Automatizálási Kutató Intézete, CDC 3300 felhasználói ismertetők 2. 1973. május) 19-20.

[3] Prékopa, A., „Sztochasztikus rendszerek optimalizálási problémáiról", doktori értekezés. Magyar Tudományos Akadémia, Budapest, 1970.

[4] Prabhu, N. U., „Recent research on the ruin problem of collective risk theory", in: *Inventory Control and Water Storage* Ed. A. Prékopa (János Bolyai Mathematical Society and North-Holland Publishing Company, Amsterdam-London, 1973) 221-228.

[5] Zoutendijk, G., *Methods of Feasible Directions* (Elseiver Publishing Company, Amsterdam and New York, 1960).

A dolgozatok szövegében az irodalmi hivatkozás számait szögletes zárójelben kell megadni, mint például [5] vagy [4, 76-78]. A szerzők a dolgozatukról 50 darab ingyenes különlenyomatot kapnak. A dolgozatok után szerzői díjat az Alkalmazott Matematikai Lapok nem fizet.

Ára: 96 Ft

TARTALOMJEGYZÉK

INDEX

Alkalmazott matematikai lapok

(10)

1990-91/3-4

A MAGYAR TUDOMÁNYOS AKADÉMIA MATEMATIKAI ÉS FIZIKAI TUDOMÁNYOK OSZTÁLYÁNAK KÖZLEMÉNYEI

15. KÖTET

ALKALMAZOTT MATEMATIKAI LAPOK

A MAGYAR TUDOMÁNYOS AKADÉMIA MATEMATIKAI ÉS FIZIKAI TUDOMÁNYOK OSZTÁLYÁNAK KÖZLEMÉNYEI

XV. kötet 3-4. szám

Szerkesztőség és kiadóhivatal: 1117 Budapest, Bogdánfy út 10/B.

Az Alkalmazott Matematikai Lapok változó terjedelmű füzetekben jelenik meg, és olyan eredeti tudományos cikkeket publikál, amelyek a gyakorlatban, vagy más tudományokban közvetlenül felhasználható új matematikai eredményt tartalmaznak, illetve már ismert, de színvonalas matematikai apparátus újszerű és jelentős alkalmazását mutatják be. A folyóirat közöl cikk formájában megírt, új tudományos eredménynek számító programokat, és olyan, külföldi folyóiratban már publikált dolgozatokat, amelyek magyar nyelven történő megjelentetése elősegítheti az elért eredmények minél előbbi, széles körű hazai felhasználását. A szerkesztőbizottság bizonyos időnként lehetővé kívánja tenni, hogy a legjobb cikkek nemzetközi folyóiratok különszámaként angol nyelven is megjelenhessenek.

A folyóirat feladata a Magyar Tudományos Akadémia III. (Matematikai és Fizikai) Osztályának munkájára vonatkozó közlemények, könyvismertetések stb. publikálása is.

A kéziratok a főszerkesztőhöz, vagy a szerkesztőbizottság bármely tagjához beküldhetők. A főszerkesztő címe:

Benczúr András, főszerkesztő
1117 Budapest, Bogdánfy út 10/B.

Közlésre el nem fogadott kéziratokat a szerkesztőség lehetőleg visszajuttat a szerzőhöz, de a beküldött kéziratok megőrzéséért vagy továbbításáért felelősséget nem vállal.

Az Alkalmazott Matematikai Lapok előfizetési ára kötetenként 192 forint (a XVI. kötettől 850 forint). Megrendelések a szerkesztőség címén lehetségesek (átutalásokat az ELTE MNB 232-90142-6207 számlaszámára a 9015567 munkaszám és 694 utalványozási kód megjelöléssel kérjük).

A Magyar Tudományos Akadémia III. (Matematikai és Fizikai) Osztálya a következő idegen nyelvű folyóiratokat adja ki:

1. Acta Mathematica Hungaricae,
2. Acta Physica Hungaricae,
3. Studia Scientiarum Mathematicarum Hungarica.

Alkalmazott Matematikai Lapok 15 (1990-91) 219–266

ELDÖNTHETŐSÉGI KÉRDÉSEK FATRANSZFORMÁCIÓ OSZTÁLYOK ÁLTAL GENERÁLT MONOIDOKBAN

FÜLÖP ZOLTÁN

Szeged

A dolgozatban determinisztikus felszálló és determinisztikus leszálló fatranszformáció osztályok által generált monoidokat vizsgálunk. Megadunk egy olyan eljárást, amellyel a fenti típusú fatranszformáció osztályoknak egy előre rögzített, véges de nem túl általános P halmaza esetén a következő algoritmus konstruálható meg. A P által a $\circ$ kompozíció művelettel generált monoid tetszőleges $X_1 \circ \ldots \circ X_m$ és $Y_1 \circ \ldots \circ Y_n$ elemei esetén ($X_i, Y_j \in P$) az algoritmussal eldönthető, hogy az

$$X_1 \circ \ldots \circ X_m \subseteq Y_1 \circ \ldots \circ Y_n$$

tartalmazás teljesül-e.

Ezután az eljárást alkalmazzuk az

$$M = \{DL, LDL, NDL, LNDL, H, LH, NH\}$$

és

$$S = \{DF, LDF, NDF, LNDF, H, LH, NH\}$$

halmazokra, ahol az M halmaz a determinisztikus leszálló fatranszformációk osztályából és ennek további hat nevezetes részosztályából áll, míg az S halmaz az M megfelelője a felszálló esetre.

1. Bevezetés

Fatranszformátorokkal a 70-es évek eleje óta foglalkoznak az elméleti számítástudományban. Ezen a kutatási területen a fa mindig az univerzális algebrából ismert *term*et (polinomszimbólumot) jelent, maga a fatranszformátor pedig olyan eszköz, amellyel — véges állapothalmaza és véges számú átírási szabálya segítségével — fákat lehet áttranszformálni fákba. Ily módon minden fatranszformátor indukál egy *term*ek feletti relációt amelyet fatranszformációnak hívunk és amely azon (p, q) *term* párokból áll amelyekre igaz, hogy a fatranszformátorral p-t át lehet transzformálni q-ba.

Több fajta fatranszformátor is ismeretes. Legelőször az ún. felszálló fatranszformátor került bevezetésre [Rou]-ban és [Tha1]-ben majd a leszálló fatranszformátor [Tha2]-ben. Ezekkel az eszközökkel leírható a magas szintű programozási nyelvek egyik igen elterjedt fordítási módja, a szintaxis–vezérelt fordítás [Eng5]. Az elnevezések pedig onnan adódnak, hogy a felszálló fatranszformátor a fa gyökerétől a levelei felé, tehát fölfelé haladva transzformálja át a fát, míg a leszálló fatranszformátor az ellenkező irányban dolgozik. Később, a transzformációs képesség fokozására, bevezettek a fenti két típusúnál hatékonyabb fatranszformátorokat is.

Ilyenek a makró fatranszformátor [CouFra], [Eng4], [EngVog1], az attributumos fatranszformátor [Eng4], [Fül1], [Bar1], a magas szintű fatranszformátor [EngVog4], a moduláris fatranszformátor [EngVog3] és a magas szintű moduláris fatranszformátor [Vog2]. Ezen fatranszformátorokkal már igen általános fatranszformációk realizálhatók. Példaként említjük meg, hogy az attributumos fatranszformátorokkal és a makró fatranszformátorokkal modellezhetők a fordítóprogramok készítésénél igen gyakran alkalmazott attributumos fordítások [Knu], valamint azt az [EngVog3]-ban bizonyított tényt, hogy a determinisztikus moduláris fatranszformátorok által indukált fatranszformációk osztálya megegyezik a fákon értelmezett primitív rekurzív függvények osztályával.

Napjainkban a fatranszformátorok elmélete, a szintaxis–vezérelt fordításokat meghaladva, az elméleti számítástudomány egyre több területéhez kapcsolódik. Ennek indoklására, a teljesség igénye nélkül, csak az egyik legfontosabbat, a *term* átíró rendszereket [Hue], [DerJou] hozzuk példaként (maga a fatranszformátor is egy speciális *term* átíró rendszer) és két olyan új eredményt, amelyek a [DTHL]-ben bevezetett *ground* fatranszformátorok tanulmányozásából eredtek: [DTHL]-ben bizonyítást nyert, hogy a *ground term* átíró rendszerek *Church–Rosser tulajdonsága* eldönthető; [FülVág8]-ban pedig gyors algoritmust adtunk a *ground term* egyenletrendszerek szóproblémájának az eldöntésére.

A fatranszformációkkal, tehát a fatranszformátorok által indukált relációkkal kapcsolatos egyik legfontosabb fogalom a kompozíció. Szinte valamennyi, ezen témakörrel foglalkozó munkában előjön és központi helyen fog állni jelen dolgozatban is. Ezért, mielőtt rátérnénk dolgozatunk eredményeinek ismertetésére, felidézzük a kompozíció fogalmát és a vele kapcsolatban elért legfontosabb eredményeket csoportosítjuk.

Ha adott két (mindegy, hogy milyen típusú) A és B fatranszformátor amelyek rendre a τ_A és τ_B fatranszformációkat indukálják, akkor ezen két fatranszformáció $\tau_A \circ \tau_B$-vel jelölt kompozícióján értjük azon (t, s) *term* párok halmazát melyekre teljesül, hogy van olyan r *term* amelyre $(t, r) \in \tau_A$ és $(r, s) \in \tau_B$. Tehát a fatranszormációk kompozíciója nem más, mint relációk szokásos kompozíciója. A kompozíció természetes módon kiterjeszthető fatranszformációkból álló osztályokra is: az Y és Z osztályok $Y \circ Z$ kompozícióján értjük azon $\tau \circ \sigma$ fatranszformációk osztályát, ahol $\tau \in Y$ és $\sigma \in Z$. Azt mondjuk, hogy az Y osztály zárt a kompozícióra, ha $Y \circ Y \subseteq Y$. Továbbá tetszőleges n természetes szám esetén az Yn-edik hatványát a következő módon értelmezzük: $Y^n = Y$, ha $n = 1$ és $Y^n = Y^{n-1} \circ Y$ különben.

A továbbiakban már konkrét fatranszformáció osztályokról is beszélünk. A felszálló fatranszformációk osztályát F-fel, a leszálló fatranszformációk osztályát L-lel, a homomorfizmus fatranszformációk osztályát pedig H-val jelöljük. Ezen jelölések előtt a D, L, N és az LN prefixek is lehetnek amelyekkel együtt a jelölés a szóban forgó osztálynak rendre a determinisztikus, a lineáris, a nemtörlő és a lineáris nemtörlő részosztályát jelenti. A prefixeket lehet kombinálni is, például LDF a lineáris, determinisztikus felszálló fatranszformációk osztályát jelenti.

A fatranszformáció osztályok kompozícióira vonatkozó eredmények legtöbbje az alábbi, egymást átfedő négy problémakör valamelyikére vonatkozik.

(a) Előállítható-e egy fatranszformáció osztály mint más, nála egyszerűbb struktúrájú fatranszformáció osztályok kompozíciója? Az ilyen előállításokat dekompozíciós eredményeknek is szokás nevezni. Sok dekompozíció található a [Bak], [Engl, 2], [EngFil], [EngVogl], [Füll-3], [FülVágl,4] dolgozatokban és a [GécSte] könyvben. Konkrét példák a $DL = LNDL \circ H$ és a $H = NH \circ LH$ dekompozíciók, l. [Engl] és [FülVágl].

(b) Melyek azok a fatranszformáció osztályok amelyek zártak a kompozícióra nézve? Ha egy osztály nem zárt, akkor a növekvő hatványai végtelen hierarchiát alkotnak-e a tartalmazásra nézve, vagy pedig a hierarchia valahonnan kezdve összeesik? Ezekre nézve az [ArnDau], [Bar2], [DTHL], [Eng2,3,6], [EngVogl,3], [FülVágl,3,4], [GécSte], [VágFül] és a [Vogl] munkákat ajánljuk. Konkrét példaként az [Eng6]-ban igazolt nevezetes hierarchia tételeket említjük, melyek szerint $L^n \subset L^{n+1}$ és $F^n \subset F^{n+1}$ minden $n \geq 1$-re. Ugyanakkor $DL^2 = DL$, tehát DL zárt a kompozícióra nézve, l. [Engl].

(c) Ha az előbbi hierarchia végtelen, akkor adjuk meg az osztály n-edik hatványának jellemzését valamilyen más típusú fatranszformátorral! Ezzel foglalkoznak a [Dau], [EngVog2,4] és a [Vág] dolgozatok.

(d) Adjunk meg olyan egyenlőségeket, tartalmazásokat amelyek fatranszformáció osztályok között fennállnak! Ez a feladat lényegében az (a) és (b) összevonásának és általánosításának tekinthető. Ilyen típusú eredmények találhatók az [ArnDau], [Bak], [Dau], [Engl,2], [EngVogl], [Füll-3], [FülVágl] és [GécSte] munkákban. Példaként itt a [FülVágl]-ben igazolt $DF^2 = NDF \circ LH$ egyenlőséget hozzuk.

A már meglevő egyenlőségekből, tartalmazásokból helyettesítéssel újabbak nyerhetők. Például az (a) pontban említett két egyenlőségből és az $NDL = LNDL \circ NH$-ból, ami ezek közül az elsőnek egy speciális esete, nyerjük, hogy $NDL \circ LH = LNDL \circ NH \circ LH = LNDL \circ H = DL$. Hasonló módon kapjuk a $DF \subset NDF \circ LH$ tartalmazást is a $DF \subset DF^2$ tartalmazásból [Rou] és a $DF^2 = NDF \circ LH$ egyenlőségből [FülVágl].

Felmerül tehát a kérdés, hogy ismerünk-e már valamennyi egyenlőséget és tartalmazást. Pontosabban, ha adott fatranszformáció osztályoknak egy véges P halmaza és adott két

$$(*) \qquad Y_1 \circ \ldots \circ Y_m \quad \text{és} \quad Z_1 \circ \ldots Z_n$$

alakú kifejezés, ahol $Y_i, Z_j \in P$ minden $1 \leq i \leq m$ és $1 \leq j \leq n$ esetén, akkor a már ismert P feletti egyenlőségek és tartalmazások segítségével el tudjuk-e dönteni, hogy az alábbi négy, egymást kizáró és minden lehetséges esetet kimerítő feltétel közül melyik teljesül.

(i) $Y_1 \circ \ldots \circ Y_m = Z_1 \circ \ldots \circ Z_n$

(ii) $Y_1 \circ \ldots \circ Y_m \subset Z_1 \circ \ldots \circ Z_n$

(iii) $Z_1 \circ \ldots \circ Z_n \subset Y_1 \circ \ldots \circ Y_m$

(iv) $Y_1 \circ \ldots \circ Y_m$ és $Z_1 \circ \ldots \circ Z_n$ összehasonlíthatatlanok.

Ha pedig az eddig ismert egyenlőségek és tartalmazások nem elegendőek, akkor megadható-e ezeknek egy olyan rendszere, amellyel a fenti kérdés mindig megválaszolható.

Vegyünk egy pozitív példát, legyen $P = \{DF\}$. Ekkor nyilvánvaló az, hogy a $DF^3 = DF^2$ egyenlőség [FülVág1] és a $DF \subset DF^2$ tartalmazás [Rou] elegendőek ahhoz, hogy tetszőleges $m, n \geq 1$ esetén el tudjuk dönteni, hogy a DF^m és DF^n osztályok között az (i)-(iv) relációk közül melyik áll fenn. (Természetesen nem csak ilyen triviálisan választott P halmazok iránt érdeklődünk majd.)

Ezen értekezés egyik fő eredménye, hogy kifejlesztettünk egy olyan módszert, metaeljárást, amelyet általánosabban P halmazokra is alkalmazhatunk. A módszer lényege abban áll, hogy olyan, a P halmaztól függő objektumokat konstruálunk meg, amelyek ismeretében algoritmikusan eldönthető, hogy tetszőleges két, $(*)$ alakú kifejezés között az (i)-(iv) relációk melyike áll fenn. Ezek az objektumok a következők lesznek

(a) egy P feletti véges T *Thue rendszer*

(b) a T által P^*-on generált $\overset{*}{\underset{T}{\longleftrightarrow}}$ Thue kongruenciának egy N reprezentáns rendszere,

(c) az N által reprezentált fatranszformáció osztályok tartalmazási diagramja,

(d) egy olyan algoritmus amelyik tetszőleges $w \in P^*$ szóhoz megadja a w osztályának a reprezentánsát.

Meg kell azonban jegyeznünk, hogy amennyiben a P halmazt „túl általánosra" választjuk, például, ha P az összes eddig definiált fatranszformáció osztályok halmaza, akkor a feladat olyan áttekinthetetlenné válik, hogy a fenti objektumok megkonstruálása reménytelennek tűnik.

Mindenesetre az eljárás eléggé széles körben alkalmazható. Ennek igazolására, az értekezés második és harmadik fő eredményeként, sikeresen alkalmaztuk az

$$M = \{DL, LDL, NDL, LNDL, H, LH, NH\}$$

és az

$$S = \{DF, LDF, NDF, LNDF, H, LH, NH\}$$

halmazokra, tehát mindkét esetben sikerült megkonstruálni az (a)-(d) objektumokat. Itt említjük meg, hogy az M halmazra vonatkozó eredmények a [Fül3] dolgozatban az S halmazzal kapcsolatos, jóval több munkát igénylő vizsgálatok pedig a [FülVág1,2,5-7] dolgozatokban kerültek publikálásra.

A dolgozat felépítése a következő. A 2. fejezetben a felhasznált fogalmakat és jelöléseket vezetjük be. A 3. fejezet tartalmazza az itt vázolt probléma pontos megfogalmazását és a megoldásra javasolt módszer megadását. A 4. fejezetben ezt a módszert alkalmazzuk az M halmazra, míg az 5. fejezetben az S halmazra.

2. Fogalmak, jelölések

2.1 Halmazok, relációk.

Tetszőleges A és B halmazok esetén $A \subseteq B$-vel jelöljük azt, hogy A ***részhalmaza*** B-nek, $A \subset B$-vel, hogy A ***valódi részhalmaza*** B-nek és $A \not\subseteq B$-vel azt, hogy A ***nem részhalmaza*** B-nek. Az egyelemű halmazt gyakran azonosítani fogjuk egyetlen elemével, tehát $\{a\}$ helyett a-t írunk.

Legyen H egy olyan parciálisan rendezett halmaz, amelynek az elemei halmazok és amelyen a rendezés a halmazok közötti $\subseteq$ tartalmazás. Ekkor H ***tartalmazási diagramján*** a H-nak, mint a $\subseteq$ relációval parciálisan rendezett halmaznak, a ***Hasse diagramját*** (l. [BurSan] 5. old.) értjük.

Ha adott két halmaz A és B, akkor az $A \times B$ direkt szorzat tetszőleges θ részhalmazát ***A-ból B-be való relációnak*** nevezzük. Amennyiben valamely $a \in A$ és $b \in B$ elemekre teljesül az $(a,b) \in \theta$ tartalmazás, akkor ezt a tényt a rövidebb $a\theta b$ írásmóddal jelöljük. A θ ***értelmezési tartományát*** dom(θ)-val, ***értékkészletét*** ran(θ)-val jelöljük és ezeket a szokásos

$$\mathrm{dom}(\theta) = \{a \mid a\theta b \text{ valamely } b \in B \text{ -re}\}$$

és

$$\mathrm{ran}(\theta) = \{b \mid a\theta b \text{ valamely } a \in A \text{ -ra}\}$$

egyenlőségekkel értelmezzük. Azt mondjuk, hogy θ ***totális***, vagy ***teljesen definiált***, ha $\mathrm{dom}(\theta) = A$. Tetszőleges $a \in A$ esetén bevezetjük a

$$\theta(a) = \{b \mid a\theta b\}$$

jelölést és ezt általánosítjuk tetszőleges $L \subseteq A$-ra úgy mint

$$\theta(L) = \bigcup(\theta(a) \mid a \in L).$$

Továbbá, amennyiben Y relációk egy osztálya, H pedig halmazok egy osztálya, akkor

$$Y(H) = \{\theta(L) \mid \theta \in Y, L \in H\},$$

tehát $Y(H)$ elemei is halmazok.

Most bevezetjük a relációk kompozíciójának a fogalmát. Ha θ A-ból B-be való, σ pedig B-ből C-be való reláció, akkor θ-nak és σ-nak a ***kompozícióján*** a

$$\theta \circ \sigma = \{(a,c) \mid a\theta b \text{ és } b\sigma c \text{ valamely } b \in B\text{-re}\},$$

A-ból C-be való relációt értjük. A kompozíció fogalma kiterjeszthető relációkból álló osztályokra is, ha Y és Z ilyen osztályok, akkor

$$Y \circ Z = \{\theta \circ \sigma \mid \theta \in Y \text{ és } \sigma \in Z\}.$$

Továbbá, legyen $Y^1 = Y$ és tetszőleges $n > 1$ esetén legyen $Y^n = Y \circ Y^{n-1}$.

Az A-ból A-ba való relációkat *A feletti relációknak* nevezzük. Az A feletti $\{(a,a) \mid a \in A\}$ *identikus relációt* $Id(A)$-val jelöljük. Amennyiben θ egy A feletti reláció akkor θ-nak az *n-edik hatványát* θ^n-nel jelöljük és a $\theta^0 = Id(A)$, $\theta^n = \theta \circ \theta^{n-1}$, ha $n > 0$ összefüggésekkel definiáljuk. Bevezetjük a $\theta^* = \bigcup(\theta^n \mid n \geq 0)$ jelölést is, tehát θ^* nem más, mint a *θ reflexív, tranzitív lezártja.*

Legyen A tetszőleges halmaz, θ pedig A feletti ekvivalencia reláció. Az $N \subseteq A$ halmazt a *θ reprezentáns rendszerének* mondjuk, ha teljesül, hogy minden $a \in A$-hoz létezik pontosan egy olyan $b \in N$, amelyre $a\theta b$.

Végül bevezetjük a hierarchia fogalmát. Osztályoknak egy $\{C_k \mid k \geq 0\}$ rendszerét *hierarchiának* nevezzük, ha $C_k \subseteq C_{k+1}$, minden $k \geq 0$ esetén. Azt mondjuk, hogy a hierarchia *valódi*, ha minden tartalmazás valódi, tehát $C_k \subset C_{k+1}$, minden $k \geq 0$-ra. Az $\bigcup(C_k \mid k \geq 0)$ osztályt a hierarchia *szuprémumának* nevezzük és röviden $\bigcup C_k$-val jelöljük.

2.2 Szavak, Thue rendszerek és sztring átíró rendszerek.

Ha Σ egy *ábécé*, tehát egy véges, nemüres halmaz, akkor Σ^*-gal jelöljük a Σ elemeiből képezhető összes véges hosszúságú jelsorozatok, *szavak* halmazát. Értelmezhető a szavak szorzata, két w és z Σ^*-beli szó *szorzatát* wz-vel jelöljük. A szorzás egységeleme az *üres szó*, amit mindig λ-val fogunk jelölni. Így Σ^* nem más, mint a Σ által generált szabad monoid. Egy $w \in \Sigma^*$ *szó hosszát* $l(w)$-vel jelöljük és a következőképpen definiáljuk. Ha $w = \lambda$ akkor $l(w) = 0$, különben, ha $w = va$ valamely $v \in \Sigma^*$-ra és $a \in \Sigma$-ra, akkor $l(w) = l(v) + 1$.

A $\Sigma^* \times \Sigma^*$ direkt szorzat tetszőleges véges T részhalmazát Σ feletti *Thue rendszernek* nevezzük. Bevezetjük a Σ^* feletti $\underset{T}{\longleftrightarrow}$ relációt a következőképpen: minden $w, z \in \Sigma^*$ esetén $w \underset{T}{\longleftrightarrow} z$ akkor és csakis akkor, ha van olyan $x, y \in \Sigma^*$ és $(u,v) \in T$, amelyre vagy az teljesül, hogy $w = xuy$ és $z = xvy$, vagy pedig az, hogy $w = xvy$ és $z = xuy$. A $\overset{*}{\underset{T}{\longleftrightarrow}}$ relációt a T által generált *Thue kongruenciának* nevezzük, mint ismeretes, $\overset{*}{\underset{T}{\longleftrightarrow}}$ a T-t tartalmazó legszűkebb kongruencia Σ^*-on.

A $\underset{T}{\longrightarrow}$ relációt Σ^* felett úgy értelmezzük, hogy tetszőleges $w, z \in \Sigma^*$ esetén $w \underset{T}{\longrightarrow} z$ akkor és csakis akkor áll fenn, ha $w \underset{T}{\longleftrightarrow} z$ és $l(w) > l(z)$. Ha valamely $w \in \Sigma^*$-hoz nincsen olyan z, amelyre $w \underset{T}{\longrightarrow} z$, akkor azt mondjuk, hogy *w T-irreducibilis* elem. Az összes T-irreducibilis elemek halmazát $IRR(T)$-vel jelöljük.

Legyen T egy Σ feletti *Thue rendszer.* Azt mondjuk, hogy T

(a) *Church-Rosser,* ha minden $w, z \in \Sigma^*$ esetén $w \overset{*}{\underset{T}{\longleftrightarrow}} z$-ből következik, hogy $w \overset{*}{\underset{T}{\longrightarrow}} x$ és $z \overset{*}{\underset{T}{\longrightarrow}} x$ valamely $x \in \Sigma^*$-ra, és

(b) *konfluens,* ha minden w, z és x Σ^*-beli szavak esetén a $w \overset{*}{\underset{T}{\longrightarrow}} z$ és $w \overset{*}{\underset{T}{\longrightarrow}} x$ relációkból következik, hogy van olyan $y \in \Sigma^*$, amelyre $z \overset{*}{\underset{T}{\longrightarrow}} y$ és $x \overset{*}{\underset{T}{\longrightarrow}} y$.

Könnyen látható, hogy ha egy T *Thue rendszer Church-Rosser,* akkor konfluens is, de a fordítottja általában nem igaz. Például a $T = \{(aa, bb)\}$, $\{a, b\}$ feletti *Thue* *rendszer* konfluens, de nem *Church-Rosser.*

Azt mondjuk, hogy a T és a T' Σ feletti *Thue rendszerek* ekvivalensek, ha $\underset{T}{\overset{*}{\longleftrightarrow}} = \underset{T'}{\overset{*}{\longleftrightarrow}}$.

Tetszőleges Σ ábécé esetén Σ feletti *sztring átíró rendszernek* nevezzük a $\Sigma^* \times \Sigma^*$ tetszőleges véges R részhalmazát. Az R elemeit *átírási szabályoknak* hívjuk, a $\underset{R}{\Longrightarrow}$ (közvetlen) *átírási relációt* pedig az alábbi módon definiáljuk: minden $w, z \in \Sigma^*$ esetén, $w \underset{R}{\Longrightarrow} z$, ha van olyan $x, y \in \Sigma^*$ és $(u, v) \in R$, amelyre $w = xuy$ és $z = xvy$. Tehát, eltérően a *Thue rendszerektől, R* elemei most csak az egyik irányban használhatók átírásra.

Jelöljük a $\underset{R}{\Longrightarrow}$ reláció reflexív, szimmetrikus és tranzitív lezártját $\underset{R}{\overset{*}{\Longleftrightarrow}}$-rel. (Nyilvánvaló, hogy $\underset{R}{\overset{*}{\Longleftrightarrow}}$ az R által generált kongruencia Σ^*-on, továbbá, hogy $\underset{R}{\overset{*}{\Longleftrightarrow}} = \underset{T}{\overset{*}{\longleftrightarrow}}$, ahol $T = R$.)

Azt mondjuk, hogy az R, Σ feletti sztring átíró rendszer

(a) *noetherikus,* ha nem létezik $x_1 \underset{R}{\Longrightarrow} x_2 \underset{R}{\Longrightarrow} \ldots$ alakú végtelen sorozat,

(b) *Church-Rosser,* ha minden $w, z \in \Sigma^*$ esetén a $w \underset{R}{\overset{*}{\Longleftrightarrow}} z$ relációból következik, hogy van olyan $x \in \Sigma^*$, amelyre $w \underset{R}{\overset{*}{\Longrightarrow}} x$ és $z \underset{R}{\overset{*}{\Longrightarrow}} x$,

(c) *konfluens,* ha minden w, z és x Σ^*-beli szó esetén a $w \underset{R}{\overset{*}{\Longrightarrow}} z$ és a $w \underset{R}{\overset{*}{\Longrightarrow}} x$ relációk teljesülése maga után vonja, hogy $z \underset{R}{\overset{*}{\Longrightarrow}} y$ és $x \underset{R}{\overset{*}{\Longrightarrow}} y$ valamely $y \in \Sigma^*$-ra.

Ellentétben a *Thue rendszerekkel,* egy R sztring átíró rendszer akkor és csakis akkor *Church-Rosser,* ha konfluens, l. [Boo2], [Hue], [Jan].

A konfluens és noetherikus sztring átíró rendszert *teljesnek* nevezzük. A Σ^*-beli R-irreducibilis elemek halmazát $IRR(R)$-rel jelöljük, ahol egy $w \in \Sigma^*$ szót R-irreducibilisnek hívunk, ha nincsen olyan $z \in \Sigma^*$, amelyre $w \underset{R}{\Longrightarrow} z$.

Végül megemlítünk egy olyan feltételt, amelyik elegendő ahhoz, hogy R noetherikus legyen. Nevezzük súlyfüggvénynek egy $\rho : \Sigma \to \{1, 2, \ldots\}$ leképezést, ahol $a \in \Sigma$ esetén $\rho(a)$ az a súlya. Egy $w \in \Sigma^*$ szó súlyát a természetes $\rho(w) = 0$, ha $w = \lambda$ és $\rho(w) = \rho(v) + \rho(a)$, ha $w = va, v \in \Sigma^*$ és $a \in \Sigma$ összefüggésekkel definiáljuk. Ha például $\rho(a) = 1$ minden $a \in \Sigma$ esetén, akkor $\rho(w) = l(w)$. Azt mondjuk, hogy R *súlycsökkentő,* ha minden $(u, v) \in R$ esetén $\rho(u) > \rho(v)$. Nyilvánvaló, hogy minden súlycsökkentő sztring átíró rendszer egyben noetherikus is.

2.3 Fák és fatranszformációk.

Rangolt ábécének nevezünk egy olyan Σ véges halmazt, amelynek minden σ eleméhez hozzá van rendelve egy nemnegatív egész szám amit σ *rangjának* nevezünk. Tetszőleges $m \geq 0$ esetén Σ_m-mel jelöljük Σ azon elemeinek a halmazát, amelyeknek a rangja m. Ha σ rangja m, tehát $\sigma \in \Sigma_m$, akkor σ-ra a $\sigma^{(m)}$ jelölést is fogjuk használni. Például $\Sigma = \{a^{(0)}, \sigma^{(2)}\}$ egy olyan rangolt ábécé, amelyben $\Sigma_0 = \{a\}, \Sigma_2 = \{\sigma\}$ és $\Sigma_m = \emptyset$ minden egyéb m-re.

A Σ feletti *termek* vagy *fák* halmazát T_Σ-val jelöljük és értjük alatta a legszűkebb olyan U halmazt, amire teljesül az alábbi két feltétel

(a) $\Sigma_0 \subseteq U$,

(b) $\sigma(t_1, \ldots, t_m) \in U$ minden $m \geq 1$, $\sigma \in \Sigma_m$ és $t_1, \ldots, t_m \in U$ esetén.

Fák esetén gyakran fogjuk használni a *struktúrális indukcióval* (röviden indukcióval) történő bizonyítást, mely a következő elven működik. Legyen P egy T_Σ feletti predikátum. Ha P-re teljesül, hogy

(a) minden $t \in \Sigma_0$ esetén $P(t)$ és

(b) minden $m \geq 0, \sigma \in \Sigma_m$ és $t_1, \ldots, t_m \in T_\Sigma$ esetén abból, hogy $P(t_1), \ldots, P(t_m)$ következik, hogy $P(\sigma(t_1, \ldots, t_m))$,

akkor ebből arra következtetünk, hogy minden $t \in T_\Sigma$-ra $P(t)$.

Szükségünk lesz az alábbi két, fákon értelmezett függvényre. Tetszőleges $t \in T_\Sigma$ esetén $lf(t)$-vel jelöljük *t leveleinek a számát,* $h(t)$-vel pedig a *t magasságát.* A függvények definíciója a következő:

(a) ha $t \in \Sigma_0$, akkor $lf(t) = 1$ és $h(t) = 0$,

(b) ha $t = \sigma(t_1, \ldots, t_m)$ valamely $m \geq 1$, $\sigma \in \Sigma_m$ és $t_1, \ldots, t_m \in T_\Sigma$-ra, akkor $lf(t) = \sum_{i=1}^{m} lf(t_i)$ és $h(t) = 1 + \max\{h(t_i) \mid 1 \leq i \leq m\}$.

A T_Σ tetszőleges L részhalmazát *fanyelvnek* vagy *erdőnek* nevezzük. Ha Δ egy további (nem feltétlenül különböző) rangolt ábécé, akkor a $T_\Sigma \times T_\Delta$ tetszőleges részhalmazát *fatranszformációnak* hívjuk. Tehát egy fatranszformáció nem más, mint egy T_Σ-ból T_Δ-ba való reláció. Az összes $Id(T_\Sigma)$ alakú, tehát totális és identikus fatranszformációk osztályát I-vel jelöljük.

Legyen H tetszőleges halmaz. Akkor $T_\Sigma(H)$-val jelöljük a $T_{\Sigma \cup H}$ halmazt, mely utóbbi esetében H valamennyi elemét 0 rangúnak tekintjük. (Természetesen amennyiben H végtelen, úgy $\Sigma \cup H$ nem rangolt ábécé, de a $\Sigma \cup H$ feletti *termek* $T_{\Sigma \cup H}$ halmaza hasonló módon értelmezhető.)
Most és a továbbiakban X-szel jelöljük változóknak az $\{x_1, x_2, \ldots\}$ halmazát, X_m pedig az $\{x_1, \ldots, x_m\}$ halmazt jelenti, ahol $m \geq 0$. A $T_\Sigma(X_m)$ halmazt röviden $T_{\Sigma,m}$-mel fogjuk jelölni, tehát $T_{\Sigma,m}$ az olyan $\Sigma \cup X_m$ feletti termek halmaza, amelyeknél X_m elemeit 0 aritásunak tekintjük. Bevezetjük a $T_{\Sigma,m}$-nek egy részhalmazát, a $\hat{T}_{\Sigma,m}$ halmazt is. Egy $t \in T_{\Sigma,m}$ fa pontosan akkor van $\hat{T}_{\Sigma,m}$-ben, ha t-ben az X_m minden eleme pontosan egyszer fordul elő, mégpedig az $x_1, \ldots, x_m$ sorrendben. Például, ha $\Sigma = \{a^{(0)}, \sigma^{(2)}\}$, akkor $\sigma(x_1, \sigma(a, x_1) \in T_{\Sigma,1}$, de $\sigma(x_1, \sigma(a, x_1)) \notin \hat{T}_{\Sigma,1}$. Ugyanakkor $\sigma(x_1, \sigma(a, x_2)) \in \hat{T}_{\Sigma,2}$.

Végül bevezetjük a *fa helyettesítés* fogalmát. Legyen $t \in T_{\Sigma,m}$ és legyenek $a_1, \ldots, a_m$ egy tetszőleges H halmaznak az elemei. Ekkor $t(a_1, \ldots, a_m)$-mel jelöljük azt a fát, amelyet úgy kapunk t-ből, hogy benne az x_i változó minden egyes előfordulásának helyére az a_i-t helyettesítjük minden $1 \leq i \leq m$ esetén. Nyilvánvaló, hogy $t(a_1, \ldots, a_m) \in T_\Sigma(H)$. Ugyanakkor megjegyezzük, hogy egy a_i nem feltétlenül fordul elő $t(a_1, \ldots, a_m)$-ben, mivel x_i sem fordul elő feltétlenül t-ben. Ha azonban $t \in \hat{T}_{\Sigma,m}$, akkor $a_1, \ldots, a_m$ mindegyike pontosan egyszer fordul elő $t(a_1, \ldots, a_m)$-ben, mégpedig a fenti sorrendben.

2.4. Fatranszformátorok.

Leszálló fatranszformátornak nevezünk egy $A = (\Sigma, Q, \Delta, R, Q')$ rendszert, ahol

(a) Σ és Δ rangolt ábécék, az *input* és az *output rangolt ábécé,*

(b) Q, az *állapothalmaz,* egy olyan rangolt ábécé, amelyre $Q = Q_1$ és $Q \cap (\Sigma \cup \Delta \cup X) = \emptyset$,

(c) $Q' \subseteq Q$ a *végállapotok halmaza,*

(d) R pedig $\sigma(q_1(x_1), \ldots, q_m(x_m)) \to q(r)$ alakú *átírási szabályok* véges halmaza, ahol $m \geq 0$, $\sigma \in \Sigma_m$, $q, q_1, \ldots, q_m \in Q$ és $r \in T_{\Delta,m}$.

Az előbb specifikált átírási szabályban (röviden szabályban) $\sigma(q_1(x_1), \ldots, q_m(x_m))$-et a szabály *bal oldalának* $q(r)$-et pedig a szabály *jobb oldalának* hívjuk.

Most bevezetjük az *A által indukált fatranszformáció* fogalmát. Definiáljuk először a $T_\Sigma(Q(T_\Delta))$ halmaz felett, ahol $Q(T_\Delta) = \{q(t) \mid q \in Q, t \in T_\Delta\}$, a $\underset{A}{\Longrightarrow}$ relációt a következőképpen: tetszőleges $t, s \in T_\Sigma(Q(T_\Delta))$ esetén $t \underset{A}{\Longrightarrow} s$ akkor és csakis akkor áll fenn, ha R-ben van egy (d) alakú szabály, s-et pedig úgy kaphatjuk t-ből, hogy abban egy $\sigma(q_1(t_1), \ldots, q_m(t_m))$ alakú részfa valamely előfordulásának a helyére a $q(r(t_1, \ldots, t_m))$ fát helyettesítjük, ahol $t_1, \ldots, t_m \in T_\Delta$. Ezután, az A által indukált fatranszformáción a

$$\tau_A = \{(t, s) \in T_\Sigma \times T_\Delta \mid t \overset{*}{\underset{A}{\Longrightarrow}} q(s) \text{ valamely } q \in Q'\text{-re }\}$$

fatranszformációt értjük.

A továbbiakban a leszálló fatranszformátor kifejezés helyett röviden *l transzformátort* fogunk mondani.

Most bevezetjük az l transzformátorok néhány speciális tipusát.

Legyen $A = (\Sigma, Q, \Delta, R, Q')$ tetszőleges l transzformátor. Azt mondjuk, hogy A

(a) *teljesen definiált,* ha minden $m \geq 0$, $\sigma \in \Sigma_m$ és $q_1, \ldots, q_m$ esetén R-ben van legalább egy olyan szabály, amelynek a bal oldala $\sigma(q_1(x_1), \ldots, q_m(x_m))$;

(b) *determinisztikus,* ha R-ben nincs két olyan szabály, amelyeknek a bal oldala ugyanaz;

(c) *lineáris,* ha minden $\sigma(q_1(x_1), \ldots, q_m(x_m)) \to q(r)$ R-beli szabályra teljesül, hogy az $x_1, \ldots, x_m$ változók mindegyike legfeljebb egyszer fordul elő r-ben;

(d) *nemtörlő,* ha minden, a fenti alakú szabályra teljesül, hogy az $x_1, \ldots, x_m$ változók mindegyike legalább egyszer előfordul r-ben;

(e) *lineáris nemtörlő,* ha A mind a lineáris és mind a nemtörlő tulajdonságokkal rendelkezik (tehát minden (c) alakú szabályra az $x_1, \ldots, x_m$ változók mindegyike pontosan egyszer fordul elő r-ben);

(f) *leszálló homomorfizmus,* ha Q egyelemű halmaz, $Q = Q'$, A teljesen definiált és determinisztikus;

(g) *leszálló faautomata,* ha $\Sigma = \Delta$ és R-ben minden szabály $\sigma(q_1(x_1), \ldots, q_m(x_m)) \to q(\sigma(x_1, \ldots, x_m))$ alakú.

A determinisztikus, lineáris és nemtörlő jelzőket rendre a d, l és n betűkkel fogjuk rövidíteni és ezeket alkalmazzuk az l transzformátor kifejezés előtt. Ily módon beszélni fogunk *dl, ldl, ndl, lndl transzformátorokról,* melyek közül például a legutolsó a lineáris, nemtörlő és determinisztikus leszálló fatranszformátor kifejezés rövidítése.

Megemlítjük azt a két nyilvánvaló dolgot, hogy ha A teljesen definiált, akkor τ_A totális fatranszformáció, míg ha A *dl* transzformátor, akkor τ_A egy T_Σ-ból T_Δ-ba való parciális függvény. Továbbá, ha A nemtörlő, akkor minden $(t,s) \in \tau_A$ párra teljesül, hogy $lf(s) \geq lf(t)$.

Ugyancsak nyilvánvaló, hogy ha $A = (\Sigma, Q, \Sigma, R, Q')$ egy leszálló faautomata, akkor τ_A egy T_Σ feletti parciális identitás. Egy $L \subseteq T_\Sigma$ erdőt *felismerhetőnek* nevezünk, ha van olyan A leszálló faautomata, amelyre $L = \mathrm{dom}(\tau_A)$ teljesül. Az összes *felismerhető erdők osztályát* pedig REC-vel jelöljük.

Legyen most x a $dl, ldl, ndl, lndl$ jelzők valamelyike. Egy τ fatranszformációról azt mondjuk, hogy x fatranszformáció, ha van olyan x típusú A transzformátor, amelyre $\tau = \tau_A$. Az összes dl, ldl, ndl, és $lndl$ fatranszformációk osztályát rendre DL, LDL, NDL, és $LNDL$-el jelöljük. Tehát, például, DL jelenti az összes olyan fatranszformációk osztályát, amelyek *dl* transzformátorokkal indukálhatók. Megjegyezzük még, hogy I valódi részosztálya a DL, LDL, NDL és $LNDL$ osztályok mindegyikének.

Most rátérünk a felszálló fatranszformátor fogalmának bevezetésére.

Felszálló fatranszformátornak nevezünk egy $A = (\Sigma, Q, \Delta, R, q_0)$ rendszert, ahol Σ, Q és Δ ugyanazt jelenti mint a leszálló esetben. A q_0 Q-nak egy eleme, amelyet *kezdőállapotnak* hívunk, R pedig most $q(\sigma(x_1, \ldots, x_m)) \to r$ alakú szabályok véges halmaza, ahol $m \geq 0$, $\sigma \in \Sigma_m$, $q \in Q$ és $r \in T_\Delta(Q(X_m))$. (A $Q(X_m)$ a $\{q(x_i) \mid q \in Q,\ 1 \leq i \leq m\}$ halmazt jelöli.)

Az A *által indukált fatranszformációt* a következőképpen értelmezzük. Most a $T_\Delta(Q(T_\Sigma))$ halmazon vezetjük be a $\underset{A}{\Rightarrow}$ relációt: tetszőleges $t, s \in T_\Delta(Q(T_\Sigma))$ esetén $t \underset{A}{\Rightarrow} s$ akkor és csakis akkor áll fenn, ha R-ben van egy $q(\sigma(x_1, \ldots, x_m)) \to r$ alakú szabály, s-et pedig úgy kapjuk t-ből, hogy abban egy $q(\sigma(t_1, \ldots, t_m))$ alakú részfa valamely előfordulásának helyére az $r(t_1, \ldots, t_m)$ fát helyettesítjük, ahol $t_1, \ldots, t_m \in T_\Sigma$. Az A által indukált fatranszformáció pedig a

$$\tau_A = \{(t,s) \in T_\Sigma \times T_\Delta \mid q_0(t) \overset{*}{\underset{A}{\Rightarrow}} s\}$$

reláció.

A rövidség kedvéért a felszálló fatranszformátor kifejezés helyett f transzformátort fogunk mondani.

A továbbiakban az f transzformátorok speciális típusait vezetjük be. Legyen $A = (\Sigma, Q, \Delta, R, q_0)$ egy f transzformátor. Azt mondjuk, hogy A

(a) *teljesen definiált,* ha minden $m \geq 0$, $\sigma \in \Sigma_m$ és $q \in Q$ esetén R-ben van legalább egy $q(\sigma(x_1, \ldots, x_m))$ bal oldalú szabály;

(b) *determinisztikus,* ha R-ben nincs két különböző szabály, amelyeknek a bal oldala ugyanaz;

(c) *lineáris,* ha minden $q(\sigma(x_1, \ldots, x_m)) \to r$ R-beli szabályra teljesül, hogy az $x_1, \ldots, x_m$ változók mindegyike legfeljebb egyszer fordul elő r-ben;

(d) *nemtörlő,* ha (c)-ben a legfeljebb helyett legalábbot mondunk;

(e) *lineáris nemtörlő,* ha lineáris és nemtörlő;

(f) *felszálló homomorfizmus,* ha $Q = \{q_0\}$, A teljesen definiált és determinisztikus;

(g) *uniform,* ha minden $q(\sigma(x_1,\ldots,x_m)) \rightarrow r$, R-beli szabályra és minden $1 \leq i \leq m$-re teljesül, hogy amennyiben $p(x_i)$ és $p'(x_i)$ előfordul r-ben valamely $p,p' \in Q$ esetén, úgy $p = p'$;

(h) *felszálló faautomata,* ha $\Sigma = \Delta$ és R-ben minden szabály $q(\sigma(x_1,\ldots,x_m)) \rightarrow \sigma(q_1(x_1),\ldots,q_m(x_m))$ alakú.

A d, l, n és az ln rövidítéseket az f transzformátorok esetében is fogjuk használni, mégpedig ugyanabban az értelemben mint a leszálló esetben.

Most is érvényes az, hogy minden teljesen definiált f transzformátor totális fatranszfomrációt indukál és hogy a df transzformátorok parciális függvényt indukálnak.

A felszálló faautomaták is alkalmasak erdők felismerésére: egy $L \subseteq T_\Sigma$ erdőről akkor mondjuk, hogy felismerhető felszálló faautomatával, ha van olyan A felszálló faautomata, amelyre $L = \mathrm{dom}(\tau_A)$. Jól ismert az a tény, hogy a felszálló faautomatákkal felismerhető erdők osztálya REC-vel egyezik meg. Ugyanakkor a determinisztikus felszálló faautomatákkal felismerhető erdők $DREC$ osztálya valódi részosztálya REC-nek.

A df, ldf, ndf és $lndf$ fatranszformáció fogalmát a leszálló esettel analóg módon értelmezzük. Az összes fenti típusú fatranszformációk osztályát rendre DF, LDF, NDF és $LNDF$-fel jelöljük. Megjegyezzük, hogy I most is mind a négy osztálynak valódi része.

Befejezésül a homomorfizmus fatranszformátorokat tekintjük. Ugyancsak jól ismert az a tény, hogy minden A leszálló homomorfizmushoz van olyan B felszálló homomorfizmus, amelyre $\tau_A = \tau_B$ és megfordítva. Ezért homomorfizmusok esetén nem teszünk különbséget a leszálló és felszálló esetek között, egyszerűen csak *h transzformátorról* és *h transzformációról* beszélünk. Ezekre szintén alkalmazzuk az l, n és ln jelzőket, tehát például lh transzformátoron lineáris (felszálló vagy leszálló) homomorfizmus transzformátort értünk. Az összes *h, lh, nh és lnh transzformációk* osztályát pedig rendre H, LH, NH és LNH-val jelöljük.

Végül legyen $A = (\Sigma, Q, \Delta, R, q_0)$ egy f transzformátor, p pedig Q-nak egy eleme. Akkor $A(p)$-vel fogjuk jelölni a $(\Sigma, Q, \Delta, R, p)$ f transzformátort.

3. A probléma megfogalmazása és megoldása

A következő problémát vetjük fel. Legyen adott fatranszformáció osztályoknak egy véges P halmaza. Adjunk meg olyan algoritmust, amely tetszőleges $m, n \geq 0$ valamint

$$(*) \qquad Y_1 \circ \ldots \circ Y_m \quad \text{és} \quad Z_1 \circ \ldots \circ Z_n$$

alakú kifejezések esetén (ahol $Y_i, Z_j \in P$ minden $1 \leq i \leq m$-re és $1 \leq j \leq n$-re, az üres kompozíciót pedig I-nek értelmezzük) el tudja dönteni, hogy az alábbi négy, egymást kizáró és minden lehetséges esetet kimerítő reláció közül melyik áll fenn.

(i) $Y_1 \circ \ldots \circ Y_m = Z_1 \circ \ldots \circ Z_n$

(ii) $Y_1 \circ \ldots \circ Y_m \subset Z_1 \circ \ldots \circ Z_n$

(iii) $Z_1 \circ \ldots \circ Z_n \subset Y_1 \circ \ldots \circ Y_m$

(iv) $Y_1 \circ \ldots \circ Y_m$ és $Z_1 \circ \ldots \circ Z_n$ összehasonlíthatatlanok.

Nyilvánvaló, hogy ezen négy eset szétválasztásához szükséges és elegendő, ha az algoritmus azt el tudja dönteni, hogy

(v) $$Y_1 \circ \ldots \circ Y_m \subseteq Z_1 \circ \ldots \circ Z_n$$

fennáll-e.

A fejezet hátralevő részében megadunk egy olyan módszert, amellyel (nem túl általános P halmaz esetén) ez az algoritmus megkonstruálható. Előbb azonban bevezetünk néhány jelölést.

Észrevesszük, hogy a fenti kifejezések lényegében P feletti szavak, ezért tekintjük a P^* szabad monoidot és ugyanakkor a P elemei által a kompozíció műveletével generált

$$[P] = \{Y_1 \circ \ldots \circ Y_m \mid m \geq 0, \quad Y_i \in P \text{ minden } 1 \leq i \leq m\text{-re}\}$$

monoidot is, aminek egységeleme I. A $[P]$ monoid elemei tehát fatranszformáció osztályok, és minden $[P]$-beli fatranszformáció osztály reprezentálható valamely (nem egyértelműen meghatározott) P^*-beli szóval. Ez a

$$|\;| : P^* \to [P]$$

homomorfizmussal írható le, ami nem más, mint a P feletti identitás egyértelmű homomorf kiterjesztése P^*-ból $[P]$-be. Vagyis, a P^*-beli műveletet $\cdot$-al jelölve, egy tetszőleges $Y_1 \cdot \ldots \cdot Y_m$ P^*-beli szó az $|Y_1 \cdot \ldots \cdot Y_m| = Y_1 \circ \ldots \circ Y_m$, $[P]$-beli fatranszformáció osztályt reprezentálja, ahol $|Y_i|$ helyett magát Y_i-t írjuk, $1 \leq i \leq m$. Így $|\lambda| = I$. Két P^*-beli szó, mondjuk $Y_1 \cdot \ldots \cdot Y_m$ és $Z_1 \cdot \ldots \cdot Z_n$ pedig akkor és csakis akkor reprezentálja ugyanazt a $[P]$-beli fatranszformáció osztályt, ha $Y_1 \circ \ldots \circ Y_m = Z_1 \circ \ldots \circ Z_n$, tehát $|Y_1 \cdot \ldots \cdot Y_m| = |Z_1 \cdot \ldots \cdot Z_n|$, vagyis akkor és csakis akkor, ha $Y_1 \cdot \ldots \cdot Y_m \,\theta\, Z_1 \cdot \ldots \cdot Z_n$, ahol θ jelöli a P^*-on a $|\;|$ homomorfizmus által indukált kongruencia relációt.

Ezen jelölésekkel élve módszerünk a következő három feladat megoldásából áll.

(a) Adjuk meg θ-nak egy N reprezentáns rendszerét. (N elemeit normálformáknak fogjuk hívni.)

(b) Adjuk meg a normálformák által reprezentált fatranszformáció osztályoknak — vagyis az $|N| = \{|u| \mid u \in N\}$ halmaznak, ami nyilvánvalóan parciálisan rendezett a tartalmazásra nézve — egy effektív tartalmazási diagramját. (Az effektív szón azt értjük, hogy a diagram segítségével tetszőleges $u, v \in N$ elemekről el tudjuk dönteni, hogy $|u| \subseteq |v|$ fennáll-e.)

(c) Adjuk meg θ-nak egy véges $T \subseteq P^* \times P^*$ generátorrendszerét (tehát egy olyan, P feletti T *Thue rendszert*, amelyre $\xleftrightarrow[T]{*} = \theta$) és adjunk meg egy olyan algoritmust, amely tetszőleges $w \in P^*$-hoz (T-beli egyenletek alkalmazásával) kiszámítja w normálformáját vagyis azt az egyértelműen meghatározott $u \in N$-t, amelyre $w \xleftrightarrow[T]{*} u$.

Most belátjuk, hogy amennyiben ezt a három objektumot sikerült megkonstruálni a P halmazhoz, akkor rendelkezésünkre áll a keresett algoritmus is.

Valóban, tekintsük a $(*)$ kifejezéseket. Első lépésként az $Y_1 \cdot \ldots \cdot Y_m$ és $Z_1 \cdot \ldots \cdot Z_n$ szavakhoz a (c) pontban szereplő algoritmus segítségével konstruáljuk meg azokat az u és v N-beli normálformákat, amelyekre

$$Y_1 \cdot \ldots \cdot Y_m \xleftrightarrow[T]{*} u \quad \text{és} \quad Z_1 \cdot \ldots \cdot Z_n \xleftrightarrow[T]{*} v.$$

Mivel (c) szerint $\xleftrightarrow[T]{*} = \theta$, az is teljesül, hogy

$$Y_1 \circ \ldots \circ Y_m = |u| \quad \text{és} \quad Z_1 \circ \ldots \circ Z_n = |v|.$$

Ezért (v) akkor és csakis akkor áll fenn, ha $|u| \subseteq |v|$, ami (b) szerint $|N|$ tartalmazási diagramjáról eldönthető. A következőt kaptuk.

3.1 ÁLLÍTÁS. *Tegyük fel, hogy a fenti (a)–(c) feladatokat megoldottuk. Akkor tetszőleges két, P feletti $(*)$ alakú kifejezésről eldönthető, hogy közöttük az (v) reláció fennáll-e.*

A fenti módszernek konkrét P halmazokra történő alklamazása azt mutatta, hogy az (a)–(c) pontokban specifikált feladatokat az alábbi lépésekben célszerű végrehajtani.

(1) Megadunk egy véges $T \subseteq P^* \times P^*$ relációt (*Thue rendszert*), amelyről azt sejtjük, hogy θ-t generálja. (Itt nagy mértékben támaszkodhatunk a már ismert dekompozíciós eredményekre.)

(2) Bebizonyítjuk, hogy minden $(u, v) \in T$ esetén $|u| = |v|$. (Mivel T elemeinek P feletti egyenlőségeket kell reprezentálniuk, a szokásos (u, v) írásmód helyett majd az $u \doteq v$-t fogjuk használni.)

Abból, hogy T rendelkezik a fenti tulajdonsággal már nyilvánvalóan következik, hogy $\xleftrightarrow[T]{*} \subseteq \theta$, hiszen $T \subseteq \theta$ és $\xleftrightarrow[T]{*}$ a T-t tartalmazó legszűkebb kongruencia P^*-on. Ezt a tényt az alábbi lemmában deklaráljuk.

3.2 LEMMA. *Tegyük fel, hogy $T \subseteq P^* \times P^*$-ra teljesül, hogy minden $u \doteq v \in T$ esetén $|u| = |v|$. Akkor $\xleftrightarrow[T]{*} \subseteq \theta$.*

(3) Megadjuk P^*-nak egy olyan N részhalmazát, amiről azt sejtjük, hogy θ-nak reprezentáns rendszere. (Megjegyezzük, hogy N-nek szükségképpen rendelkeznie

kell azzal a tulajdonsággal, hogy tetszőleges $u, v \in N$ esetén $|u| = |v|$ akkor és csakis akkor igaz, ha $u = v$; ellenkező esetben N nem lehetne reprezentáns rendszer.)

(4) Megadjuk az N elemei által reprezentált fatranszformáció osztályok $|N| = \{|u| \mid u \in N\}$ halmazának egy effektív tartalmazási diagramját.

(5) Végül megadunk egy olyan algoritmust, amely tetszőleges $w \in P^*$-hoz (T-beli egyenletek alkalmazásával) kiszámítja azt az $u \in N$ normálformát, amelyre $w \overset{*}{\underset{T}{\longleftrightarrow}} u$.

Most belátjuk, hogy a fenti öt pont végrehajtása az (a)–(c) feladatok megoldását is jelenti.

3.3 LEMMA. *Tegyük fel, hogy az (1)–(5) pontokat végrehajtottuk. Akkor az (a)–(c) pontokban kitűzött feldatokat is megoldottuk.*

Bizonyítás. Először azt igazoljuk, hogy $\theta \subseteq \overset{*}{\underset{T}{\longleftrightarrow}}$ is teljesül, tehát, hogy T valóban θ-t generálja. Evégett legyenek $w, w' \in P^*$ olyanok, hogy $w\theta w'$. Konstruáljuk meg az (5)-ben megadott algoritmus segítségével azon $u, u' \in N$ normálformákat, amelyekre $w \overset{*}{\underset{T}{\longleftrightarrow}} u$ és $w' \overset{*}{\underset{T}{\longleftrightarrow}} u'$. Ekkor, a már igazolt $\overset{*}{\underset{T}{\longleftrightarrow}} \subseteq \theta$ tartalmazás miatt, $w\theta u$ és $w'\theta u'$ is érvényes, tehát $u\theta u'$ is, ami viszont (3) értelmében azt jelenti, hogy $u = u'$. Mivel $\overset{*}{\underset{T}{\longleftrightarrow}}$ kongruencia, azt kapjuk, hogy $w \overset{*}{\underset{T}{\longleftrightarrow}} u = u' \overset{*}{\underset{T}{\longleftrightarrow}} w'$, vagyis $w \overset{*}{\underset{T}{\longleftrightarrow}} w'$. Következésképpen $\overset{*}{\underset{T}{\longleftrightarrow}} = \theta$, tehát T θ-t generálja, mint ahogy annak (c)-ben teljesülnie kell. Ebből azonban (5) értelmében az is következik, hogy N valóban reprezentáns rendszere θ-nak, tehát az (a) pontban kitűzött feladatot is megoldottuk. Végül (4) a (b) megoldását jelenti. □

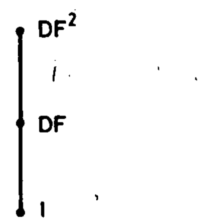
A módszer könnyebb érthetősége végett számoljuk végig az (1)–(5) pontokat a már ismert példán, a $P = \{DF\}$ halmazon.

(1') Legyen $T = \{DF^3 \doteq DF^2\}$. (Természetesen itt a [FülVágl]-ben igazolt $DF^3 = DF^2$ egyenlőség motiválta T választását, ahol DF most a *df* transzformációk osztályát jelöli.)

(2') Nem kell semmit bizonyítani, [FülVágl]-re hivatkozhatunk.

(3') Legyen $N = \{\lambda, DF, DF^2\}$.

(4') Nyilvánvaló, hogy $I \subset DF$, az pedig [Rou]-ban nyert igazolást, hogy $DF \subset DF^2$. Ezért a normálformák által reprezentált fatranszformáció osztályok tartalmazási diagramja az alábbi:



(5') A normálformát kiszámító algoritmus pedig a következő. Legyen $w \in P^*$, tehát $w = DF^m$ valamely $m \geq 0$-ra.

(i) Ha $m = 0$, akkor legyen $u = \lambda$,
(ii) ha $m = 1$, akkor legyen $u = DF$,
(iii) ha $m \geq 2$, akkor legyen $u = DF^2$.
Nyilvánvaló, hogy mindhárom esetben $w \xleftrightarrow[T]{*} u$.

Ezek után a 3.1 állításnak megfelelően könnyedén el tudjuk dönteni azt, hogy tetszőleges két

$$DF^m \text{ és } DF^n$$

osztály között, ahol $m, n \geq 0$, fennáll-e az (v)-nek megfelelő tartalmazás.

A dolgozat következő két fejezetében a fenti (1)–(5) pontokat számoljuk végig a P halmaz két, kevésbé triviális választása esetén.

Befejezésül kitérünk még egy problémára, amelyhez azonban felidézünk két, *Thue rendszerekkel* és sztring átíró rendszerekkel kapcsolatos tételt.

3.4 TÉTEL ([JAN], [Boo2]). *Egy T Thue rendszer (noetherikus R sztring átíró rendszer) akkor és csakis akkor Church–Rosser, ha $\xleftrightarrow[T]{*}$ $\left(\overset{*}{\underset{R}{\Longleftrightarrow}}\right)$ kongruencia minden blokkjában pontosan egy T-irreducibilis (R-irreducibilis) elem van.*

3.5. TÉTEL ([Bool]). *Minden T Thue rendszer (súlycsökkentő R sztring áíró rendszer) esetén megadható olyan algoritmus, amely minden w szóhoz lineáris idő alatt megkonstruál egy olyan u T-irreducibilis (R-irreducibilis) szót, amelyre $w \xrightarrow[T]{*} u$ $\left(w \overset{*}{\underset{R}{\Longrightarrow}} u\right)$.*

Megjegyezzük, hogy amennyiben a 3.5 tételben szereplő T (R) még *Church–Rosser* is akkor, a 3.4 tétel szerint, u a w kongruencia osztályában szereplő egyetlen irreducibilis elem lesz.

Ezért kívánatos lenne, ha az (1) pontban megadott T *Thue rendszer Church–Rosser* tulajdonságú lenne, a normálformák (3)-ban megadott N halmazára pedig az teljesülne, hogy $N = IRR(T)$. Ekkor ugyanis az (5)-ben szereplő algoritmust már nem kellene megadnunk, mivel a fent említett, [Bool]-ben szereplő algoritmus minden $w \in P^*$ szóhoz kiszámítaná w-nek az $u \in N$ normálformáját (mert ez az $N = IRR(T)$ egyenlőség miatt éppen a w osztályában lévő egyetlen T-irreducibilis elem lenne).

Természetesen az (1)-ben megadott T általában nem lesz ilyen tulajdonságú. Mégis elkerülhetjük az (5) algoritmus használatát, ha T-hez keresünk egy olyan R sztring átíró rendszert, amelyik súlycsökkentő, ekvivalens T-vel (tehát $\xleftrightarrow[T]{*} = \overset{*}{\underset{R}{\Longleftrightarrow}}$) és amelyikre $N = IRR(R)$. Ekkor ugyanis R *Church–Rosser* lesz (hiszen $\overset{*}{\underset{R}{\Longleftrightarrow}}$ minden blokkjában pontosan egy R-irreducibilis elem van), a [Bool]-ben szereplő algoritmus pedig minden w-hez megadja az $u \in N$ normálformáját (hiszen $N = IRR(R)$).

A következő fejezetekben mind a leszálló mind a felszálló esetben kapott T-hez sikerült megadni a fenti három tulajdonsággal rendelkező R-t.

4. A determinisztikus leszálló eset

Ebben a fejezetben a 3. fejezet (1)–(5) pontjaiban specifikált feladatokat oldjuk meg az

$$M = \{DL, LDL, NDL, LNDL, H, LH, NH\}$$

halmaz esetében. Használni fogjuk a 3. fejezet jelöléseit, tehát a $|\ | : M^* \to [M]$ homomorfizmust, ahol $[M]$ az

$$[M] = \{Y_1 \circ \ldots \circ Y_m \mid m \geq 0,\ Y_i \in M \text{ minden } 1 \leq i \leq m\text{-re}\}$$

egyenlőséggel definiált monoid; és a $|\ |$ által M^*-on generált θ kongruencia relációt. Előbb azonban felidézzük a leszálló fatranszformátorok kompozíciójának a fogalmát, ami igen hasznos segítséget fog nyújtani ezen feladatok megoldásában.

4.1. Determinisztikus leszálló fatranszformátorok kompozíciói.

A most bevezetésre kerülő fogalmat a [Bak] és [GécSte] munkákból vettük.

Legyenek $A = (\Sigma, P, \Delta, R, P')$ és $B = (\Delta, Q, \Omega, R', Q')$ *dl* transzformátorok. Ezen két *dl* transzformátor kompozícióján értjük azt a $C = (\Sigma, P \times Q, \Omega, R'', P' \times Q')$ *dl* transzformátort melynek R'' szabályhalmaza az alábbi két feltétellel van definiálva:

(a) minden $\sigma \in \Sigma_0$, $p \in P$ és $q \in Q$ esetén a $\sigma \to \langle p, q\rangle\ (r)$ szabály akkor és csakis akkor van R''-ben, ha az R-ben van olyan $\sigma \to p\ (r')$ szabály, amelyre teljesül hogy $r' \overset{*}{\underset{B}{\Rightarrow}} q\ (r)$;

(b) minden $m \geq 1$, $\sigma \in \Sigma_m$, $p, p_1, \ldots, p_m \in P$ és $q, q_1, \ldots, q_m \in Q$ esetén a $\sigma(\langle p_1, q_1\rangle(x_1), \ldots, \langle p_m, q_m\rangle(x_m)) \to \langle p, q\rangle\ (r)$ szabály akkor és csakis akkor van R''-ben, ha R-ben van olyan $\sigma(p_1(x_1), \ldots, p_m(x_m)) \to p\ (r')$ alakú szabály, amelyre teljesül, hogy $r'(q_1(x_1), \ldots, q_m(x_m)) \overset{*}{\underset{B}{\Rightarrow}} q\ (r)$.

Könnyen igazolható, l. [Bak], hogy ezen *dl* transzformátorok által indukált fatranszformációkra teljesül, hogy $\tau_C = \tau_A \circ \tau_B$. Ebből a tényből következik az, hogy $DL \circ DL \subseteq DL$ vagyis, hogy DL zárt a kompozícióra. (Megjegyezzük, hogy a $\tau_C = \tau_A \circ \tau_B$ egyenlőség általában nem igaz, ha A és B nemdeterminisztikus *l* transzformátorok.)

Ezt a kompozíciós technikát fel lehet használni más, $Y \circ Z \subseteq V$ alakú tartalmazások megmutatására is, ahol Y, Z és V az M halmaznak tetszőleges elemei. Valóban, vesszük az A és B transzformátorokat, amelyek rendre y és z típusúak és megkonstruáljuk A-nak és B-nek a C kompozícióját. Ha ez a C v típusú, akkor ez elegendő ahhoz, hogy $Y \circ Z \subseteq V$ teljesüljön. Például így igazolható, hogy $LDL \circ LDL \subseteq LDL$.

Egy másik észrevétel az, hogy ha már igazoltunk egy $Y \circ Z \subseteq Z$ alakú tartalmazást, ahol Y és Z ugyancsak az M halmaz tetszőleges elemeit jelentik, akkor ebből következik, hogy $Y \circ Z = Z$ is teljesül. Ez azért van, mert $I \subseteq Y$ és mert minden Z-beli fatranszformáció felírható mint egy alkalmas I-beli identitásnak és önmagának a kompozíciója. Tehát a fentiekből az is adódik, hogy $DL \circ DL = DL$ és $LDL \circ LDL = LDL$.

4.2 A T Thue rendszer megadása és a $\overset{*}{\underset{T}{\longleftrightarrow}} \subseteq \theta$ *tartalmazás..*

Azt állítjuk, hogy az alábbi véges, M feletti T reláció (tehát *Thue rendszer*) éppen θ-t generálja. A T elemeit a szokásos (u, v) írásmód helyett most is $u \doteq v$ alakban írjuk fel. Álljon tehát T az alábbi 13 elemből.

(1)	$NH \cdot LH \doteq H$	(8)	$LNDL \cdot H \doteq DL$
(2)	$LH \cdot NH \doteq H$	(9)	$NH \cdot NDL \doteq NDL$
(3)	$NH \cdot NH \doteq NH$	(10)	$LH \cdot LDL \doteq LDL$
(4)	$LH \cdot LH \doteq LH$	(11)	$NH \cdot LDL \doteq DL$
(5)	$LNDL \cdot LNDL \doteq LNDL$	(12)	$LDL \cdot LNDL \doteq LDL$
(6)	$LNDL \cdot LH \doteq LDL$	(13)	$NDL \cdot LNDL \doteq NDL$
(7)	$LNDL \cdot NH \doteq NDL$		

Először azt igazoljuk, hogy T elemei valóban $[M]$-beli egyenlőségeket reprezentálnak, röviden, hogy T minden eleme érvényes $[M]$-ben. Ez például a (8) esetében azt jelenti, hogy $LNDL \circ H = DL$ is teljesül.

4.2.1 TÉTEL. *Minden $u \doteq v \in T$ esetén $|u| = |v|$ is teljesül.*

Bizonyítás. Nem kell valamennyi esetet igazolnunk, mivel a T elemeit tulajdonképpen már ismert vagy az ismertekre könnyen visszavezethető dekompozíciókból gyűjtöttük össze. Továbbá az (1) és (2) érvényességét majd az 5.1.1 tételben fogjuk igazolni, mivel ott ez egyszerűbben megtehető. Kezdjük azzal, hogy (8) érvényessége következik az [Engl]-ben található 4.1 lemmából, de a [GécSte] 158. oldalán levő 3.3 tételből is, a (6) és a (7) pedig (8)-nak speciális eseteiként adódnak. Továbbá, (3), (4), (5), (9), (10), (12) és (13) érvényessége könnyen bizonyítható a 4.1 pontban ismertetett, fatranszformátorok kompozícióira vonatkozó technikával. Egyedül (11) érvényességét kell igazolnunk, vagyis azt a tényt, hogy $NH \circ LDL = DL$.

Észrevesszük, hogy az $NH \circ LDL \subseteq DL$ tartalmazás ugyancsak a 4.1-ben leírt módon bizonyítható.

A fordított tartalmazás igazolása végett vegyünk egy tetszőleges $A = (\Sigma, Q, \Delta, R, Q')$ *dl* transzformátort. Jelöljük rhs(R)-rel az R-beli szabályok jobb oldalainak a halmazát és minden $t \in \text{rhs}(R)$ és $i \geq 1$ esetén legyen $t(i)$ az x_i változó előfordulásainak a száma T-ben. Legyen továbbá $n = \max\{t(i) \mid t \in \text{rhs}(R),\ i \geq 1\}$. Meg fogunk adni egy B *h* transzformátort és egy C *ldl* transzformátort amelyekre $\tau_A = \tau_B \circ \tau_C$.

Legyen $B = (\Sigma, p, \Sigma', R', p)$ ahol Σ' a legszűkebb olyan rangolt ábécé amelyre $\Sigma'_{n \cdot m} = \{\sigma' \mid \sigma \in \Sigma_m\}$ minden $m \geq 0$ esetén és ahol R' az összes

$$\sigma(p(x_1), \ldots, p(x_m)) \to p(\sigma'(x_1, \ldots, x_1, \ldots, x_m, \ldots, x_m))$$

alakú szabályokból áll úgy, hogy $m \geq 0$, $\sigma \in \Sigma_m$ és az $x_1, \ldots, x_m$ változók mindegyike n-szer fordul elő a jobboldalon.

Továbbá vezessük be a $C = (\Sigma', Q, \Delta, R'', Q')$ *ldl* transzformátort, ahol egy

$$\sigma'(q_1(x_1), \dots, q_1(x_n), \dots, q_m(x_{n\cdot(m-1)+1}), \dots, q_m(x_{n\cdot m})) \to$$
$$q(r(x_1, \dots, x_{n_1}, \dots, x_{n\cdot(m-1)+1}, \dots, x_{n\cdot(m-1)+n_m}))$$

szabály akkor és csakis akkor van R''-ben, ha a

$$\sigma(q_1(x_1), \dots, q_m(x_m)) \to q(r(x_1, \dots, x_1, \dots, x_m, \dots, x_m))$$

szabály R-ben, ahol $r \in \hat{T}_{\Delta,k}$, $k = n_1 + \dots + n_m$ és az x_i változó n_i-szer fordul elő ezen szabály jobboldalán minden $1 \le i \le m$-re.

Ekkor, t szerinti struktúrális indukcióval, a következő ekvivalencia igazolható könnyen: minden $t \in T_\Sigma, s \in T_\Delta$ és $q \in Q$ esetén $t \overset{*}{\underset{A}{\Longrightarrow}} q(s)$ akkor és csakis akkor, ha van olyan $s' \in T_{\Sigma'}$ amelyre $t \overset{*}{\underset{B}{\Longrightarrow}} p(s')$ és $s' \overset{*}{\underset{C}{\Longrightarrow}} q(s)$. Ezen ekvivalenciából q-t Q'-belinek választva azt kapjuk, hogy $\tau_A = \tau_B \circ \tau_C$, amivel egyúttal a $DL \subseteq NH \circ LDL$ tartalmazást is igazoltuk. Tehát $NH \circ LDL = DL$ és így (11) is érvényes $[M]$-ben. □

4.2.2 KÖVETKEZMÉNY. $\overset{*}{\underset{T}{\longleftrightarrow}} \subseteq \theta$.

Bizonyítás. L. 3.2 lemma. □

4.3 Reprezentáns rendszer és tartalmazási diagram.

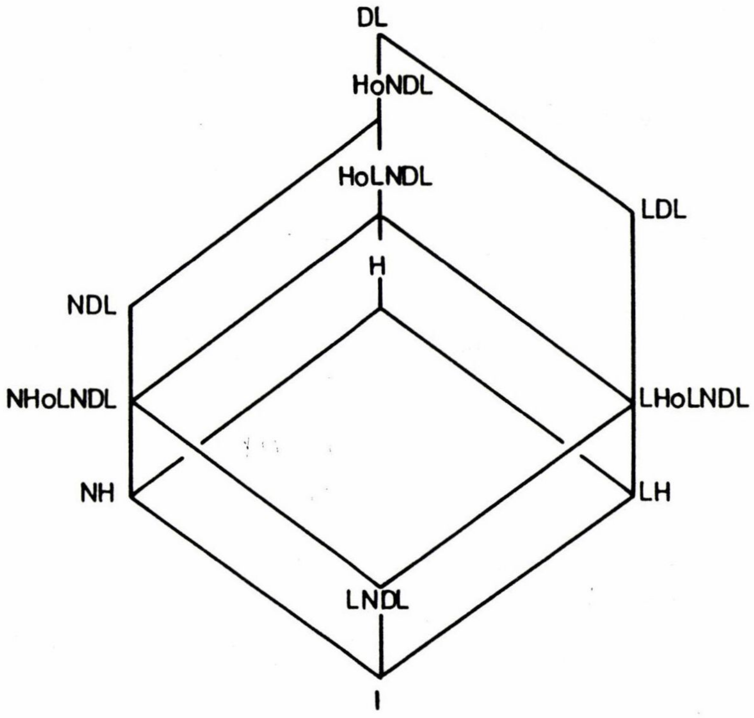
Most megadjuk M^*-nak egy véges N részhalmazát, amiről később belátjuk, hogy θ-nak egy reprezentáns rendszere. Az N halmazt a következőképpen definiáljuk:

$$N = M \cup \{\lambda,\ H \cdot NDL,\ H \cdot LNDL,\ NH \cdot LNDL,\ LH \cdot LNDL\}.$$

Először bebizonyítjuk, hogy N különböző elemeinek a $|\ |$ melletti képei is különbözőek vagyis azt, hogy minden $w, z \in N$ esetén ha $w \neq z$ akkor $|w| \neq |z|$. Ezt úgy érjük el, hogy megadjuk az $|N| = \{|u| \mid u \in N\}$ halmaznak a tartalmazási diagramját, amiről előbbi állításunk helyessége leolvasható. Az 1. ábra diagramjáról fogjuk igazolni, hogy $|N|$ tartalmazási diagramja.

4.3.1 TÉTEL. *Az 1. ábrán látható diagram az $|N|$ tartalmazási diagramja.*

Bizonyítás. Először is azt vesszük észre, hogy ha egy Y osztály a Z osztály fölött van, akkor $Z \subseteq Y$. Ez könnyedén következik az M halmaz elemei között fennálló alapvető tartalmazási relációkból és abból a tényből, hogy minden $Y_1, Y_2 \in M$ esetén az $Y_1, Y_2 \subseteq DL$; $Y_1, Y_2 \subseteq LDL$ és $Y_1, Y_2 \subseteq NDL$ tartalmazásokból rendre következik, hogy $Y_1 \circ Y_2 \subseteq DL$; $Y_1 \circ Y_2 \subseteq LDL$ és $Y_1 \circ Y_2 \subseteq NDL$. Ez könnyedén látható a 4.1 pontban leírtak értelmében. Ezért azt kell még belátni, hogy a diagram által mutatott valamennyi tartalmazás valódi, és hogy a relációba nem állított osztályok pedig összehasonlíthatatlanok. Ehhez szükségünk lesz a következő két lemmára.



1. ábra

4.3.2 LEMMA. $LDL \not\subseteq H \circ NDL$.

Bizonyítás. Megadunk egy olyan *ldl* fatranszformációt, amelyik nincs $H \circ NDL$ -ben. Evégett vezessük be a $\Sigma = \{a^{(0)}, \sigma^{(2)}\}$ rangolt ábécét és jelöljük T_Σ^o-val (T_Σ^e-vel) T_Σ-nak azt a részhalmazát amelyik az összes páratlan (páros) számú a-t tartalmazó fából áll. Megjegyezzük, hogy $a \in T_\Sigma^o$. Továbbá, definiáljuk az odd : $T_\Sigma \to T_\Sigma$ függvényt a következőképpen:

(a) $t = a$ esetén legyen $\mathrm{odd}(t) = t$,

(b) $t = \sigma(t_1, t_2)$ esetén, ahol $t_1, t_2 \in T_\Sigma$, legyen

(i) $\mathrm{odd}(t) = a$, ha $t_1, t_2 \in T_\Sigma^o$ vagy $t_1, t_2 \in T_\Sigma^e$,

(ii) $\mathrm{odd}(t) = \sigma(a, \mathrm{odd}(t_2))$, ha $t_1 \in T_\Sigma^e$ és $t_2 \in T_\Sigma^o$,

(iii) $\mathrm{odd}(t) = \sigma(\mathrm{odd}(t_1), a))$, ha $t_1 \in T_\Sigma^o$ és $t_2 \in T_\Sigma^e$.

Informálisan szólva, odd azt a transzfomrációt valósítja meg, amely minden $t \in T_\Sigma$ és $t \neq a$ esetén a t-nek minden olyan t' részfáját a-val helyettesíti, amely eleget tesz az alábbi feltételeknek:

(a) t' páros számú a-t tartalmaz,

(b) t' nem valódi részfája egy másik olyan részfának, ami páros számú a-t tartalmaz.

Könnyű azt igazolni, hogy odd indukálható *ldl* transzformátorral. Valóban, tekintsük azt az A *ldl* transzformátort amelynek szabályai a következők:

(a) $a \to p_o(a)$,

(b) $\sigma(p_e(x_1), p_e(x_2)) \to p_e(a)$,

(c) $\sigma(p_o(x_1), p_o(x_2)) \to p_e(a)$,

(d) $\sigma(p_e(x_1), p_o(x_2)) \to p_o(\sigma(a, x_2))$,

(e) $\sigma(p_o(x_1), p_e(x_2)) \to p_o(\sigma(x_1, a))$.

Nyilvánvaló, hogy ha mind p_o-t és p_e-t végállapotnak veszzük, akkor $\tau_A =$ odd, tehát odd $\in LDL$.

Megmutatjuk, hogy odd $\notin H \circ NDL$. Ez azért lesz így, mert a h transzformátor nem tudja felismerni, hogy az input fa egy részfája páratlan vagy páros számú a-t tartalmaz-e, az *ndl* transzformátor pedig, habár képes erre a felismerésre, nem tud törölni.

Tételezzük fel, hogy odd $\in H \circ NDL$ vagyis azt, hogy odd $= \tau_B \circ \tau_C$ teljesül valamilyen $B = (\Sigma, p, \Delta, R, p)$ *h* transzformátorra és $C = (\Delta, Q, \Sigma, R', Q')$ *ndl* transzformátorra. Ekkor B nem lehet nemtörlő. Valóban, ha B nemtörlő lenne, akkor a 4.1 pontban mondottak miatt $\tau_B \circ \tau_C \in NDL$ is igaz lenne. De ekkor azt kapnánk, hogy $lf(s) \geq lf(t)$ minden $(t, s) \in \tau_B \circ \tau_C$ esetén, ami pedig ellentmondana annak, hogy odd $= \tau_B \circ \tau_C$.

Tehát B törlő. Megmutatjuk, hogy ez is ellentmondáshoz vezet. Vegyük B-nek a σ-t átíró $\sigma(p(x_1), p(x_2)) \to p(r)$ alakú szabályát, amelyre nézve három eset lehetséges, nevezetesen

(a) x_1 előfordul r-ben de x_2 nem,

(b) x_2 előfordul r-ben de x_1 nem,

(c) sem x_1 sem x_2 nem fordul elő r-ben.

Tegyük fel, hogy az első eset áll fenn és vegyünk egy teszőleges $t \in T_\Sigma^o$ fát, de úgy, hogy $t \neq a$. Legyen t' az a fa, amelyet úgy kapunk t-ből, hogy benne az a legjobboldalibb előfordulását a $\sigma(a, a)$ fával helyettesítjük. Ekkor nyilvánvaló, hogy $t' \in T_\Sigma^e$ és hogy $\tau_B(t) = \tau_B(t')$, mivel B alkalmazása invariáns lesz erre a helyettesítésre. Ezért $\tau_B \circ \tau_C(t) = \tau_B \circ \tau_C(t')$, ami ellentmondás, mert odd$(t') = a$, de odd$(t) \neq a$. Hasonló módon igazolható, hogy a második eset sem állhat fenn, míg teljesen nyilvánvaló az, hogy a harmadik is lehetetlen. Ezért odd $\notin H \circ NDL$, ami a bizonyítás végét jelenti. □

4.3.3 Lemma. $NDL \not\subseteq H \circ LNDL$.

Bizonyítás. Tekintsük a $\Sigma = \{a^{(0)}, \sigma^{(1)}\}$ és a $\Delta = \{a^{(0)}, \sigma^{(1)}, \gamma^{(2)}\}$ rangolt ábécéket, és vezessük be az $A = (\Sigma, Q, \Delta, R, Q')$ *ndl* transzformátort, ahol $Q = Q' = \{p, q\}$ és R az alábbi három szabályból áll:

(a) $a \to p(a)$,

(b) $\sigma(p(x_1)) \to q(\gamma(x_1, x_1))$ és $\sigma(q(x_1)) \to p(\sigma(x_1))$.

Nem specifikáljuk az A által indukált fatranszformációt, csupán megállapítjuk azt, hogy $\tau_A(T_\Sigma)$ nem felismerhető erdő, valamint, hogy minden $m \geq 0$-ra, ha $(\sigma^m(a), t) \in \tau_A$, akkor $lf(t) = 2^n$, ahol $n = \lfloor (m+1)/2 \rfloor$.

Meg fogjuk mutatni, hogy $\tau_A \notin H \circ LNDL$. Indirekt módon tegyük fel, hogy $\tau_A = \tau_B \circ \tau_C$, ahol $B = (\Sigma, u, \Omega, R, u)$ h transzformátor és $C = (\Omega, Q, \Delta, R', Q')$ pedig *lndl* transzformátor. Először is észrevesszük, hogy B nem lehet lineáris. Valóban, ha B lineáris lenne, akkor a $\tau_B(T_\Sigma)$ erdő felismerhető lenne és ezért a $\tau_B \circ \tau_C(T_\Sigma)$ is; a bizonyítást lásd a [GécSte] könyv 174–175 oldalain. Ez pedig ellentmondana annak, hogy $\tau_A(T_\Sigma)$ nem felismerhető.

Ezért B-nek a σ-t átíró $\sigma(u(x_1)) \to u(r)$ szabályára annak kell teljesülnie, hogy x_1 $k > 1$-szer fordul elő r-ben. Ekkor azonban, ha $(\sigma^m(a), t') \in \tau_B$ valamilyen $m \geq 1$-re, akkor $lf(t') \geq k^m$. Továbbá, mivel C nemtörlő, minden $(t', t) \in \tau_C$ esetén $lf(t) \geq lf(t')$. Ezért minden $m \geq 1$ és $(\sigma^m(a), t) \in \tau_B \circ \tau_C$-re azt kapjuk, hogy $lf(t) \geq k^m$, ahol $k > 1$. Ellentmondás, mivel $\tau_A = \tau_B \circ \tau_C$ és $lf(t) = 2^n$ minden $(\sigma^m(a), t) \in \tau_A$-ra, ahol $n = \lfloor (m+1)/2 \rfloor$. □

Most már be tudjuk fejezni a 4.3.1 tétel bizonyítását is. Mint láttuk, $LDL \not\subseteq H \circ NDL$; továbbá $H \circ LNDL \not\subseteq LDL$, mivel a bal oldal tartalmaz nem lineáris *dl* transzformátorok által indukált fatranszformációkat is. Következésképpen $LDL \not\subseteq H \circ LNDL$ és $H \circ NDL \not\subseteq LDL$. Így kapjuk, hogy $H \circ NDL \subset DL$, $LDL \subset DL$ (ami egyébként is nyilvánvaló), $LH \circ LNDL \subset LDL$ és $LH \circ LNDL \subset H \circ LNDL$.

Hasonlóan láttuk, hogy $NDL \not\subseteq H \circ LNDL$, míg $H \circ LNDL \not\subseteq NDL$ nyilvánvaló. Ezért $H \circ LNDL \subset H \circ NDL$, $NDL \subset H \circ NDL$, $NH \circ LNDL \subset NDL$ és $NH \circ LNDL \subset H \circ LNDL$.

Nem jelent problémát annak a ténynek az igazolása, hogy az $NH \circ LNDL$, $LH \circ LNDL$ és a H osztályok páronként összehasonlíthatatlanok. Ebből kapjuk, hogy $H \subset H \circ LNDL$, $NH \subset H$, $LH \subset H$, $LNDL \subset NH \circ LNDL$ és $LNDL \subset LH \circ LNDL$.

Ugyancsak triviálisan teljesülnek az $I \subset NH$, $I \subset LH$ és az $I \subset LNDL$ tartalmazások.

Végül megjegyezzük, hogy az $NDL - NH \circ LNDL - NH$ vonal bármelyik osztálya összehasonlíthatatlan az $LDL - LH \circ LNDL - LH$ vonal bármelyik osztályával. Valóban, az előbbiek tartalmaznak nem lineáris fatranszformációkat, az utóbbiak pedig törlő fatranszformációkat. Ezzel befejeztük a 4.3.1 tétel bizonyítását. □

4.3.4 KÖVETKEZMÉNY. *Tetszőleges $u, u' \in N$ esetén $u = u'$ akkor és csakis akkor, ha $|u| = |u'|$.*

Bizonyítás. Pontosan ez olvasható le az 1. ábráról. □

4.4 Algoritmus a normálforma megadására.

Most igazoljuk, hogy tetszőleges $w \in M^*$-hoz van olyan $u \in N$ normálforma, amelyre $w \overset{*}{\underset{T}{\leftrightarrow}} u$; sőt meg is adjuk azt az algoritmust, ami ezen u-t szolgáltatja.

4.4.1 TÉTEL. *Tetszőleges $w \in M^*$ esetén effektíven megadható olyan $u \in N$, amelyre teljesül, hogy $w \overset{*}{\underset{T}{\leftrightarrow}} u$.*

Bizonyítás. Először megadunk további 6 olyan formális egyenletet, amelyek mindegyike levezethető T-ből.

$$
\begin{array}{llll}
(14) & H \cdot NH \doteq H & (17) & LH \cdot H \doteq H \\
(15) & NH \cdot H \doteq H & (18) & LH \cdot NH \doteq NH \cdot LH \\
(16) & H \cdot LH \doteq H & (19) & H \cdot H \doteq H
\end{array}
$$

Tehát a fenti $u \doteq v$ alakú egyenletek mindegyikére teljesül, hogy $u \overset{*}{\underset{T}{\longleftrightarrow}} v$, ami könnyen ellenőrizhető. Például (14)-et a $H \cdot NH \doteq LH \cdot NH \cdot NH \doteq LH \cdot NH \doteq H$ levezetéssel kapjuk rendre a (2), (3) és (2) egyenletek alkalmazásával.

A keresett algoritmust a 2. ábrán látható táblázat (ún. *Cayley táblázat*) segítségével fogjuk megadni, ami a következő információkat tartalmazza. Először is vegyük észre, hogy minden $u \in N$-hez tartozik a táblázatnak egy sora és minden $Y \in M$-hez egy oszlopa.
Az u-hoz tartozó sor és az Y-hoz tartozó oszlop találkozásánál lévő téglalapban pedig N-nek az a v eleme van amire $u \cdot Y \overset{*}{\underset{T}{\longleftrightarrow}} v$. Továbbá, ha $u \cdot Y = v$ is igaz, akkor semmi más nincs a téglalapban. Ha pedig $u \cdot Y \neq v$, akkor azon T-beli és T-ből levezetett formális egyenletek sorszáma is ott található (beleértve a (14)–(19) egyenleteket is), amelyekkel a feltüntetett sorrendben igazolható, hogy $u \cdot Y \overset{*}{\underset{T}{\longleftrightarrow}} v$. Így a táblázatról elmondottak helyessége utánaszámolással ellenőrizhető.
Például, az $u = NH \cdot LNDL$ és $Y = LH$ esetben DL az a normálforma, amelyikre $NH \cdot LNDL \cdot LH \overset{*}{\underset{T}{\longleftrightarrow}} DL$. Még pontosabban, a (6) egyenlettel kapjuk, hogy $NH \cdot LNDL \cdot LH \underset{T}{\longleftrightarrow} NH \cdot LDL$, a (11)-gyel pedig, hogy $NH \cdot LDL \underset{T}{\longleftrightarrow} DL$. (Annak a megállapítását azonban az olvasóra bízzuk, hogy egy feltüntetett egyenletet melyik rész sztringre alkalmazzuk, és hogy milyen irányítással használjuk az egyenletet.)

Végül rátérünk magának a tételnek az igazolására. Megmutatjuk a w hossza szerinti indukcióval, hogy mindig meg tudjuk konstruálni azt az $u \in N$-t, amire $w \overset{*}{\underset{T}{\longleftrightarrow}} u$.

Ha $l(w) = 0$, tehát $w = \lambda$, akkor legyen $u = \lambda$.

Ha $l(w) > 0$, vagyis $w = x \cdot Y$ valamely $x \in M^*$ és $Y \in M$ esetén, akkor pedig járjunk el a következőképpen. Konstruáljuk meg azt az (indukció feltevés szerint létező) $u \in N$-t, amelyre $x \overset{*}{\underset{T}{\longleftrightarrow}} u$, majd a táblázat szerint konstruáljuk meg azt a $v \in N$-t, amelyre $u \cdot Y \overset{*}{\underset{T}{\longleftrightarrow}} v$. Így kapjuk, hogy $w = x \cdot Y \overset{*}{\underset{T}{\longleftrightarrow}} u \cdot Y \overset{*}{\underset{T}{\longleftrightarrow}} v$ és v effektíven megkonstruálható. $\square$

Ezzel a 3. fejezet (1)–(5) pontjaiban kitűzött feladatokat megoldottuk az M halmaz esetében. Így, a 3.3 lemma értelmében az (a)–(c) feladatokat is, tehát alkalmazhatjuk a 3.1 állítást.

Befejezésül a 3.1 állításban szereplő algoritmus alkalmazására hozunk egy példát. Legyen a két, M feletti kifejezés

$$LNDL \circ NH \circ LDL \quad \text{és} \quad NH \circ LNDL \circ NDL.$$

	DL	NDL	LDL	$LNDL$
DL	DL 8, 11, 12, 11, 8, 19, 8	DL 7, 11, 12, 11 8, 14, 8	DL 8, 1, 10, 11 8, 5, 8	DL 11, 12, 11
NDL	DL 8, 13, 7, 15, 8	NDL 7, 9, 7, 5 7	DL 7, 11, 8, 5 8	NDL 13
LDL	DL 8, 12, 6, 17, 8	DL 7, 12 6, 2, 8	LDL 6, 12 6, 4, 6	LDL 12
$LNDL$	DL 8, 5, 8	NDL 7, 5, 7	LDL 6, 5, 6	$LNDL$ 5
H	DL 11, 4, 1 10, 11	$H \cdot NDL$	DL 1, 10, 11	$H \cdot LNDL$
NH	DL 11, 3, 11	NDL 9	DL 11	$NH \cdot LNDL$
LH	DL 11, 2, 1 10, 11	$H \cdot NDL$ 9, 2	LDL 10	$LH \cdot LNDL$
$H \cdot NDL$	DL 8, 13, 7, 15, 8 11, 14, 1, 10, 11	$H \cdot NDL$ 7, 9 7, 5, 7	DL 7, 11, 8, 5, 8 11, 14, 1, 10, 11	$H \cdot NDL$ 13
$H \cdot LNDL$	DL 8, 5, 8, 11, 14, 1, 10, 11	$H \cdot NDL$ 7, 5, 7	DL 6, 5, 6 1, 10, 11	$H \cdot LNDL$ 5
$NH \cdot LNDL$	DL 8, 5, 8 11, 3, 11	NDL 7, 5, 7 9	DL 6, 5, 6 11	$NH \cdot LNDL$ 5
$LH \cdot LNDL$	DL 8, 5, 8, 11, 2, 1, 10, 11	$H \cdot NDL$ 7, 5, 7 9, 2	LDL 6, 5, 6, 10	$LH \cdot LNDL$ 5

2/a. ábra

A 2. ábrán lévő táblázatnak a 4.4.1 tételben leírt módon történő alkalmazásával kapjuk, hogy $LNDL \cdot NH \underset{T}{\longleftrightarrow} NDL$ és $NDL \cdot LDL \overset{*}{\underset{T}{\longleftrightarrow}} DL$, vagyis, hogy $LNDL \cdot$

	H	NH	LH
DL	DL $8,19,8$	DL $8,14,8$	DL $8,16,8$
NDL	DL $7,15,8$	NDL $7,3,7$	DL $7,1,8$
LDL	DL $6,17,8$	DL $6,2,8$	LDL $6,4,6$
$LNDL$	DL 8	NDL 7	LDL 6
H	H 19	H 14	H 16
NH	H 15	NH 3	H 1
LH	H 17	H 2	LH 4
$H \cdot NDL$	DL $7,15,8$ $11,14,1,10,11$	$H \cdot NDL$ $7,3,7$	DL $7,1,8$ $11,14,1,10,11$
$H \cdot LNDL$	DL $8,11,14$ $1,10,11$	$H \cdot NDL$ 7	DL $6,1$ $10,11$
$NH \cdot LNDL$	DL $8,11$ $3,11$	NDL $7,9$	DL $6,11$
$LH \cdot LNDL$	DL $8,11,2$ $1,10,11$	$H \cdot NDL$ $7,9,2$	LDL $6,10$

2/b. ábra

$NH \cdot LDL \overset{*}{\underset{T}{\longleftrightarrow}} DL$. Hasonló módon $NH \cdot LNDL \cdot NDL \overset{*}{\underset{T}{\longleftrightarrow}} NDL$ is adódik. Másrészt, az 1. ábrán lévő tartalmazási diagram értelmében $NDL \subset DL$, tehát a keresett reláció

$$NH \circ LNDL \circ NDL \subset LNDL \circ NH \circ LDL.$$

4.5 A T-vel ekvivalens teljes R sztring átíró rendszer megadása.

Mindenekelőtt megjegyezzük, hogy T nem *Church–Rosser tulajdonságú.* Például, könnyen igazolható, hogy $NDL \cdot NH \xleftrightarrow[T]{*} LNDL \cdot NH$, de nincsen olyan $z \in M^*$, amelyre $NDL \cdot NH \xrightarrow[T]{*} z$ és $LNDL \cdot NH \xrightarrow[T]{*} z$.

Tekintsük a 2. ábrán látható táblázatot és vezessük be a következő R sztring átíró rendszert. A táblázat minden u sorához és Y oszlopához rendeljük hozzá az $(u \cdot Y, v)$ párt, ahol v a szóban forgó sor és oszlop találkozásánál lévő elem. (Például az $u = H \cdot NDL$ és $Y = LDL$ esetben a kapott rendezett pár $(H \cdot NDL \cdot LDL, DL)$.) Legyen R az összes ilyen módon kapható párok halmaza, amelyekre $u \cdot Y \neq v$. (Nem tartozik tehát R-be az $u = H$ és $Y = NDL$ esethez tartozó pár, mivel itt $v = H \cdot NDL$.) Válasszuk továbbá azt a súlyozást amely szerint $\rho(LH) = 2$ és $\rho(Y) = 1$ minden $Y \in M$-re, amelyre $Y \neq LH$. Állítjuk, hogy R megfelelő lesz.

4.5.1 TÉTEL. *R súly csökkentő, $\xleftrightarrow[T]{*} = \xLeftrightarrow[R]{*}$ és $N = IRR(R)$.*

Bizonyítás. A tétel első állítása könnyen ellenőrizhető a táblázat (és így R definíciója) alapján. Ami a második állítást illeti, ugyancsak a táblázatról látható, hogy $T \subseteq R$ ezért $\xleftrightarrow[T]{*} \subseteq \xLeftrightarrow[R]{*}$. Továbbá, ahogy azt már láttuk, R elemei levezethetők T elemeiből, vagyis $R \subseteq \xleftrightarrow[T]{*}$, tehát $\xLeftrightarrow[R]{*} \subseteq \xleftrightarrow[T]{*}$.

Most azt mutatjuk meg, hogy $N = IRR(R)$. Az $N \subseteq IRR(R)$ tartalmazás nyilvánvaló R definíciójából: nincsen olyan N-beli elem, amelyre alkalmazható lenne valamilyen R-beli szabály. A fordított tartalmazás igazolása végett jelöljük $IRR_n(R)$-el az n hosszúságú R-irreducibilis elemek halmazát, majd mutassuk meg, hogy $IRR_n(R) \subseteq N$ minden $n \geq 0$ esetén. Nyilvánvaló, hogy $IRR_0(R) = \{\lambda\}$ és $IRR_1(R) = M$. Továbbá R definíciójából, tehát a 2. ábrán lévő táblázatból látható, hogy csak a $H \cdot NDL$, $H \cdot LNDL$, $NH \cdot LNDL$ és $LH \cdot LNDL$ kettő hosszúságú szavakra nem alkalmazható R-beli szabály, ezért $IRR_2(R) = \{H \cdot NDL, H \cdot LNDL, NH \cdot LNDL$ és $LH \cdot LNDL\}$. Végül igazoljuk, hogy $IRR_n(R) = \emptyset$ minden $n \geq 3$-ra. Ez abból az észrevételből adódik, hogy amennyiben $IRR_3(R)$-nek lenne eleme, akkor az csak $u \cdot Y$ vagy $Y \cdot u$ alakú lehetne, ahol $u \in IRR_2(R)$ és $Y \in M$. Könnyű azonban ellenőrizni, hogy minden ilyen szóra alkalmazható valamely R-beli szabály, tehát nem lehet R irreducibilis. Tehát $IRR_n(R) \subseteq N$ minden $n \geq 0$-ra. □

A most igazolt tétel, továbbá a 3.4 és 3.5 tételek alkalmazásaként kapjuk, hogy a 4.4.1 tételben megadott algoritmus helyett alkalmazható a 3.5 tételben szereplő standard algoritmus, és így N és R önmagában is elegendő $[M]$ teljes leírására.

5. A determinisztikus felszálló eset

Ebben a fejezetben a 3. fejezetben megadott módszert alkalmazzuk az

$$S = \{DF, LDF, NDF, LNDF, H, LH, NH\}$$

halmazra. A 3. fejezetben leírtakkal analóg módon definiáljuk az $[S]$ monoidot, a $|\ | : S^* \to [S]$ homomorfizmust és a θ kongruencia relációt S^*-on. Mielőtt azonban rátérnénk a módszer alkalmazására, fel kell idéznünk néhány, a felszálló fatranszformátorok kompozícióira vonatkozó tételt.

5.1 Determinisztikus felszálló fatranszformátorok kompozíciói.

A következő kompozíció fogalom először a [Bak] dolgozatban került bevezetésre, de megtalálható a [GécSte] könyvben is.

Legyenek $A = (\Sigma, P, \Delta, R, p_0)$ és $B = (\Delta, Q, \Omega, R', q_0)$ *df* transzformátorok. Az A és B kompozícióján éretjük azt a $C = (\Sigma, P \times Q, \Omega, R'', \langle p_0, q_0\rangle)$ *df* transzformátort, ahol R'' a legszűkebb olyan halmaz, amire teljesül, hogy ha R-ben van egy

$$p(\sigma(x_1, \ldots, x_m)) \rightarrow t(p_1(x_{i_1}), \ldots, p_n(x_{i_n}))$$

szabály, ahol $t \in \hat{T}_{\Delta,n}$ és

$$q(t) \overset{*}{\underset{B}{\Longrightarrow}} r(q_{11}(x_1), \ldots, q_{1v_1}(x_1), \ldots, q_{n1}(x_n), \ldots, q_{nv_n}(x_n)),$$

valamely $q \in Q$, $v_j \geq 0$, $q_{j1}, \ldots, q_{jv_j} \in Q$ $(1 \leq j \leq n)$ és $t \in \hat{T}_{\Omega,v}$ esetén, ahol $v = v_1 + \ldots + v_n$; akkor a

$$\langle p, q\rangle(\sigma(x_1, \ldots, x_m)) \rightarrow r(\langle q_{11}, p_1\rangle(x_1), \ldots, \langle q_{nv_n}, p_n\rangle(x_n))$$

szabály R''-ben van. Könnyen igazolható, hogy az így megkonstruált C *df* transzformátorra érvényes lesz $\tau_A \circ \tau_B \subseteq \tau_C$, l. [Bak]; sőt ennél több is igaz, amit a következő lemmában foglalunk össze.

5.1.1 LEMMA.

(i) Ha A teljesen definiált vagy ha B nemtörlő, akkor $\tau_A \circ \tau_B = \tau_C$.

(ii) Ha mind A, mind B rendelkezik az x tulajdonsággal, ahol x lehet a teljesen definiált, az egyállapotú, a lineáris, a nemtörlő és a lineáris nemtörlő tulajdonságok bármelyike, akkor C is x tulajdonságú.

(iii) Ha A lineáris, B pedig homomorfizmus transzformátorok, akkor C uniform.

(iv) Vezessük be az $\iota_A = Id(\mathrm{dom}(\tau_A))$ jelölést. Ekkor $\iota_A \circ \tau_C = \tau_A \circ \tau_B$.

Bizonyítás. Az (i) állítás következik [Bak] 1. tételéből míg (ii) és (iii) nyilvánvaló a C konstrukciója miatt. Ezért csak (iv)-vel kell foglalkoznunk.

Az egyik irányú tartalmazás könnyű, mivel $\iota_A \circ \tau_A = \tau_A$ és mint már említettük $\tau_A \circ \tau_B \subseteq \tau_C$. Ezért $\tau_A \circ \tau_B = \iota_A \circ \tau_A \circ \tau_B \subseteq \iota_A \circ \tau_C$.

Az $\iota_A \circ \tau_C \subseteq \tau_A \circ \tau_B$ tartalmazáshoz a következőt kell igazolni. Minden $p \in P$, $q \in Q$ és $t \in T_\Sigma$ esetén, ha $t \in \mathrm{dom}(\tau_{A(p)})$, úgy a $\langle p, q\rangle(t) \overset{*}{\underset{C}{\Longrightarrow}} r$ deriváció akkor és csakis akkor érvényes valamely $r \in T_\Delta$-re, ha a $t' = \tau_{A(p)}(t)$ fára teljesül, hogy $q(t') \overset{*}{\underset{B}{\Longrightarrow}} r$. Az állítás t szerinti strukturális indukcióval igazolható. (Megjegyezzük, hogy a $t \in \mathrm{dom}(\tau_{A(p)})$ feltétel fontos, az ekvivalencia nem igaz tetszőleges $t \in T_\Sigma$-ra, mert ekkor nem biztos, hogy a t' fa létezik. Emiatt nem igaz általában a $\tau_C \subseteq \tau_A \circ \tau_B$ tartalmazás sem.) □

Vegyük észre, hogy az előbbi lemma (i) és (ii) pontjainak (együttes) alkalmazásával több $Y \circ Z \subseteq V$ alakú tartalmazás igazolható, ahol $Y, Z, V \in S$. Ilyenek például a $DF \circ NDF \subseteq DF$, $H \circ DF \subseteq DF$ tartalmazások, amelyek (i)-ből következnek, és az $LH \circ LNDF \subseteq LDF$ tartalmazás, amihez (ii)-t is felhasználtuk (különben csak azt tudnánk, hogy $LH \circ LNDF \subseteq DF$).

Egy további észrevétel, hogy — csakúgy mint a leszálló esetben — az $Y \circ Z \subseteq Z$ alakú tartalmazásokból itt is következik, hogy $Y \circ Z = Z$, mivel $I \subseteq Y$ minden $Y \in S$ esetén.

5.2 A T Thue rendszer megadása és a $\overset{*}{\underset{T}{\longleftrightarrow}} \subseteq \theta$ tartalmazás.

Legyen T (amiről később igazoljuk, hogy θ-t generálja) az alábbi 19 formális egyenlet halmaza.

(1)	$DF \cdot NDF \doteq DF$	(6)	$NDF \cdot LH \doteq DF^2$
(2)	$DF \cdot NH \doteq DF$	(7)	$LDF \cdot LNDF \doteq LDF$
(3)	$NDF^2 \doteq NDF$	(8)	$LNDF \cdot DF \doteq DF^2$
(4)	$NDF \cdot LNDF \doteq NDF$	(9)	$LNDF \cdot NDF \doteq NDF$
(5)	$NDF \cdot NH \doteq NDF$	(10)	$LNDF^2 \doteq LNDF$
(11)	$LNDF \cdot LDF \doteq LDF^2$	(16)	$NH \cdot LH \doteq H$
(12)	$LNDF \cdot LH \doteq LDF^2$	(17)	$LH \cdot LDF \doteq LDF$
(13)	$NH \cdot NDF \doteq NDF$	(18)	$LH^2 \doteq LH$
(14)	$NH \cdot LDF \doteq DF$	(19)	$LH \cdot NH \doteq H$
(15)	$NH^2 \doteq NH$		

A következő tételben igazoljuk, hogy T elemei érvényesek $[S]$-ben.

5.2.1 TÉTEL. *Minden $u \doteq v \in T$ esetén $|u| = |v|$ teljesül.*

Bizonyítás. Természetesen a formális egyenletek legtöbbjét most is jól ismert dekompozíciós eredményekből származtattuk. Ily módon az (1)–(5), (7), (9), (10), (13), (15), (17) és (18) egyenletek érvényessége az 5.1.1 lemma és az azt követő megjegyzések értelmében nyilvánvaló. A [Bak] dolgozat 11. tételében került igazolásra, hogy (14) is érvényes és a tétel bizonyításából kiderült, hogy (16) nem más, mint (14) speciális esete; nevezetesen a 11. tétel h transzformációkra való alkalmazásaként kapjuk. Ezért már csak (6), (8), (11), (12) és (19) érvényességének igazolása van hátra. Ez utóbbival kezdjük.

5.2.2 LEMMA. *$LH \circ NH = H$.*

Bizonyítás. Ugyancsak az 5.1.1 lemmából és az azt követő megjegyzésekből következik, hogy $LH \circ NH \subseteq H$. A fordított irányú tartalmazás igazolása végett vegyünk egy $A = (\Sigma, q, \Delta, R, q)$ h transzformátort. Írjuk az R-beli szabályokat

$$(*) \qquad q(\sigma(x_1, \ldots, x_m)) \to r(q(x_{i_1}), \ldots, q(x_{i_n}))$$

alakba, ahol $m, n \geq 0$, $\sigma \in \Sigma_m$, $1 \leq i_1 < \ldots < i_n \leq m$, $r \in T_{\Delta,n}$ és minden $1 \leq j \leq n$-re teljesül, hogy az x_j változó legalább egyszer előfordul r-ben. (Nyilvánvaló, hogy az R-beli szabályok mindig ilyen alakra hozhatók.) Minden $\sigma \in \Sigma$ esetén legyen σ' egy új szimbólum, aminek aritása (az előbb definiált) n és legyen $\Sigma' = \{\sigma' \mid \sigma \in \Sigma\}$.

Most vezessük be a $B = (\Sigma, p, \Sigma', R', p)$ és a $C = (\Sigma', u, \Delta, R'', u)$ h transzformátorokat, ahol R' és R'' a legszűkebb olyan halmazok, amelyekre teljesül, hogy ha egy $(*)$ alakú szabály R-ben van, akkor a $p(\sigma(x_1, \ldots, x_m)) \to \sigma'(p(x_{i_1}), \ldots, p(x_{i_n}))$ szabály R'-ben, az $u(\sigma'(x_1, \ldots, x_n)) \to r(u(x_1), \ldots, u(x_n))$ szabály pedig R''-ben van. A szabályhalmazok konstrukciója miatt nyilvánvaló az, hogy B lineáris, C pedig nemtörlő. Továbbá, t szerinti indukcióval igazolható, hogy minden $t \in T_\Sigma$ esetén a $q(t) \overset{*}{\underset{A}{\Longrightarrow}} s$ deriváció akkor és csakis akkor teljesül valamely $s \in T_\Delta$-re, ha van olyan $t' \in T_{\Sigma'}$, amelyre $p(t) \overset{*}{\underset{B}{\Longrightarrow}} t'$ és $u(t') \overset{*}{\underset{C}{\Longrightarrow}} s$, ami azt bizonyítja, hogy $\tau_A = \tau_B \circ \tau_C$ tehát, hogy $H \subseteq LH \circ NH$. □

Most a (8) formális egyenlet érvényességét bizonyítjuk be.

5.2.3 Lemma. $LNDF \circ DF = DF^2$.

Bizonyítás. Csak azt kell belátnunk, hogy $DF^2 \subseteq LNDF \circ DF$, mivel a másik irányú tartalmazás nyilvánvalóan fennáll. Vegyünk két tetszőleges $A = (\Sigma, P, \Delta, R, p_0)$ és $B = (\Delta, Q, \Omega, R', q_0)$ df transzformátort és konstruáljuk meg ezek C kompozícióját, ahogyan 5.1-ben tettük. Ekkor az 5.1.1 lemma szerint az $\iota_A = Id(\mathrm{dom}(\tau_A))$ identikus fatranszformációra fennáll az $\iota_A \circ \tau_C = \tau_A \circ \tau_B$ egyenlőség. Másrészt jól ismert az a tény, hogy ι_A indukálható egy determinisztikus felszálló faautomatával, lásd [GécSte] 213. oldal, ami viszont egy speciális $lndf$ transzformátornak tekinthető. Ezzel a lemma bizonyítását befejeztük. □

Vegyük észre, hogy ha az előbbi lemmában A és B ldf transzformátorok, akkor az 5.1.1 lemma (ii) pontja értelmében C is az lesz. Ezért az $LNDF \circ LDF = LDF^2$ egyenlőség, tehát (11) is érvényes.

Hátra van még (6) és (12) érvényességének igazolása.

5.2.4 Lemma. $NDF \circ LH = DF^2$.

Bizonyítás. Először azt bizonyítjuk be, hogy $DF \subseteq NDF \circ LH$. Tekintsünk ezért egy tetszőleges $A = (\Sigma, Q, \Delta, R, q_0)$ df transzformátort. Legyen az R szabály halmaz számossága k. Számozzuk meg a szabályokat 1-től k-ig, és írjuk le a szabályok listáját

$$i : q(\sigma(x_1, \ldots, x_m)) \to r(q_{11}(x_1), \ldots, q_{1n_1}(x_1), \ldots, q_{m1}(x_m), \ldots, q_{mn_m}(x_m))$$

alakban, ahol $1 \leq i \leq k$, $m \geq 0$, $\sigma \in \Sigma_m$, $q \in Q$ és minden $1 \leq j \leq m$-re $n_j \geq 0$, $q_{j1}, \ldots, q_{jn_j} \in Q$. Továbbá $r \in \hat{T}_{\Delta,n}$, ahol $n = n_1 + \ldots + n_m$. (Megjegyezzük, hogy n_j szemléletes jelentése az, hogy x_j hányszor fordul elő az i-edik szabály jobboldalán $(1 \leq j \leq m)$ és azt, hogy n_j természetesen függ i-től is, de ezt a jobb áttekinthetőség végett nem specifikáltuk.)

Meg fogunk adni egy B *ndf* transzformátort és egy C *lh* transzformátort amelyekre $\tau_A = \tau_B \circ \tau_C$.

Először minden $1 \leq i \leq k$-ra és $1 \leq j \leq m$-re vezessük be az

$$u_j = \begin{cases} n_j, & \text{ha } n_j > 0 \\ 1, & \text{ha } n_j = 0 \end{cases}$$

számokat. Vegyünk továbbá minden $1 \leq i \leq k$-ra egy σ_i műveleti szimbólumot aminek az aritása $u = u_1 + \ldots + u_m$ és legyen Σ' az ezen szimbólumokból álló rangolt ábécé.

Most konstruáljuk meg a $B = (\Sigma, Q', \Omega, R', q_0)$ *df* transzformátort, ahol

(a) $Q' = Q \cup \{p\}$, $p \notin Q$,

(b) $\Omega = \Sigma \cup \Sigma'$,

(c) R' pedig az alábbi két szabály szerint épül fel:

(i) minden $1 \leq i \leq k$-ra R' tartalmazza a

$$q(\sigma(x_1, \ldots, x_m)) \to \sigma_i(p_{11}(x_1), \ldots, p_{1u_1}(x_1), \ldots, p_{m1}(x_m), \ldots, p_{mu_m}(x_m))$$

szabályt, ahol $q(\sigma(x_1, \ldots, x_m))$ az R-beli i-edik szabály baloldala, σ_i az i-edik szabályhoz tartozó új műveleti szimbólum, a $p_{j1}, \ldots, p_{ju_j}$ sorozatok, $1 \leq j \leq m$-re, pedig az alábbi egyenlőséggel vannak definiálva:

$$p_{j1}, \ldots, p_{ju_j} = \begin{cases} q_{j1}, \ldots, q_{jn_j}, & \text{ha } n_j > 0 \\ p, & \text{ha } n_j = 0. \end{cases}$$

(ii) minden $m \geq 0$, $\sigma \in \Sigma_m$ esetén R' tartalmazza a $p(\sigma(x_1, \ldots, x_m)) \to \sigma(p(x_1), \ldots, p(x_m))$ szabályt.

Mindenekelőtt jegyezzük meg azt, hogy B nemtörlő *df* transzformátor lesz. Ez az (i) típusú szabályokat nézve abból következik, hogy $u_j \geq 1$ minden $1 \leq j \leq m$-re, míg az (ii) típusú szabályok esetén nyilvánvaló. Ugyancsak nyilvánvaló az, hogy $\tau_{B(p)} = Id(T_\Sigma)$.

Ezután vezessük be a $C = (\Omega, c, \Delta, R'', c)$ *h* transzformátort, ahol R'' tartalmazza az összes

$$c(\sigma_i(x_1, \ldots, x_u)) \to$$
$$r(c(x_1), \ldots, c(x_{n_1}), \ldots, c(x_{u_1+\ldots+u_{m-1}+1}), \ldots, c(x_{u_1+\ldots+u_{m-1}+n_m}))$$

alakú szabályt, ahol $1 \leq i \leq k$, σ_i az i-edik R-beli szabályhoz tartozó új műveleti szimbólum, r pedig az i-edik R-beli szabály jobboldalán álló fa. Továbbá, hogy C teljesen definiált legyen, minden $m \geq 0$ és $\sigma \in \Sigma$ esetén vegyük még fel R''-be a $c(\sigma(x_1, \ldots, x_m)) \to \sigma(c(x_1), \ldots, c(x_m))$ szabályt is. Látható, hogy C nem többszöröz meg egyetlen részfát sem, tehát lineáris.

Végül, t szerinti indukcióval igazolható, hogy minden $t \in T_\Sigma$ és $q \in Q$ esetén a

$$q(t) \overset{*}{\underset{A}{\Longrightarrow}} s$$

deriváció akkor és csakis akkor teljesül valamilyen $s \in T_\Delta$ esetén, ha van olyan $t' \in T_\Omega$, amelyre

$$q(t) \underset{B}{\overset{*}{\Longrightarrow}} t' \quad \text{és} \quad c(t') \underset{C}{\overset{*}{\Longrightarrow}} s,$$

amiből a $q = q_0$ választással azt kapjuk, hogy $\tau_A = \tau_B \circ \tau_C$.

Ezzel igazoltuk, hogy $DF \subseteq NDF \circ LH$. Így az 5.2.3 lemma felhasználásával kapjuk, hogy $DF^2 = LNDF \circ DF \subseteq LNDF \circ NDF \circ LH$, ahonnan a (9) formális egyenlet már igazolt érvényessége miatt adódik, hogy $DF^2 \subseteq NDF \circ LH$. Mivel az ellentétes irányú tartalmazás nyilvánvalóan teljesül, az 5.2.4 lemma bizonyítását befejeztük. □

Befejezésül észrevesszük, hogy amennyiben az előző lemmában A *ldf* transzformátor, akkor B *lndf* transzformátor lesz, ezért azt kapjuk, hogy $LDF \subseteq LNDF \circ LH$. Így a (11) és (10) érvényessége miatt adódik, hogy $LDF^2 = LNDF \circ LDF \subseteq LNDF \circ LNDF \circ LH = LNDF \circ LH$. A fordított irányú tartalmazás itt is nyilvánvaló, tehát (12) is érvényes $[S]$-ben.

Ezzel az 5.2.1 tétel bizonyítását befejeztük. □

5.2.5 KÖVETKEZMÉNY. $\underset{T}{\overset{*}{\longleftrightarrow}} \subseteq \theta$.

Bizonyítás. L. 3.2 lemma.

5.3. Reprezentáns rendszer és tartalmazási diagram.

Bevezetjük azt az N halmazt, amiről később bebizonyítjuk, hogy θ-nak reprezentáns rendszere. Evégett legyen
$N_1 = \{DF, LDF, NDF, LNDF, DF^2, LDF{\cdot}NDF, LDF^2, H{\cdot}NDF, LDF^2{\cdot}NDF\}$.
és
$N_2 = \{H, NH, LH, LDF \cdot NH, LDF \cdot H\}$.

Továbbá, definiáljuk a z_k szavakat minden $k \geq 0$-ra a következőképpen. Legyen $z_0 = \lambda$, legyen $z_{k+1} = z_k \cdot LNDF$, ha k páros szám és $z_{k+1} = z_k \cdot NH$, ha k páratlan szám.

Azt állítjuk, hogy az

$$N = N_1 \cup N_2 \cup \{z_k \mid k \geq 0\} \cup \{z \cdot z_k \mid z \in N_2,\ k \geq 1\}$$

halmaz a θ reprezentáns rendszere. Az eddigiekből már tudjuk, hogy ehhez először meg kell adnunk az $|N| = \{|u| \mid u \in N\}$ halmaznak a tartalmazási diagramját. Ezt fogjuk megtenni ebben a fejezetben, ami egyébként az értekezés leghosszabb önálló része. Módszerünk most is az lesz, mint a leszálló esetben, vagyis mnegadunk egy diagramot, majd bebizonyítjuk, hogy ez az $|N|$ tartalmazási diagramja.

Legelőször vezessük be $k \geq 0$ esetén a $C_k = |z_k|$ jelölést; tehát $C_0 = I$, $C_{k+1} = C_k \circ LNDF$, ha k páros és $C_{k+1} = C_k \circ NH$, ha k páratlan.

Ezután felidézünk három olyan tételt, amelyek elengedhetetlenül szükségesek a tartalmazási diagram megadásához. A tételek bizonyításait azonban terjedelmi okok miatt el kell hagynunk abban a reményben, hogy a dolgozat enélkül még érthető marad.

5.3.1 TÉTEL. ([VágFül], 3. tétel). *Minden* $k \geq 1$ *esetén van olyan* $\tau_k \in C_{2k}$ *fatranszformáció, amelyre* $\tau_k \notin C_{2k-2}$. *Ezért a* $\{C_{2k} \mid k \geq 0\}$ *rendszer valódi hierarchia. Továbbá* τ_k *teljesen definiált és megőrzi a fák magasságát, tehát minden* $(t,s) \in \tau_k$ *-ra* $h(t) = h(s)$.

Megjegyezzük, hogy az utóbbi feltétel biztosítja azt, hogy τ_k nem vihet végtelen sok fát ugyanabba az output fába. Ezt a tulajdonságát τ_k-nak a későbbiekben fel fogjuk használni.

5.3.2 TÉTEL. ([VágFül], 17. lemma). *Van olyan* A *teljesen definiált ndf transzformátor, amelyre* $\tau_A \notin C_k$ *semmilyen* $k \geq 0$ *esetén, ezért* $\bigcup C_k \subset NDF$. *Továbbá* τ_A *injektív leképezés.*

Itt is fel fogjuk használni azt a tényt, hogy τ_A — lévén injektív — nem vihet végtelen sok fát ugyanazon output fába.

A következő tétel azt fejezi ki, hogy az *ndf* transzformációkkal kapott output halmazok osztálya szűkebb, mint a (tetszőleges) *df* transzformációk esetén kapott osztály.

5.3.3 TÉTEL. ([FülVág2], 3.2 lemma). $NDF(REC) \subset DF(REC)$.

Ezután tekintsük a 3. ábrán lévő diagramot. Erről fogjuk bebizonyítani, hogy $|N|$ tartalmazási diagramja.

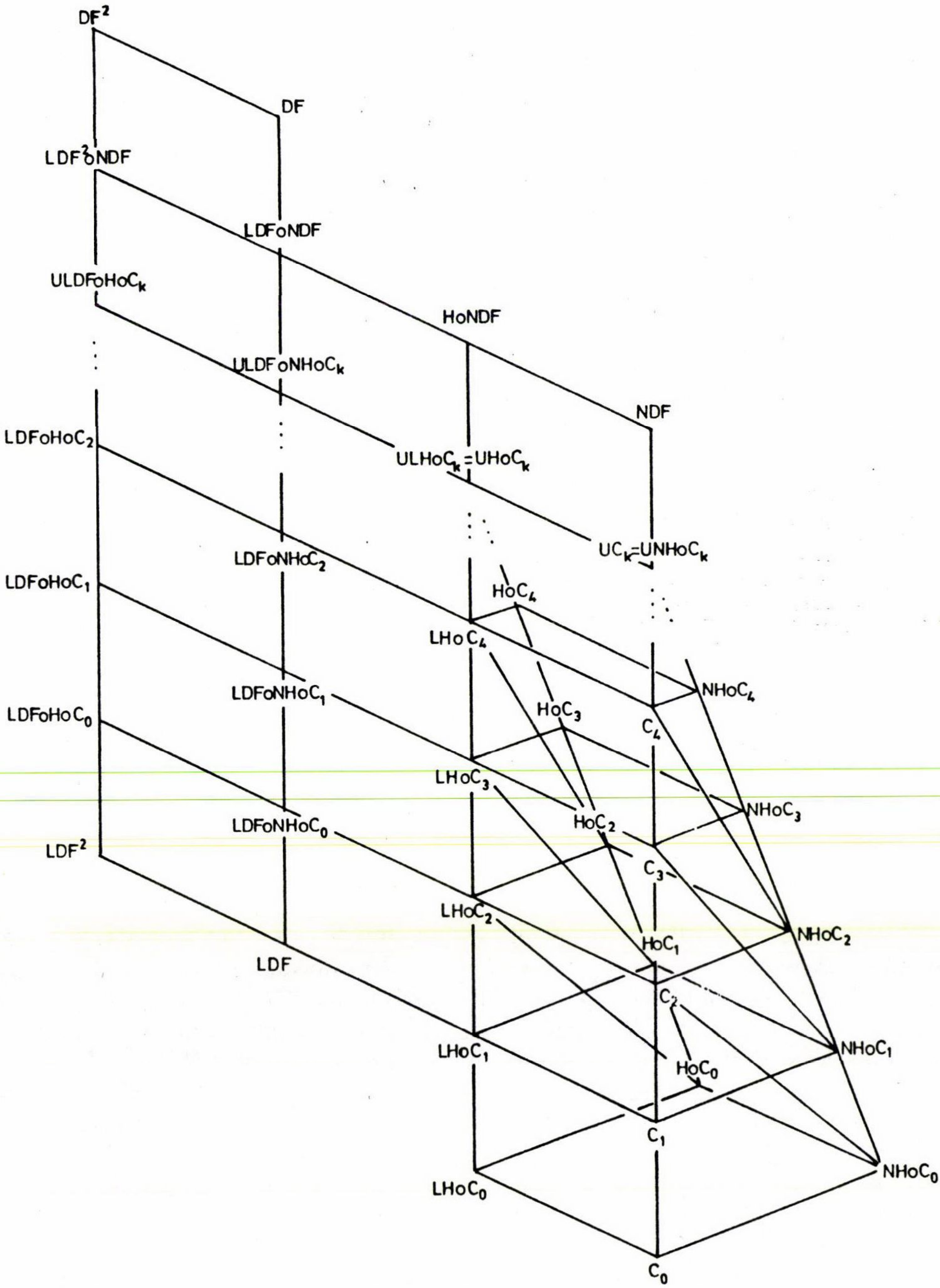
Először is látható, hogy a diagram $|N|$ elemein kívül tartalmazza hat hierarchiának a szuprémumát is. Ezeket azért tüntettük fel, mert azt is igazolni tudjuk, hogy ezeket a szuprémumokat valódi módon tartalmazzák a felettük lévő elemek, pl. $\bigcup LH \circ C_k = \bigcup H \circ C_k \subset H \circ NDF$, stb.

A következő lemmával kezdjük.

5.3.4 LEMMA. *Ha egy* Z *osztály egy* Y *osztály felett van és él vezet* Z*-ből* Y*-ba, akkor* $Y \subseteq Z$.

Bizonyítás. (a) Először tekintsük a diagramnak az $LH \circ C_0$ és a $H \circ NDF$ osztályokat összekötő vonaltól jobbra eső részét. Nyilvánvaló az, hogy $C_k \subseteq LH \circ C_k$ és $C_k \subseteq NH \circ C_k$ minden $k \geq 0$-ra. Továbbá, mivel $LH \subseteq H$ és $NH \subseteq H$, az $LH \circ C_k \subseteq H \circ C_k$ és az $NH \circ C_k \subseteq H \circ C_k$ tartalmazások is teljesülnek. Ugyancsak nyilvánvaló az $NH \circ C_k \subseteq C_{k+2}$ tartalmazás, mivel $C_{k+2} = LNDF \circ NH \circ C_k$, ezért $\bigcup NH \circ C_k \subseteq \bigcup C_k$ és így $\bigcup NH \circ C_k = \bigcup C_k$. Hasonló módon a $H = LH \circ NH$ egyenlőséget felhasználva, kapjuk, hogy $H \circ C_k \subseteq LH \circ C_{k+1}$ ezért $\bigcup LH \circ C_k = \bigcup H \circ C_k$ is fennáll. Végül a definíciókból következik, hogy fennállnak az $NH \subseteq NDF$ és az $LNDF \subseteq NDF$ tartalmazások és NDF zárt a kompozícióra. Következésképpen $C_k \subseteq NDF$ minden $k \geq 0$-ra és így $\bigcup C_k \subseteq NDF$ is. Ugyanezen okok miatt az $\bigcup H \circ C_k \subseteq H \circ NDF$ tartalmazás is igaz.

(b) Vegyük az LDF és a DF osztályokat összekötő vonalat. Az 5.1.1 lemma (i) és (ii) pontjai miatt fennáll az $LH \circ LNDF \subseteq LDF$ tartalmazás, tehát $LH \circ C_1 \subseteq LDF$. Továbbá a (7) egyenlet miatt $LH \circ C_k = LH \circ LNDF \circ NH \circ C_{k-2} \subseteq LDF \circ NH \circ C_{k-2}$ minden $k \geq 2$-re, ezért $\bigcup LH \circ C_k \subseteq \bigcup LDF \circ NH \circ C_k$. Az



DF²
DF
LDF²oNDF
LDFoNDF
ULDFoHoCk
HoNDF
ULDFoNHoCk
NDF
LDFoHoC2
ULHoCk=UHoCk
UCk=UNHoCk
LDFoNHoC2
LDFoHoC1
HoC4
LHoC4
LDFoNHoC1
NHoC4
LDFoHoC0
HoC3
C4
LHoC3
NHoC3
LDFoNHoC0
HoC2
LDF²
C3
LHoC2
NHoC2
HoC1
LDF
C2
LHoC1
NHoC1
HoC0
C1
NHoC0
LHoC0
C0

3. ábra

$\bigcup LDF \circ NH \circ C_k \subseteq LDF \circ NDF$ tartalmazás ugyancsak annak a ténynek a következménye, hogy NDF zárt a kompozícióra, míg $LDF \circ NDF \subseteq DF$ az 5.1.1 lemmából következik. Végül, az 5.2.1 tételben bizonyított egyenlőségek értelmében $H \circ NDF = LH \circ NH \circ NDF = LH \circ NDF$, ezért $H \circ NDF \subseteq LDF \circ NDF$.

(c) Befejezésül tekintsük az LDF^2 és a DF^2 osztályokat összekötő vonalat. Az $LDF \subseteq LDF^2$ tartalmazás nyilvánvaló és $NH \subseteq H$ miatt az is, hogy $LDF \circ NH \circ C_k \subseteq LDF \circ H \circ C_k$ minden $k \geq 0$-ra. Következésképpen $\bigcup LDF \circ NH \circ C_k \subseteq \bigcup LDF \circ H \circ C_k$. A $H = LH \circ NH$ egyenlőség és az (a) pontban igazolt $\bigcup NH \circ C_k \subseteq NDF$ tartalmazás miatt $\bigcup LDF \circ H \circ C_k \subseteq LDF^2 \circ NDF$, mely utóbbi az $LDF \circ NDF \subseteq DF$ tartalmazás miatt, l. 5.1.1 lemma, a DF^2 részhalmaza. □

A fejezet hátralevő részében azt kell igazolnunk, hogy a diagram által mutatott valamennyi tartalmazás valódi, és hogy a relációba nem állított osztályok összehasonlíthatatlanok. Azzal fogunk kezdeni, hogy igazoljuk: a diagramban szereplő hat hierarchia mindegyike valódi.

5.3.5 LEMMA. *A $\{C_k \mid k \geq 0\}$ hierarchia valódi.*

Bizonyítás. Nyilvánvaló az, hogy $C_0 \subset C_1$. Indirekt módon tegyük fel, hogy $C_k = C_{k+1}$ valamilyen, mondjuk páros, $k \geq 1$ esetén. Akkor $C_k \circ NH = C_{k+1} \circ NH$ is teljesül, tehát $C_k = C_{k+2}$. Mivel $C_{k+1} = C_{k+2}$, azt is kapjuk, hogy $C_{k+1} \circ LNDF = C_{k+2} \circ LNDF$, vagyis $C_{k+1} = C_{k+3}$. Tovább folytatva, nyerjük, hogy $C_k = C_{k+l}$, minden $l \geq 1$-re, ami ellentmond annak a ténynek, l. 5.3.1 tétel, hogy $\{C_{2k} \mid k \geq 0\}$ valódi hierarchia. Hasonlóan ellentmondáshoz vezet az az eset is, amikor k páratlan, tehát a lemmát bebizonyítottuk. □

5.3.6 LEMMA. *A $\{H \circ C_k \mid k \geq 0\}$ hierarchia valódi.*

Bizonyítás. Először megmutatjuk, hogy minden $k \geq 2$ esetén $H \circ C_{2k-4} \subset H \circ C_{2k}$. Legyen evégett $k \geq 2$, és tekintsük az 5.3.1 tételben szereplő τ_k fatranszformációt. Mivel $\tau_k \in C_{2k}$ ezért $\tau_k \in H \circ C_{2k}$ is fennáll. Bebizonyítjuk, hogy $\tau_k \notin H \circ C_{2k-4}$. Feltéve az ellenkezőjét, van olyan A h transzformátor és egy $\tau \in C_{2k-4}$ fatranszformáció, amelyre $\tau_k = \tau_A \circ \tau$. Állítjuk, hogy A nemtörlő. Valóban, ha A törlő, akkor τ_A végtelen sok input fát visz ugyanabba a fába. Másrészt $\mathrm{ran}(\tau_A) \subseteq \mathrm{dom}(\tau)$, mivel τ_k teljesen definiált, ezért $\tau_A \circ \tau$ is végtelen sok fát visz ugyanabba a fába, ami ellentmond 5.3.1 tételnek. Tehát A nh transzformátor, vagyis $\tau_k \in NH \circ C_{2k-4} \subseteq C_{2k-2}$, ami ugyancsak ellentmondásban van az 5.3.1 tétellel, ezért valóban $H \circ C_{2k-4} \subset H \circ C_{2k}$.

Innen a bizonyítás az előző lemma bizonyításához hasonlóan fejezhető be. A $H \circ C_0 \subset H \circ C_1$ tartalmazás ugyancsak triviális. Továbbá, a $H \circ C_k = H \circ C_{k+1}$ egyenlőség megint csak maga után vonja, hogy $H \circ C_k = H \circ C_{k+l}$, minden $l \geq 1$-re, ami viszont az előbb bizonyítottak szerint lehetetlen. □

5.3.7 KÖVETKEZMÉNY. *Az $\{NH \circ C_k \mid k \geq 0\}$ és az $\{LH \circ C_k \mid k \geq 0\}$ valódi hierarchiák.*

Bizonyítás. Ha valamilyen $k \geq 0$-ra $NH \circ C_k = NH \circ C_{k+1}$ lenne, akkor balról szorozva LH-val, és felhasználva a $H = LH \circ NH$ egyenlőséget azt kapnánk, hogy $H \circ C_k = H \circ C_{k+1}$, ami az előző lemma szerint nem lehet. Hasonló módon ellentmondáshoz vezetne az $LH \circ C_k = LH \circ C_{k+1}$ egyenlőség is. □

5.3.8 LEMMA. *Az $\{LDF \circ H \circ C_k \mid k \geq 0\}$ hierarchia valódi.*

Bizonyítás. Először azt bizonyítjuk be, hogy minden $k \geq 1$ esetén $LDF \circ H \circ C_{2k-2} \subset LDF \circ H \circ C_{2k}$. Rögzítsünk egy tetszőlegesen választott $k \geq 1$ egész számot és tekintsük az 5.3.1 tételben szereplő $\tau_{k+1} \in C_{2k+2}$ fatranszformációt, amelyre természetesen $\tau_{k+1} \in LDF \circ H \circ C_{2k}$ is teljesül. Meg fogjuk mutatni, hogy $\tau_{k+1} \notin LDF \circ H \circ C_{2k-2}$. Indirekt módon, tegyük fel az ellenkezőjét. Ekkor létezik egy A *ldf* transzformátor, egy B *h* transzformátor és egy $\tau \in C_{2k-2}$ fatranszformáció amelyekre $\tau_{k+1} = \tau_A \circ \tau_B \circ \tau$. Észrevesszük, hogy A teljesen definiált kell legyen, mivel τ_{k+1} is teljesen definiált. Konstruáljuk meg A-nak és B-nek a C kompozícióját, amelyre az 5.1.1 tétel (i), (ii) és (iii) pontjai értelmében azt kapjuk, hogy $\tau_C = \tau_A \circ \tau_B, C$ teljesen definiált és uniform. Ezért $\tau_{k+1} = \tau_C \circ \tau$ is érvényes lesz.

Azt állítjuk, hogy C nemtörlő. Valóban, ha ez nem lenne igaz, akkor τ_C végtelen sok fát vinne ugyanabba a fába. Továbbá, mivel τ_{k+1} teljesen definiált, annak is teljesülnie kell, hogy $\mathrm{ran}(\tau_C) \subseteq \mathrm{dom}(\tau)$. Azonban ebből a két dologból az következne, hogy $\tau_C \circ \tau$ is végtelen sok fát vinne ugyanabba a fába, ami lehetetlen, tehát C mégis csak nemtörlő.

Mivel C uniform is, van olyan A' *lndf* transzformátor és B' *nh* transzformátor, hogy $\tau_C = \tau_{A'} \circ \tau_{B'}$, l. [VágFül], 14. lemma utáni érvelést. Következésképpen $\tau_{k+1} = \tau_{A'} \circ \tau_{B'} \circ \tau$, vagyis $\tau_{k+1} \in LNDF \circ NH \circ C_{2k-2} = C_{2k}$, ami viszont ellentmond 5.3.1 tételnek. Ez az ellentmondás a $\tau_{k+1} \in LDF \circ H \circ C_{2k-2}$ feltevésből származik, ami azt igazolja, hogy $LDF \circ H \circ C_{2k-2} \subset LDF \circ H \circ C_{2k}$ minden $k \geq 1$-re. Innen a bizonyítás ugyanúgy fejezhető be mint az 5.3.5 lemma bizonyítása. □

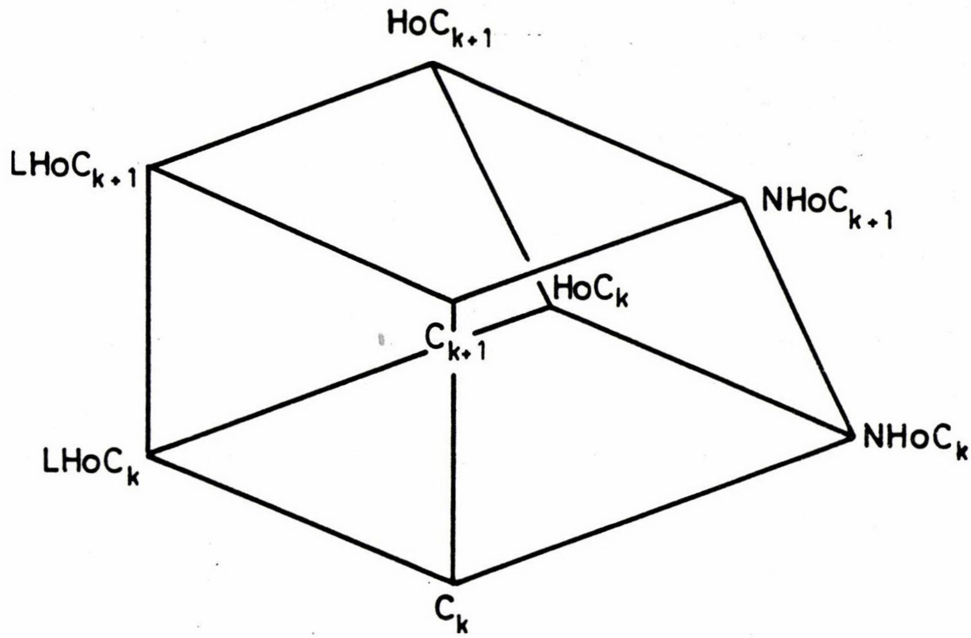
5.3.9 KÖVETKEZMÉNY. *Az $\{LDF \circ NH \circ C_k \mid k \geq 0\}$ hierarchia is valódi.*

Bizonyítás. Tegyük fel az ellenkezőjét, vagyis azt, hogy $LDF \circ NH \circ C_k = LDF \circ NH \circ C_{k+1}$ valamelyik $k \geq 0$-ra. Ekkor $LDF^2 \circ NH \circ C_k = LDF^2 \circ NH \circ C_{k+1}$ egyenlőségnek is fenn kell állnia. Másrészt megmutatható, hogy $LDF^2 \circ NH = LDF \circ H$. Valóban, a $H = LH \circ NH$, $LDF = LDF \circ LNDF$, $LNDF \circ LH = LDF^2$ és $LNDF \circ LDF = LDF^2$, 5.2.1 tételben bizonyított egyenlőségek felhasználásával a következő számolás végezhető el: $LDF \circ H = LDF \circ LH \circ NH = LDF \circ LNDF \circ LH \circ NH = LDF^3 \circ NH = LDF \circ LNDF \circ LDF \circ NH = LDF^2 \circ NH$.

A kettőt összevetve azt kapjuk, hogy $LDF \circ H \circ C_k = LDF \circ H \circ C_{k+1}$, ami ellentmond az 5.3.8 lemmának, tehát kiinduló feltevésünk helytelen volt. □

Most a diagram bizonyos részeiről mutatjuk meg, hogy tartalmazási diagramok.

5.3.10 LEMMA. *Az alábbi diagram a feltüntetett fatranszformáció osztályok tartalmazási diagramja minden $k \geq 0$ esetén.*



Bizonyítás. Azt már láttuk az 5.3.5-6 lemmákban és az 5.3.7 következményben, hogy az oldalélek által mutatott tartalmazások valódiak. Most az (a)–(c) pontokban arra mutatunk rá, hogy a lapélek is valódi tartalmazásokat reprezentálnak.

(a) Minden $k \geq 0$ esetén $C_k \subset LH \circ C_k$ és $NH \circ C_k \subset H \circ C_k$. Ezek abból a nyilvánvaló tényből következnek, hogy a jobb oldalak tartalmaznak törlő fatranszformációkat is, míg a bal oldalak nem.

(b) $LH \circ C_k \subset H \circ C_k$ tetszőleges $k \geq 0$-ra. Mivel $k \leq 1$ esetén az állítás nyilvánvaló, csak a $k > 1$ esettel foglalkozunk. Tegyük fel, hogy valamely $k > 1$-re $LH \circ C_k = H \circ C_k$. Balról szorozva $LNDF$-fel, a $H = LH \circ NH$ és az $LNDF \circ LH = LDF^2$ egyenletek felhasználásával azt kapjuk, hogy $LDF^2 \circ C_k = LDF^2 \circ NH \circ C_k$. Innen, mivel $LDF \circ LNDF = LDF$, az adódik, hogy $LDF^2 \circ NH \circ C_{k-2} = LDF^2 \circ NH \circ C_k$, ami ellentmondásban van az 5.3.8 lemmával, mert mint ahogy azt az előbb láttuk $LDF^2 \circ NH = LDF \circ H$. Tehát a feltevés helytelen volt. □

(c) A $C_k \subset NH \circ C_k$ tartalmazás közvetlen következménye (b)-nek és az $LH \circ NH = H$ egyenlőségnek.

Azt kell még igazolni, hogy a relációba nem állított osztályok összehasonlíthatatlanok. Ezt a (d)–(f) pontokban tesszük meg.

(d) $LH \circ C_k \not\subseteq NH \circ C_{k+1}$ semmilyen $k \geq 0$ esetén, mivel a bal oldal tartalmaz törlő fatranszformációkat.

(e) $NH \circ C_k \not\subseteq LH \circ C_{k+1}$ semmilyen $k \geq 0$-ra. Valóban, ha $NH \circ C_k \subseteq LH \circ C_{k+1}$ lenne, akkor LH-val szorozva balról azt kapnánk, hogy $H \circ C_k \subseteq LH \circ C_{k+1}$. Ebből viszont, NH-val vagy $LNDF$-fel szorozva jobbról, az adódna, hogy $H \circ C_{k+1} \subseteq LH \circ C_{k+1}$, aminek éppen az ellenkezője érvényes (b) szerint.

(f) $H \circ C_k \not\subseteq C_{k+1}$ semmilyen $k \geq 0$-ra. Ez az állítás nyilvánvaló, mivel a bal oldal tartalmaz törlő fatranszformációkat is.

Ezen három nem-tartalmazás segítségével be lehet látni az összes többi nem-tartalmazást is. Pl. $LH \circ C_k \not\subseteq C_{k+1}$, mert ha $LH \circ C_k \subseteq C_{k+1}$ lenne, akkor $LH \circ C_k \subseteq NH \circ C_{k+1}$ is teljesülne, ami pedig (d) szerint nem lehet. A lemma bizonyítását befejeztük. □

Ezután be fogjuk látni, hogy a diagramnak az $LH \circ C_0$ és a $H \circ NDF$ osztályokat összekötő vonaltól jobbra eső része tartalmazási diagram. Evégett helyezzük az 5.3.10 lemmában kapott tartalmazási diagramokat egymás tetejére. Ez természetesen újabb éleket is eredményezhet, mint ahogy azt az alábbi következmény mutatja.

5.3.11 KÖVETKEZMÉNY. *Minden $k \geq 0$-ra, $NH \circ C_k \subset C_{k+2}$ és $H \circ C_k \subset LH \circ C_{k+2}$.*

Bizonyítás. Az, hogy mindkét esetben tartalmazás áll fenn, triviális. Másrészt az előző lemma miatt egyenlőség nem lehet, tehát a tartalmazások valódiak. □

Több új élet azonban nem kapunk a „kis" tartalmazási diagramok egymásra rakásával. Ez a következőképpen látható be. Az $LH \circ C_0$ és az NDF osztályok összehasonlíthatatlanok, definíciójuk alapján látszik. Ezért $LH \circ C_k \not\subseteq C_l$ és $LH \circ C_k \not\subseteq NH \circ C_l$ semmilyen $k < l$ esetén. Ugyanezen ok miatt $H \circ C_k \not\subseteq C_l$ és $H \circ C_k \not\subseteq NH \circ C_l$ is teljesül. Tehát a diagram szóban forgó részének a helyességéhez még azt kell megmutatnunk, hogy az $\bigcup LH \circ C_k$, $\bigcup C_k$, NDF és a $H \circ NDF$ osztályok által alkotott paralelogrammára is érvényesek a tartalmazási diagram szabályai, vagyis minden tartalmazás valódi, és a relációba nem állított osztályok összehasonlíthatatlanok.

Először is megállapítjuk, hogy az $LH \circ C_0$ és az NDF osztályok összehasonlíthatatlansága maga után vonja, hogy $\bigcup C_k \subset \bigcup LH \circ C_k$, $NDF \subset H \circ NDF$ és $\bigcup LH \circ C_k \not\subseteq NDF$. Továbbá, az 5.3.2 tétel szerint, $\bigcup C_k \subset NDF$. Tehát azt kell még igazolni, hogy $NDF \not\subseteq \bigcup LH \circ C_k$ és $\bigcup LH \circ C_k \subset H \circ NDF$. Mindkét állítás az alábbi lemmának lesz majd következménye, amit emellett még a továbbiakban is fel fogunk használni.

5.3.12 LEMMA. *$NDF \not\subseteq \bigcup LDF \circ H \circ C_k$.*

Bizonyítás. Azt mutatjuk meg, hogy az 5.3.2 tételben szereplő τ_A fatranszformáció nincs $LDF \circ H \circ C_k$-ban semmilyen $k \geq 0$ esetén. Tételezzük fel az ellenkezőjét vagyis azt, hogy $\tau_A \in LDF \circ H \circ C_k$. Ekkor van olyan B *ldf* transzformátor, C *h* transzformátor és $\tau \in C_k$ fatranszformáció, amelyre $\tau_A = \tau_B \circ \tau_C \circ \tau$. Az általánosság megszorítása nélkül feltehető, hogy B-nek ugyanaz az input ábécéje mint A-nak, ezért B-nek is teljesen definiáltnak kell lennie, mert A is az. Megkonstruálva B-nek és C-nek a D kompozícióját, az 5.1.1 lemma alkalmazásával azt kapjuk, hogy D teljesen definiált uniform *df* transzformátor, amelyre $\tau_D = \tau_B \circ \tau_C$. Ezért $\tau_A = \tau_D \circ \tau$ is teljesül, és az is, hogy $\mathrm{ran}(\tau_D) \subseteq \mathrm{dom}(\tau)$. Így D nemtörlő, mivel ellenkező esetben $\tau_D \circ \tau$ végtelen sok fát vinne ugyanabba a fába és ez ellentmondana annak, hogy τ_A injektív. Tehát D uniform, nemtörlő *df* transzformátor, ezért van olyan B' *lndf* transzformátor és C' *nh* transzformátor, amelyre $\tau_D = \tau_{B'} \circ \tau_{C'}$, l. 5.3.8 lemmát. De akkor $\tau_A = \tau_{B'} \circ \tau_{C'} \circ \tau$, vagyis $\tau_A \in C_{k+2}$, ami ellentmondás az 5.3.2 tétel szerint. Az ellentmondás a $\tau_A \in LDF \circ H \circ C_k$ feltételezésből adódott, tehát a lemmát bebizonyítottuk. □

5.3.13 KÖVETKEZMÉNY. *$NDF \not\subseteq \bigcup LH \circ C_k$ és $\bigcup LH \circ C_k \subset H \circ NDF$.*

Bizonyítás. Az első állítás az előző lemma következménye, mert $LH \subseteq H$, a második pedig az első miatt teljesül. □

Ezzel befejeztük annak a bizonyítását, hogy a diagramnak az $LH \circ C_0$ és $H \circ NDF$ osztályokat összekötő vonaltól jobbra eső része tartalmazási diagram. Úgy megyünk tovább, hogy az LDF és a DF osztályokat összekötő vonalat illesztjük a tartalmazási diagramhoz. A következő lemmával folytatjuk.

5.3.14 LEMMA. *LDF és $H \circ NDF$ összehasonlíthatatlanok.*

Bizonyítás. Az, hogy $H \circ NDF \not\subseteq LDF$, nyilvánvaló, mivel $NDF \not\subseteq LDF$ is teljesül. A fordított $LDF \not\subseteq H \circ NDF$ nem-tartalmazásra indirekt bizonyítást adunk. Tételezzük fel tehát, hogy $LDF \subseteq H \circ NDF$. Akkor $NH \circ LDF \subseteq NH \circ H \circ NDF$, ahonnan az 5.2.1 tételben igazolt $NH \circ LDF = DF$ és a nyilvánvaló $NH \circ H = H$ miatt $DF \subseteq H \circ NDF$. Mivel $H \circ NDF = LH \circ NDF$, amit az 5.2.1 tételben bizonyított $H = LH \circ NH$ és $NH \circ NDF = NDF$ egyenlőségekkel kapunk, $DF \subseteq LH \circ NDF$ is teljesül. Mindkét oldalt alkalmazva REC-re és kihasználva azt, hogy $LH(REC) = REC$, aminek az igazolása [GécSte] 174–175 oldalain található, kapjuk, hogy $DF(REC) \subseteq NDF(REC)$, és ez ellentmond az 5.3.3 tételnek. Tehát indirekt feltevésünk helytelen volt. □

Vegyük észre, hogy a fenti lemma biztosítja, hogy az $LH \circ C_0$ és a $H \circ NDF$ osztályokat összekötő vonalnak és az LDF és $LDF \circ NDF$ osztályokat összekötő vonalnak nem lehet közös pontja. Ezt állítjuk most explicit módon.

5.3.15 KÖVETKEZMÉNY. *$LH \circ C_1 \subset LDF$ és minden $k \geq 2$ esetén $LH \circ C_k \subset LDF \circ NH \circ C_{k-2}$. Továbbá, $\bigcup LH \circ C_k \subset \bigcup LDF \circ NH \circ C_k$ és $H \circ NDF \subset LDF \circ NDF$.*

Az 5.3.14 lemma biztosítja azt is, hogy az $LDF \circ NH \circ C_k$ osztályok egyike sem lehet részosztálya egyetlen olyan osztálynak sem, amelyik a $H \circ NDF$ osztály „alatt" van, tehát $H \circ NDF$-nek részosztálya.

Azt is meg kell mutatni, hogy az LDF és az $LDF \circ NH \circ C_k$ alakú osztályok csak a diagram által mutatott osztályokat tartalmazzák. Ez az alábbi lemmából következik majd.

5.3.16 LEMMA. *$NH \circ C_0 \not\subseteq LDF^2$ és $NH \circ C_k \not\subseteq LDF \circ H \circ C_{k-1}$ minden $k \geq 1$-re.*

Bizonyítás. Az állítás első része nyilvánvaló. A második igazolását indirekt úton végezzük. Tegyük fel, hogy $NH \circ C_k \subseteq LDF \circ H \circ C_{k-1}$. Ekkor $LDF \circ NH \circ C_k \subseteq LDF^2 \circ H \circ C_{k-1} = LDF^2 \circ NH \circ LH \circ C_{k-1}$. Alkalmazva az 5.3.9 következményben bebizonyított $LDF^2 \circ NH = LDF \circ H$ egyenlőséget és a nyilvánvaló $H \circ LH = H$-t, kapjuk, hogy $LDF \circ NH \circ C_k \subseteq LDF \circ H \circ C_{k-1}$. Ismét szorozva LDF-fel balról, az $LDF^2 \circ NH \circ C_k \subseteq LDF^2 \circ H \circ C_{k-1}$ tartalmazás adódik. Mivel $LDF^2 \circ H = LDF \circ H$ is érvényes az $LDF^2 \circ H = LDF^2 \circ NH \circ LH = LDF \circ H \circ LH = LDF \circ H$ számolás miatt, azt kaptuk, hogy $LDF \circ H \circ C_k \subseteq LDF \circ H \circ C_{k-1}$, ami viszont az 5.3.8 lemma szerint lehetetlen. Ezzel a lemma bizonyítását befejeztük. □

5.3.17 KÖVETKEZMÉNY. *$NH \circ C_0 \not\subseteq LDF$ és $NH \circ C_k \not\subseteq LDF \circ NH \circ C_{k-1}$, minden $k \geq 1$-re.*

Most tekintjük az $\bigcup LDF \circ NH \circ C_k$, az $\bigcup LH \circ C_k$, a $H \circ NDF$ és az $LDF \circ NDF$ osztályok által alkotott paralelogrammát. Erre nézve a következők teljesülnek. Az 5.3.14 lemmából következik, hogy $\bigcup LDF \circ NH \circ C_k \not\subseteq H \circ NDF$. Ezért az 5.3.15 következmény miatt már csak az $NDF \not\subseteq \bigcup LDF \circ NH \circ C_k$, $H \circ NDF \not\subseteq \bigcup LDF \circ NH \circ C_k$ nem–tartalmazásokat és az $\bigcup LDF \circ NH \circ C_k \subset LDF \circ NDF$ tartalmazást kell bizonyítani. Ehhez viszont elegendő azt észrevenni, hogy a második kettő az elsőnek a következménye, az első viszont az 5.3.12 lemmából következik. Tehát a szóban forgó paralelogramma által mutatott tartalmazások is helyesek.

Azt is meg kell mutatnunk, hogy $LDF \circ NDF \subset DF$. Ennél többet bizonyítunk.

5.3.18 LEMMA. *$LDF^2 \circ NDF \subset DF^2$.*

Bizonyítás. Az 5.2.1 tételben igazolt $DF^2 = LNDF \circ DF$ és az 5.1.1 lemmából következő $DF \circ LNDF = DF$ egyenlőségek segítségével látható, hogy $DF^3 = DF \circ LNDF \circ DF = DF^2$. Ezért az, hogy a két osztály között tartalmazás áll fenn, nyilvánvaló. Megmutatjuk, hogy egyenlőség nem lehet. Valóban, ha $LDF^2 \circ NDF = DF^2$, akkor mindkét oldalt alkalmazva REC-re és kihasználva az $LDF(REC) = REC$ egyenlőséget, l. [GécSte] 174–175 old., az adódik, hogy $NDF(REC) = DF^2(REC)$. Másrészt [Rou] I.3 tétele szerint $DF^2(REC) = DF(REC)$, tehát $NDF(REC) = DF(REC)$ is fennáll. Ez viszont ellentmond az 5.3.3 tételnek, tehát $LDF^2 \circ NDF \subset DF^2$. □

5.3.19 KÖVETKEZMÉNY. *$LDF \circ NDF \subset DF$.*

Bizonyítás. A tartalmazás az 5.1.1 lemma miatt áll fenn, másrészt egyenlőség nem lehet. Ha ugyanis $LDF \circ NDF = DF$, akkor $LDF^2 \circ NDF = LDF \circ DF$, ami ellentmondásban áll az előző lemmával, mert $LDF \circ DF = DF^2$. Ez utóbbi egyenlőség igazolását majd az 5.4.1 tételben végezzük el. □

Ezzel bizonyítottuk, hogy a diagramnak az LDF és DF osztályokat összekötő vonaltól jobbra eső része helyes.

Befejezésül az LDF^2 és a DF^2 osztályokat összekötő vonalat illesztjük a már helyesnek bebizonyított részhez. Szükségünk van a következő tényre.

5.3.20 LEMMA. *LDF^2 és DF összehasonlíthatatlanok.*

Bizonyítás. Könnyű megadni olyan DF-beli fatranszformációt, amelyik nincs LDF^2-ben. Például minden olyan DF-beli fatranszformáció jó, melynek az értékkészlete nem felismerhető erdő. Másrészt [Rou]-ban az I.2 tétel bizonyítása után szerepel két olyan fatranszformáció, amelyek LDF-ben vannak, de a kompozíciójuk nem indukálható egyetlen df transzformátorral. □

Hasonlóan mint fentebb, ez a lemma szétválasztja az LDF és DF osztályokat összekötő vonalat és az LDF^2 és DF^2 osztályokat összekötő vonalat, amit formálisan is leírunk.

5.3.21 KÖVETKEZMÉNY. *$LDF \subset LDF^2$ és minden $k \geq 0$-ra $LDF \circ NH \circ C_k \subset LDF \circ H \circ C_k$. Továbbá, $\bigcup LDF \circ NH \circ C_k \subset \bigcup LDF \circ H \circ C_k$, $LDF \circ NDF \subset LDF^2 \circ NDF$ és $DF \subset DF^2$. Végül, $LDF \circ H \circ C_k \not\subseteq LDF \circ NH \circ C_l$, semmilyen $k < l$ esetén.*

Szintén az 5.3.20 lemma következménye, hogy LDF^2 és az $LDF \circ H \circ C_k$ osztályok egyike sem lehet részosztálya egyetlen DF alatti osztálynak sem. Továbbá, az 5.3.16 lemma következménye, hogy LDF^2 és $LDF \circ H \circ C_k$ osztályok csak a diagram által mutatott osztályokat tartalmazzák.

Most tekintsük az $\bigcup LDF \circ H \circ C_k$, $\bigcup LDF \circ NH \circ C_k$, $LDF \circ NDF$ és $LDF^2 \circ NDF$ osztályok által meghatározott paralelogrammát. Az 5.3.12 lemmából és az 5.3.20 lemmából következik, hogy ez a rész is tartalmazási diagram, és az is, hogy további élek hozzávétele nélkül illeszthető a diagram már helyesnek bebizonyított részéhez. Tehát már csak egy dolog van hátra, nevezetesen annak igazolása, hogy a legfelső paralelogramma is helyes, mint tartalmazási diagram. Pontosabban, azt már láttuk az 5.3.18 lemmában és az 5.3.19 következményben, hogy $LDF^2 \circ NDF \subset DF^2$ és hogy $LDF \circ NDF \subset DF$. Továbbá, az 5.3.21 következmény és az 5.3.20 lemma értelmében $DF \subset DF^2$, $LDF \circ NDF \subset LDF^2 \circ NDF$ és $LDF^2 \circ NDF \not\subseteq DF$. Ezért már csak egyetlen nem–tartalmazás van, amit meg kell mutatnunk.

5.3.22 LEMMA. *$DF \not\subseteq LDF^2 \circ NDF$.*

Bizonyítás. Ha $DF \subseteq LDF^2 \circ NDF$ lenne, akkor mindkét oldalt alkalmazva REC-re, és kihasználva a fentebb már idézett $LDF(REC) = REC$ egyenlőséget, azt kapnánk, hogy $DF(REC) \subseteq NDF(REC)$. Ez viszont ellentmond az 5.3.3 tételnek. □

Most már kimondhatjuk azt a tételt, amelyre szükségünk van a 3. fejezet (5) pontja értelmében, és amelynek a bizonyítását az 5.3.4 lemmával kezdve, és az 5.3.22 lemmával befejezve elvégeztük.

5.3.23 TÉTEL. *A 3. ábrán látható diagram az $|N|$ halmaz tartalmazási diagramja.*

5.3.24 KÖVETKEZMÉNY. *Tetszőleges $u, u' \in N$ esetén $u = u'$ akkor és csakis akkor áll fenn, ha $|u| = |u'|$.*

Bizonyítás. L. 3. ábra. □

5.4 Algoritmus a normálforma megadására.

Következő feladatunk tehát megadni egy olyan algoritmust, amelyik tetszőleges $w \in S^*$-hoz kiszámítja azt az $u \in N$-t, amelyre $w \overset{*}{\underset{T}{\longleftrightarrow}} u$.

5.4.1 TÉTEL. *Tetszőleges $w \in S^*$ esetén effektíven megadható olyan $u \in N$, amelyre teljesül, hogy $w \overset{*}{\underset{T}{\longleftrightarrow}} u$.*

Bizonyítás. A leszálló eset bizonyításához hasonlóan először megadunk néhány további formális egyenletet, amelyek levezethetők T-ből. A most megadásra kerülő formális egyenletek után azon T-beli, vagy már korábban bevezetett egyenletek sorszámát tüntetjük fel, amelyekkel ezek levezethetők. Nézzük az elsőt részletesebben. A

$$(20) \quad H \cdot NH \doteq H \qquad (19, 15, 19)$$

írásmód azt jelenti, hogy $H \cdot NH \doteq LH \cdot NH \cdot NH$ a T-beli (19) egyenletet alkalmazva, $LH \cdot NH \cdot NH \doteq LH \cdot NH$ (15)-öt alkalmazva és végül $LH \cdot NH \doteq H$ újra (19)-et alkalmazva. Hasonló módon könnyen nyerhetők

$$\begin{array}{llll}
(21) & NH \cdot H \doteq H & (22) & H \cdot LH \doteq H \\
(23) & LH \cdot H \doteq H & (24) & H \cdot H \doteq H \\
(25) & LH \cdot NH \doteq NH \cdot LH & &
\end{array}$$

Továbbá szükségünk lesz még az alábbi egyenletekre.

$$\begin{array}{lll}
(26) & DF \cdot LNDF \doteq DF & (14, 7, 14) \\
(27) & H \cdot LDF \doteq DF & (16, 17, 14) \\
(28) & DF^3 \doteq DF^2 & (8, 26) \\
(29) & LDF^3 \doteq LDF^2 & (11, 7) \\
(30) & DF \cdot LH \doteq DF^2 & (1, 6, 28) \\
(31) & NDF \cdot H \doteq DF^2 & (19, 6, 2) \\
(32) & NDF \cdot LDF \doteq DF^2 & (5, 17, 25, 6, 14, 28) \\
(33) & NDF \cdot DF \doteq DF^2 & (14, 5, 32) \\
(34) & LDF \cdot LH \doteq LDF^2 & (7, 12, 29) \\
(35) & LDF \cdot DF \doteq DF^2 & (27, 19, 34, 12, 25, 17, 14, 8)
\end{array}$$

A szóban forgó algoritmust a 4. ábrán látható táblázat felhasználásával adjuk meg, amelyet a következőképpen kell majd használni. Mint ahogy az látható, az

$N_1 \cup N_2 \cup \{z_k \mid k \geq 1\}$ halmaz minden u eleméhez tartozik a táblázatnak egy sora ($k \geq 2$ esetén a z_{2k-1} elemekhez és a z_{2k} elemekhez ugyanazok a sorok tartoznak), és S minden Y eleméhez tartozik egy oszlopa. Az u-hoz tartozó sor és az Y-hoz tartozó oszlop találkozásánál lévő téglalapban pedig N-nek az a v eleme áll amelyre $u \cdot Y \overset{*}{\underset{T}{\longleftrightarrow}} v$. Továbbá azon T-beli és T-ből levezetett formális egyenletek sorszáma is a téglalapban van, amelyekkel a feltüntett sorrendben igazolható, hogy $u \cdot Y \overset{*}{\underset{T}{\longleftrightarrow}} v$. Ha pedig $u \cdot Y = v$, akkor rajta kívül nincs más a téglalapban.

Például az $u = DF$ és az $Y = LNDF$ esetet véve

$$DF \cdot LNDF \underset{T}{\longleftrightarrow} DF \cdot NDF \cdot LNDF \text{ az (1) alkalmazásával,}$$
$$DF \cdot NDF \cdot LNDF \underset{T}{\longleftrightarrow} DF \cdot NDF \text{ a (4) alkalmazásával}$$
$$\text{és } DF \cdot NDF \underset{T}{\longleftrightarrow} DF \text{ újra az (1) alkalmazásával.}$$

Látható, hogy megint csak az olvasóra bízzuk annak a megállapítását, hogy a szóban forgó T-beli egyenleteket melyik rész sztringre és hogy milyen irányítással alkalmazzuk. Továbbá megjegyezzük, hogy az $u = z_{2k-1}$ és az $u = z_{2k}, k \geq 2$ esetekben az alkalmazandó egyenletek sorrendje sincs feltüntetve, nem is lehet, mert k értékével növekvő hosszúságú sorozatokat kapnánk. Itt csupán az alkalmazandó T-beli és T-ből levezetett egyenletek halmazát adtuk meg, növekvő sorszám szerinti sorrendben. A sorrend megállapítását ugyancsak az olvasóra bízzuk.

Most rátérünk magának a tételnek az igazolására. A w hossza szerinti indukciót alkalmazunk.

Ha $l(w) = 0$, vagyis $w = \lambda$, akkor legyen $u = \lambda$.

Ha $l(w) > 0$, tehát $w = x \cdot Y$ valamely $x \in S^*$ és $Y \in S$ esetén, akkor pedig tegyük a következőket. Konstruáljuk meg a táblázat segítségével azt a $v \in N$ normálformát, amelyre $x \overset{*}{\underset{T}{\longleftrightarrow}} v$. (Megjegyezzük, hogy ez az indukció feltevés értelmében elvégezhető.) Két eset lehetséges, amelyeket az alábbiakban részletezünk.

(a) $v \in N_1 \cup N_2 \cup \{z_k \mid k \geq 1\}$. Ekkor a táblázatból a fent leírt módon kiolvasható az az $u \in N$, amelyre $v \cdot Y \overset{*}{\underset{T}{\longleftrightarrow}} u$. Így u lesz a keresett normálforma, mert $w = x \cdot Y \overset{*}{\underset{T}{\longleftrightarrow}} v \cdot Y \overset{*}{\underset{T}{\longleftrightarrow}} u$.

(b) $v \in \{z \cdot z_k \mid z \in N_2, k \geq 1\}$, tehát $v = z \cdot z_k$ valamely $z \in N_2$ és $k \geq 1$ esetén. Ebben az esetben azt az u normálformát keressük, amelyre $z \cdot z_k \cdot Y \overset{*}{\underset{T}{\longleftrightarrow}} u$. Először a táblázat segítségével konstruáljuk meg azt a $v' \in N$-t, amelyre $z_k \cdot Y \overset{*}{\underset{T}{\longleftrightarrow}} v'$. Ekkor, a táblázat értelmében, ismét két eset állhat fenn.

(b1) $v' = z_{k+1}$. Ekkor $u = z \cdot z_{k+1}$, mert $w = x \cdot Y \overset{*}{\underset{T}{\longleftrightarrow}} v \cdot Y = z \cdot z_k \cdot Y \overset{*}{\underset{T}{\longleftrightarrow}} z \cdot v' = z \cdot z_{k+1}$.

(b2) $v' \in N_1 \cup N_2$. Ekkor pedig, mivel $l(v') \leq 3$, a táblázat legfeljebb háromszori alkalmazásával megkapjuk azt az $u \in N$-t, amelyre $z \cdot v' \underset{T}{\longleftrightarrow} u$. Ez az u megfelelő

	DF	NDF	LDF	$LNDF$	H	NH	LH
DF	DF^2	1 DF	2, 14 DF^2	1, 4, 1 DF	19, 30, 2 DF^2	2 DF	30 DF^2
LDF	35 DF^2	$LDF \cdot NDF$	LDF^2	7 LDF	$LDF \cdot H$	$LDF \cdot NH$	34 LDF^2
NDF	33 DF^2	3 NDF	32 DF^2	4 NDF	31 DF^2	5 NDF	6 DF^2
$z_1 = LNDF$	8 DF^2	9 NDF	11 LDF^2	10 $LNDF$	19, 12, 34, 19 $LDF \cdot H$	z_2	12 LDF^2
DF^2	28 DF^2	1 DF^2	2, 14, 28 DF^2	26 DF^2	16, 2, 30, 28 DF^2	2 DF^2	30, 28 DF^2
$LDF \cdot NDF$	33, 35, 28 DF^2	3 $LDF \cdot NDF$	32, 35, 28 DF^2	4 $LDF \cdot NDF$	31, 35, 28 DF^2	5 $LDF \cdot NDF$	6, 35, 28 DF^2
LDF^2	35, 35, 28 DF^2	$LDF^2 \cdot NDF$	29 LDF^2	7 LDF^2	34, 23 $LDF \cdot H$	34, 19 $LDF \cdot H$	34, 29 LDF^2
$H \cdot NDF$	33, 14, 20, 27 DF^2	3 $H \cdot NDF$	32, 14, 20, 27 DF^2	4 $H \cdot NDF$	31, 14, 20, 27 DF^2	5 $H \cdot NDF$	6, 14, 20, 27 DF^2
$LDF^2 \cdot NDF$	33, 35, 35, 28 28 DF^2	3 $LDF^2 \cdot NDF$	32, 35, 35, 28 28 DF^2	4 $LDF^2 \cdot NDF$	31, 35, 35, 28 28 DF^2	5 $LDF^2 \cdot NDF$	6, 35, 35, 28 28 DF^2

4/a. ábra

	DF	NDF	LDF	$LNDF$	H	NH	LH
H	14, 20, 27 DF	$H \cdot NDF$	27 DF	$H \cdot z_1$	24 H	20 H	22 H
NH	14, 15, 14 DF	13 NDF	14 DF	$NH \cdot z_1$	21 H	15 NH	16 H
LH	14, 19, 27 DF	13, 19 $H \cdot NDF$	17 LDF	$LH \cdot z_1$	23 H	19 H	18 LH
$LDF \cdot NH$	14, 15, 14, 35 DF^2	13 $LDF \cdot NDF$	14, 35 DF^2	$LDF \cdot NH \cdot z_1$	21 $LDF \cdot H$	15 $LDF \cdot NH$	16 $LDF \cdot H$
$LDF \cdot H$	14, 20, 27, 35 DF^2	19, 34, 13 $LDF^2 \cdot NDF$	27, 35 DF^2	$LDF \cdot H \cdot z_1$	24 $LDF \cdot H$	20 $LDF \cdot H$	22 $LDF \cdot H$
z_2	14, 15, 14, 8 DF^2	13, 9 NDF	14, 8 DF^2	z_3	21, 19, 12, 34, 19 $LDF \cdot H$	15 z_2	16, 19, 12, 34, 19 $LDF \cdot H$
$z_{2k-1}, k \geq 2$	8, 14, 15, 28 DF^2	13, 9 NDF	2, 8, 11, 14 15, 28 DF^2	10 z_{2k-1}	2, 8, 12, 14 15, 19, 28 DF^2	z_{2k}	2, 8, 12, 14 15, 28 DF^2
$z_{2k}, k \geq 2$	8, 14, 15, 28 DF^2	13, 9 NDF	8, 14, 15, 28 DF^2	z_{2k+1}	2, 8, 12, 14 15, 19, 21, 28 DF^2	15 z_{2k}	2, 8, 12, 14, 16 19, 28, 30, 34, 15 DF^2

4/b. ábra

lesz, mivel $w = x \cdot Y \underset{T}{\overset{*}{\longleftrightarrow}} v \cdot Y = z \cdot z_k \cdot Y \underset{T}{\overset{*}{\longleftrightarrow}} z \cdot v' \underset{T}{\overset{*}{\longleftrightarrow}} u$.

Ezzel a tétel bizonyítását befejeztük. □

Egyúttal a 3. fejezet (1)–(5) pontjaiban kitűzött feladatokat is megoldottuk az S halmaz esetében, tehát alkalmazhatjuk a 3.1 állítást.

Végezetül ezen állításban szereplő algoritmus működését szemléltetjük egy példán.

Legyen a két, S feletti kifejezés

$$LNDF \circ LH \circ NH \circ LNDF \quad \text{és} \quad LH \circ LNDF \circ H.$$

Az 5.4.1 tételben megadott algoritmus alkalmazásával kapjuk, hogy $LNDF \cdot LH \underset{T}{\longleftrightarrow} LDF^2$, $LDF^2 \cdot NH \underset{T}{\overset{*}{\longleftrightarrow}} LDF \cdot H$ és hogy $LDF \cdot H \cdot LNDF = LDF \cdot H \cdot z_1$ már normálforma. Hasonlóan, a másik szóra az adódik, hogy $LH \cdot LNDF \in N$ és $LNDF \cdot H \underset{T}{\overset{*}{\longleftrightarrow}} LDF \cdot H$. Most újra a táblázat szerint, l. (b2) eset, $LH \cdot LDF \underset{T}{\longleftrightarrow} LDF$ és $LDF \cdot H$ már normálforma, tehát $LH \cdot LNDF \cdot H \underset{T}{\overset{*}{\longleftrightarrow}} LDF \cdot H$. Ezért, a 3. ábrán lévő táblázat szerint a

$$LH \circ LNDF \circ H \subset LNDF \circ LH \circ NH \circ LNDF$$

tartalmazás áll fenn.

5.5 A T-vel ekvivalens, teljes R sztring átíró rendszer megadása.

Először rámutatunk arra, hogy maga a T nem *Church–Rosser tulajdonságú*. Valóban, hiszen nem is konfluens, mert az $NDF \cdot NH \cdot NDF$ szóból kiindulva azt kapjuk, hogy $NDF \cdot NH \cdot LDF \underset{T}{\longrightarrow} NDF \cdot LDF$ és $NDF \cdot NH \cdot LDF \underset{T}{\longrightarrow} NDF \cdot DF$, de nincsen olyan $x \in S^*$ szó, amelyre $NDF \cdot LDF \underset{T}{\overset{*}{\longrightarrow}} x$ és $NDF \cdot DF \underset{T}{\overset{*}{\longrightarrow}} x$.

Most, csakúgy, mint a leszálló esetben, megadunk egy R sztring átíró rendszert. Tekintsük evégett a 4. ábrán lévő táblázatot. Az utolsó három sora kivételével minden u sorához és Y oszlopához rendeljük hozzá az $(u \cdot Y, v)$ párt, ahol v az u sor és Y oszlop találkozásánál lévő elem. (Például az $u = LDF \cdot NDF$ és az $Y = LDF$ esethez az $(LDF \cdot NDF \cdot LDF,\ DF^2)$ pár tartozik.) Legyen R az összes ilyen párokból álló sztring átíró rendszer, amely párokra még az is teljesül, hogy $u \cdot Y \neq v$. (A z_1 helyett természetesen mindenütt $LNDF$-et írunk.) Adjuk meg továbbá S elemeinek az alábbi súlyozását: $\rho(DF) = 1$, $\rho(LDF) = 2$, $\rho(NDF) = 3$, $\rho(LNDF) = 4$, $\rho(H) = 5$, $\rho(LH) = 6$, $\rho(NH) = 7$. Ekkor érvényes lesz a következő tétel, amelynek bizonyítása egyébként teljesen hasonló a leszálló eset bizonyításához, és a fenti konstrukcióval együtt teljes egészében megtalálható [FülVág7]-ben.

5.5.1 TÉTEL. *R súly csökkentő,* $\underset{T}{\overset{*}{\longleftrightarrow}} = \underset{R}{\overset{*}{\longleftrightarrow}}$ *és* $N = IRR(R)$.

Bizonyítás. L. [FülVág7]. □

Csakúgy, mint a leszálló esetben, most is azt kapjuk, hogy az 5.4.1 tételben megadott algoritmus helyett alkalmazhatjuk a 3.5 tételben szereplő algoritmust a normálforma kiszámítására, így N és R megint csak elegendő az $[S]$ monoid teljes leírására.

6. Összefoglalás

Jelen dolgozatban kifejlesztettünk egy olyan eljárást, amellyel fatranszformáció osztályok egy véges, de — mint ahogy azt a bevezetésben is írtuk — nem túl általános P halmazához olyan algoritmus adható meg, amellyel a P által kompozícióval generált $[P]$ monoid tetszőleges két eleméről eldönthető, hogy közöttük fennáll-e a tartalmazási reláció. Ezáltal a P halmaz elemeire vonatkozó összes kompozíciós és dekompozíciós eredményt megkapjuk szemben az eddigi kutatásokkal, amelyek az ilyen eredményeket csak véletlenszerűen keresték és találták meg.

Ha módszerünket alkalmazni akarjuk, akkor előre el kell döntenünk azt, hogy mely fatranszformáció osztályok közötti összes kompozíciós és dekompozíciós eredményt akarjuk megkapni és ezen osztályokból kell a P halmazt megalkotni. Ez az a pont, ahol körültekintőnek kell lenni. Amennyiben túl sok eleme lesz a P-nek vagy — kissé pontatlanul fogalmazva — a kompozíció csak kevés elemet „ejt össze" úgy a normálformák halmazának megadásánál, de még inkább az ezek által reprezentált fatranszformáció osztályok tartalmazási diagramjának megadásánál nehézségekbe ütközünk. Nem tudható előre, hogy a normálformák halmaza megadható-e valamilyen rekurzív módszerrel (mint ahogy az megadható volt a felszálló esetben), hogy a tartalmazási diagram felállítható-e, bebizonyítható-e, és végül, hogy a monoid szorzótáblák milyen módon vezethetők vissza a véges esetre (mint a felszálló esetben). A szerző véleménye szerint ezen problémákkal találnánk magunkat szemben, ha például az értekezésben vizsgált determinisztikus osztályok helyett ezen osztályok nem determinisztikus megfelelőit vizsgálnánk, mely észrevételünket J. ENGELFRIETnek a már a bevezetésben is említett híres hierarchia tételei támasztják alá [Eng6].

Mindazonáltal a 4. és 5. fejezetek eredményeit elegendő evidenciának érezzük ahhoz, hogy módszerünk elég széles körben lehet eredményes. Érdekes és — szerintünk — eredményes lenne olyan P osztályokból kiindulni, amelyek az igen jól alkalmazható ún. átcímkéző fatranszformációk osztályát [GécSte] is tartalmazzák. Ugyancsak megoldhatónak tűnik a feladat a [FülVág4] dolgozatban bevezetett ún. reguláris előrenézésű fatranszformációosztályok esetén, különösen a [FülVág9]-ben igazolt hierarchia tétel ismeretében.

IRODALOM

[ArnDau] ARNOLD, A. and DAUCHET, M., "Morphismes et biomorphismes d'arbes", *Theoret. Comput. Sci.* **20** (1982), 33–93.

[Bak] BAKER, B. S., "Composition of top-down and bottom-up tree transducers", *Inform. and Control* **41** (1979), 186–213.

[Bar1] BARTHA, M., *An Algebraic definition of attributed transformations*, Proc. of the 1981 FCT Conference (Szeged, Hungary, Aug. 24–28), Lecture Notes in Computer Science, vol. 117 (Springer–Verlag, Berlin, 1981).

[Bar2] BARTHA, M., "Linear deterministic attributed transformations", *Acta Cybernet.* **6** (1983), 125–147.

[BKR] BENINGHOFEN, B., KEMMERICH, S. and RICHTER, M. M., *Systems of reductions*, Lecture Notes in Computer Science 277 (Springer–Verlag, Berlin, 1987).

[Boo1] BOOK, R. V., "Confluent and other types of Thue systems", *J. Assoc. Comput. Mach.* **29** (1982), 171–182.

[Boo2] BOOK, R. V., "Thue systems and the Church–Rosser property: replacement systems, specification of formal languages and presentations of monoids", in Progress in combinatorics on words (L. Cummings szerk.) (Academic Press, New York, 1983), pp. 1–38.

[BooODu] BOOK, V. and O'DUNLAING, P., "Testing for the Church–Rosser Property", *Theoret. Comput. Sci.* **16** (1981), 223–229.

[BurSan] BURRIS, S. and SANKAPPANAVAR, H. P., *A Course in Universal Algebra* (Springer–Verlag, New York, 1981).

[CouFra] COURCELLE, B. and FRANCHI-ZANNETTACCI, P., "Attribute grammars and recursive program schemes, Part I and II", *Theoret. Comput. Sci.* **17** (1982), 163–191; 235–257.

[Dau] DAUCHET, M, *Transductions de forets, bimorphisms de magmoides*, Thése de doctorat (University of Lille, France, 1977).

[DerJou] DERSHOWITZ, N., JOUANNAUD, J. P., "Rewrite Systems", in: *Handbook of Theoretical Computer Science* (North–Holland, Amsterdam, 1989).

[DTHL] DAUCHET, M., TISON, S., HEULLIARD, J. and LESCANNE, P., *Decidability of the confluence of ground term rewriting systems*, Technical report Nr 102 (University of Lille, France, 1987); *see also in Proc. 2nd IEEE Symp. on Logic in Computer Science* (Itacha, New York, 1987).

[Eng1] ENGELFRIET, J., "Bottom-up and top-down tree transformations — a comparison", *Math. Systems Theory* **9** (1975), 198–231.

[Eng2] ENGELFRIET, J., "Top-down tree transducers with regular look-ahead", *Math. Systems Theory* **10** (1977), 289–303.

[Eng3] ENGELFRIET, J., "On tree transducers for partial functions", *Inform. Process. Lett.* **7** (1978), 170–172.

[Eng4] ENGELFRIET, J., *Some open questions and recent results on tree transducers and tree languages*, in "Formal Language Theory; Perspectives and Open Problems" (R.V. Book szerk.) (Academic Press, New York, 1980).

[Eng5] ENGELFRIET, J., *Tree transducers and syntax-directed semantics, in Proc. of the 7th CAAP* (Lille, 1982), pp. 82–107.

[Eng6] ENGELFRIET, J., "Three hierarchies of transducers", *Math. Systems Theory* **15** (1982), 95–125.

[EngFil] ENGELFRIET, J. and FILE, G., "The formal power of one-visit attribute grammars", *Acta Inform.* **16** (1981), 275–302.

[EngVog1] ENGELFRIET, J. and VOGLER, H., "Macro tree transducers", *J. Comput. System Sci.* **31** (1985), 71–146.

[EngVog2] ENGELFRIET, J. and VOGLER, H., "Pushdown machines for the macro tree transducer", *Theoret. Comput. Sci.* **42** (1986), 251–368.

[EngVog3] ENGELFRIET, J. and VOGLER, H., "Modular tree transducers", *Aachener–Informatik Berichte*, Nr. 88–22; also *J. Comput. System Sci.* **31** (1985), 71–146.

[EngVog4] ENGELFRIET, J. and VOGLER, H., "High level tree transducers and iterated pushdown transducers", *Acta Inform.* **26** (1988), 131–192.

[Fül1] FÜLÖP, Z., "On Attributed tree transducers", *Acta Cybernet.* **5** (1981), 261–279.

[Fül2] FÜLÖP, Z., "Decomposition results concerning K–visit attributed tree transducers", *Acta Cybernet.* **6** (1983), 163–171.

[Fül3] FÜLÖP, Z., "A complete description for a monoid of deterministic bottom-up tree transformation classes", *Theoret. Comput. Sci.* 88 (1991), 253–268.

[FülVág1] FÜLÖP, Z. and VÁGVÖLGYI, S., "Results on compositions of deterministic root-to-frontier tree transfromations", *Acta Cybernet.* 8 (1987), 49–61.

[FülVág2] FÜLÖP, Z. and VÁGVÖLGYI, S., "On ranges of compositions of deterministic root-to-frontier tree transformations", *Acta Cybernet.* 8 (1988), 259–266.

[FülVág3] FÜLÖP, Z. and VÁGVÖLGYI, S., "Top-down tree transducers with deterministic top-down look-ahead", *Inform. Process. Lett.* 33 (1989/90), 3–5.

[FülVág4] FÜLÖP, Z. and VÁGVÖLGYI, S., "Variants of top-down tree transducers with look-ahead", *Math. Systems Theory* 21 (1989), 125–145.

[FülVág5] FÜLÖP, Z. and VÁGVÖLGYI, S., "A complete classification of deterministic root-to-frontier tree transfromation classes", *Theoret. Comput. Sci.* 81 (1991), 1–15.

[FülVág6] FÜLÖP, Z. and VÁGVÖLGYI, S., "A finite presentation for a monoid of tree transformation classes", *in Proc. 2nd Conference on Automata. Languages and Programming Systems* (Gécseg F. és Peák I. szerk.), *Salgótarján* (1988).

[FülVág7] FÜLÖP, Z. and VÁGVÖLGYI, S., "A complete term rewriting system for a monoid of tree transformation classes", *Inform. and Computation* 86 (1990), 195–212.

[FülVág8] FÜLÖP, Z. and VÁGVÖLGYI, S., "Ground term rewriting rules for the word problem of a ground term equation system", *Bulletin of the EATCS* 45 (1991), 186–201.

[FülVág9] FÜLÖP, Z. and VÁGVÖLGYI, S., "Iterated deterministic top-down look-ahead", Proc. of the 1989 FCT Conference (Szeged, Aug. 21–25), *Lectures Notes in Computer Sciences*, vol. 380 (Springer–Verlag, Berlin, 1989).

[GécSte] GÉCSEG, F. and STEINBY, M., *Tree automata* (Akadémiai Kiadó, Budapest, 1984).

[Hue] HUET, G., "Confluent reductions: Abstract properties and applications to term rewriting systems", *Journal of the ACM* 27 (1980), 797–821.

[Jan] JANTZEN, M., *Confluent String Rewriting* (Springer–Verlag, Berlin–Heidelberg–New York, 1988).

[Knu] KNUTH, D. E., "Semantics of context-free languages", *Math. Syst. Theory* 2 (1968), 127–145.

[Rou] ROUNDS, W. C., "Mappings and grammars on trees", *Math. Systems Theory* 4 (1970), 257–287.

[Tha1] THATCHER, J. W., "Generalized^2 sequential machine maps", *J. Comput. System Sci.* 4 (1970), 339–367.

[Tha2] THATCHER, J. W., *Tree automata: an informal survey*, Currents in the Theory of Computing (A.V. Aho szerk.) (Prentice-Hall, Englewood Cliffs, 1973).

[Vág] VÁGVÖLGYI, S., "On compositions of root-to-frontier tree transformations", *Acta Cybernet.* 7 (1986), 443–480.

[VágFül] VÁGVÖLGYI, S. and FÜLÖP, Z., "An infinite hierarchy of tree transformations in the class NDR", *Acta Cybernet.* 8 (1987), 153–168.

[Vog1] VOGLER, H., "Basic tree transducers", *J. Comput. System Sci.* 34 (1987), 87–128.

[Vog2] VOGLER, H., "High level modular tree transducers", *in: "1978–1988, Ten Years IIG, Institutes for Information Processing"* (H.J. Pongratz és W. Schinnerl szerk.), Report 260, Institutes for Information Processing, Graz University of Technology.

(Beérkezett: 1991. január 18.)

FÜLÖP ZOLTÁN
MTA AUTOMATAELMÉLETI TANSZÉKI KUTATÓCSOPORT
6720 SZEGED, ARADI VÉRTANUK TERE 1.

DECIABILITY QUESTIONS IN MONOIDS GENERATED BY TREE TRANSFORMATION CLASSES

Z. FÜLÖP

In this paper monoids generated by deterministic top-down and deterministic bottom-up tree transformation classes with composition are considered. A procedure is described, which, given a finite and not too general set P of such tree transformation classes, yields an algorithm with the following property. For any two elements $X_1 \circ \cdots \circ X_m$ and $Y_1 \circ \cdots \circ Y_n$ of the monoid generated by P with composition $\circ$, the algorithm can decide if the inclusion

$$X_1 \circ \cdots \circ X_m \subseteq Y_1 \circ \cdots \circ Y_n$$

holds or not.

The procedure is applied to two particular choices of P, namely to the sets

$$M = \{DL, LDL, NDL, LNDL, H, LH, NH\}$$

and

$$S = \{DF, LDF, NDF, LNDF, H, LH, NH\}$$

where the firs set consists of the class of deterministic bottom-up tree transformations and its six fundamental subclasses; and the second one consists of the top-down counterparts of the elements of the first one.

Alkalmazott Matematikai Lapok 15 (1990–91) 267–287

ABSZTRAKT ADATTÍPUSOK MEGVALÓSÍTÁSÁNAK LEHETŐSÉGEI KÜLÖNBÖZŐ PROGRAMOZÁSI NYELVEKBEN

KOZICS SÁNDOR

Budapest

A programok megbízhatósága szempontjából központi helyet foglal el a típus fogalma. A dolgozatban azt a folyamatot kísérjük végig, ahogyan kialakultak a nyelvekben a fogalom legfontosabb tulajdonságai. Az elemzést a SIMULA-val kezdjük, ahol a típus még csak az eljárások csoportosításának az eszköze, s az Adáig folytatjuk, ahol már absztrakt típusosztály definiálására is van mód.

1. Bevezetés

A programozási nyelvek fejlődési irányát az utóbbi húsz évben elsősorban az jellemzi, hogy egyre hatékonyabb nyelvi eszközökkel támogatják a megbízható programok létrehozását. A megbízhatóság növelésében alapvető szerepet játszik az absztrakció. A nyelvi eszközök az absztrakt fogalom definiálásának és megvalósításának a leírását szolgálják. Ebben a dolgozatban a programozási nyelveknek ezt a fejlődési irányát kívánjuk megvizsgálni az absztrakt adattípusok szempontjából.

Akkor mondhatjuk, hogy a programozási nyelv támogatja az absztrakt típusok kezelését, ha teljesíti az alábbi két feltételt [5]:

a) A nyelv programegységet biztosít a megvalósítás keretéül. A megvalósítás magába foglalja a típus reprezentációját és a típusműveletek implementációját.

b) A nyelv a típusműveletekre korlátozza a reprezentációhoz való hozzáférést.

A továbbiakban végigkísérjük a nyelvek fejlődését a 60-as évek végétől kezdve (a SIMULA 67–től az Adáig). A jellegzetességeket példákon mutatjuk be, példáinkban a heap típussal és ennek különféle reprezentációival foglalkozunk. A heap típust KNUTH [1] definiálta, mi lényegében ezt a definíciót fogadjuk el. Vizsgálódásunkhoz az elméleti hátteret olyan, absztrakt típusokkal foglalkozó írások adják, mint pl. [5], [7] és [8].

2. A heap típus

A heap egyetlen komponens típust tartalmaz. Egy heap érték ebből a típusból, az ún. alaptípusból vett értékek sorozata. A heap alaptípusa csak olyan típus lehet, amelyen értelmezve van egy rendezési reláció. A típushoz öt műveletet definiálunk:

create: → Heap
push: Heap × Basetype → Heap
top: Heap → Basetype
pop: Heap → Heap
empty: Heap → Boolean

A típus sajátossága, hogy a top és a pop művelet mindig a sorozat legnagyobb elemét választja le. Jelentését tehát algebrai specifikációs módszerrel a következőképpen adhatjuk meg:

```
empty(create) = true
empty(push(h,x)) = false
top(create) = error
pop(create) = error
push(pop(h), top(h)) = h
pop(push(h,x)) = if empty(h) or top(h) ≤ x then
h
else
push(pop(h), x)
top(push(h,x)) = if empty (h) or top(h) ≤ x then
x
else
top(h)
```

3. SIMULA 67

A SIMULA 67 tervezői, O.-J. DAHL és csoportja, az ALGOL 60–at választották kiindulási pontként az új nyelv kialakításakor. Központi fogalomnak, a dekompozíció elsődleges eszközének a blokk fogalmát fogadták el. [2]

A blokk adatszerkezeteket és hozzájuk kapcsolódó alprogramokat foglal magába. A blokkban definiált mennyiségek lokálisak, azaz teljesen függetlenek a program többi részétől. A blokk maga csak egy minta, amelynek egy példánya a blokk végrehajtásakor jön létre. A példány létrejötte annak a memóriaterületnek a lefoglalását jelenti, ami a blokk lokális változóinak és a végrehajtásához feljegyzendő információnak a tárolásához szükséges.

Egy ALGOL 60 program végrehajtása a blokkok szemszögéből az egymás végrehajtását kezdeményező blokkpéldányok sorozatából áll. A SIMULA 67 a blokk fogalmának általánosításával ennél bonyolultabb kölcsönhatási mechanizmust hoz létre, amelyben a blokkpéldányokból már tetszőleges listaszerkezeteket lehet kialakítani.

3.1. A class fogalma

A class egy paraméterezhető blokk, ami változók és alprogramok gyűjteménye. Az osztálydefiníció voltaképpen egy minta (típus!), amiből egy erre szolgáló művelettel (generálással) állíthatók elő az osztályba tartozó példányok, vagyis az objektumok.

Az alábbi SZEMELY osztály egy egyszerű class, ami tevékenységet nem is tartalmaz

```
class SZEMELY (N); integer N;
begin
    character array (1:N) NEV;
    integer ADAT;
end;
```

Új példányt a new generátor segítségével hozhatunk létre.

```
new SZEMELY (24)
```

Ehhez hasonlóan tetszőleges számú SZEMELY objektum is előállítható.

Azért, hogy az egyes példányokra külön-külön is hivatkozni lehessen, a nyelv bevezeti a hivatkozás típusú (azaz pointer) változót.

```
ref (SZEMELY) S;
```

Így S egy olyan változó, aminek értékül adható a SZEMELY osztály egy példányának címe.

```
S :-- new SZEMELY (24);
```

(A :- az értékadásnak a hivatkozási értékekhez tartozó formája.) Az osztályon kívül S.NEV formában hivatkozhatunk az osztály komponenseire.

Ha egy osztálydeklarációban tevékenységek is szerepelnek, akkor azokat az osztályhoz tartozó valamennyi objektum végrehajthatja.

```
class PONT (X, Y); real X, Y;
begin
    ref (PONT) procedure TAVOLSAG (P) REF (PONT) P;
        TAVOLSAG :- new
            PONT(SQRT((X-P.X)**2+(Y-P.Y)**2));
end PONT;
ref(PONT) Q, R;
```

Így az R pontnak a Q ponttól való távolságát a Q.TAVOLSAG (R) kifejezés adja meg.

Ha azt akarjuk elérni, hogy az osztály valamely komponensei (tehát változói vagy alprogramjai) az osztályon kívül ne legyenek láthatóak, úgy ezeket a protected hidden minősítéssel kell ellátni.

A SIMULA 67 class fogalmát tekinthetjük a legkorábbi olyan nyelvi eszköznek, amely típusmegvalósítás leírására alkalmas.

3.2. Példa

Az alábbi példában a heap–et egy H(1..SIZE) vektorral és egy N egésszel ábrázoljuk. N a heap elemeinek száma, N=0 jelenti az üres heap–et. A sorozatot a H(1), ... ,H(N) elemekkel fejezzük ki. A vektorhoz hozzárendeljük még a

$$\forall i \in [1,N] \; : \; H(1) \geq H(i)$$

invariánst. Így a H(1) érték a vektor legnagyobb eleme. Egy új elem beillesztése olcsón, egy elem kivétele azonban csak N-nel arányos költséggel végezhető el.

```
class HEAP (SIZE); integer  SIZE;
    protected hidden N, H;
begin
    integer N;
    integer array (1:SIZE) H;
    procedure PUSH (X); value X; integer X;
    begin
        if N = SIZE then TERMINATE_PROGRAM;
        N:= N+1;
        if  X > H(1) then
            begin H(N) := H(1); H(1) := X; end
        else
            H(N) := X;
    end PUSH;

    integer procedure TOP;
    begin
        if N = 0 then TERMINATE_PROGRAM;
        TOP := H(1);
    end TOP;

    procedure POP;
    begin
        integer I, M;
        if N = 0 then TERMINATE_PROGRAM;
        M := N;
        for I := 2 step 1 until N do
            if H(I) > H(M) then M := I;
        H(1) := H(M); H(M) := H(N); N := N - 1;
    end POP;

    Boolean procedure EMPTY;
    begin
        EMPTY := (N = 0);
    end EMPTY;
    N := 0;
    end HEAP;
```

Példánkban a típusból a CREATE műveletet elhagytuk, mert az más nyelvi utasításokkal helyettesíthető. A SIMULA 67 nyelvben egy új heap bevezetése két lépésben történik: definiálni kell egy pointer-változót, amivel a létrehozandó heap-et azonosítani fogjuk,

```
ref (HEAP) H;
```

majd létre kell hozni egy heap-et a new generátor segítségével, aminek „címét" a H változó kapja értékül (itt paraméterként megadható ennek heap-nek a maximális mérete).

```
H :- new HEAP (10);
```

Azt, hogy egy típusművelet melyik heap-re vonatkozik úgy választjuk ki, hogy a kívánt heap-nek a típusműveletét használjuk.

```
integer Z;
H.PUSH(Z);
Z := H.TOP;
```

Minden egyes generálás alkalmával, az új példányban azonnal végrehajtódik a deklarációk után álló utasítássorozat (itt N := 0;). Ez a tulajdonság felhasználható arra, hogy egy új heap-változó bevezetésekor azt inicializáljuk. A programok megbízhatósága szempontjából ez egy igen pozitív vonása a class-nak.

A class típussal nem paraméterezhető.

4. MODULA-2

Amikor N. WIRTH 1977-ben elvállalta, hogy a tervezés alatt álló *Lilith személyi számítógépet* felszereli a szükséges software-rel, nem utolsósorban annak bizonyítására, hogy alap-software is készíthető magasszintű nyelven, úgy döntött, hogy a teljes operációs rendszert a kiszolgáló programokkal együtt ugyanazon a magasszintű nyelven készíti el. [3]

Az alkalmazási céloknak megfelelően a tervezett nyelvnek alkalmasnak kellett lennie algoritmus, azaz bizonyos fokú absztrakció leírására, ugyanakkor az operációs rendszer magja (pl. a perifériakezelő driver-ek) megírásához a hardware regiszterek közvetlen elérését is lehetővé kellett tennie. A Pascal nyilvánvalóan nem felelt meg erre a célra, és nem látszott elegendőnek a nyelv egyszerű kibővítése sem. A WIRTH által néhány évvel korábban megalkotott Modula túlságosan primitív nyelv volt, erre a célra közvetlenül szintén nem használható, a modul-fogalma viszont igen alkalmasnak látszott a magasszintű nyelvben a gépközeli információk elrejtésére. Az új nyelv, a Modula-2 ebből a két nyelvből született a Pascal szintaxisának némi javításával és a Modula néhány fogalmának átvételével.

4.1. A modul

A Modul-2 nemcsak lehetővé teszi a program feldarabolását és a részenkénti fordítást, hanem a Modula mintájára eszközöket ad a modulok közötti kapcsolatok leírására is. A nyelvek között elsőként nem triviálisan használja a láthatóság fogalmát: a modul definíciós és implementációs részre válik szét. A definíciós modul tartalmazza a más fordítási egységekből elérhető típusok, objektumok és alprogramok specifikációit, az implementációs modulban pedig megtalálható ezek megvalósítása, törzse. Az implementációs rész rejtve marad a többi modul előtt. Ezzel lehetővé válik absztrakt típusok definiálása, a típus specifikációjának ill. reprezentációjának és implementációjának szétválasztása is.

Azért, hogy a fordítóprogram ellenőrizni tudja a modulok közötti kapcsolatokat, a lefordított modulokat egy adatbázisban (könyvtárban) helyezi el, a modulok fordítási sorrendjére pedig az alábbi megszorításokat teszi:

1) Egy modul definíciós részét előbb kell fordítani, mint a hozzá tartozó implementációs részt.
2) Ha az A definíciós vagy implementációs rész hivatkozik egy B modulban definiált névre, akkor a B modul definíciós részét előbb kell fordítani, mint a hivatkozó A részt.
3) Ha egy definíciós részt újra lefordítunk, a könyvtárban érvénytelenné válik minden olyan definíciós és implementációs rész fordítása, amely — közvetlenül vagy közvetve — az újrafordított definíciós részre hivatkozik.

4.2. Az opaque típus

A MODULA-2 tervezésekor célul tűzték ki nyelvi elemek definiálását a típusmegvalósítás leírására. Ez az ún. átlátszatlan (opaque) típus. Az átlátszatlan típus specifikációja a definíciós modulban, reprezentációja és a típusműveletek implementációja az ehhez tartozó implementációs modulban van. Mivel az implementációs modul módosításakor nem szükséges újrafordítani a típust felhasználó modulokat (hiszen a típus specifikációja nem változott!), ez a megoldás egy biztonságos és könnyen alakítható programot eredményez.

A MODULA-2 tartalmaz egy szigorú megkötést: az átlátszatlan típus reprezentációja mindig a `WORD` típussal kompatibilis típus kell legyen. Erre a szabályra a nyelvben azért van szükség, mert a definíciós modul semmiféle információt sem tartalmaz a típus reprezentációjáról, így azt sem lehet tudni, hogy egy ilyen típusú változó definiálásakor az mekkora tárolóterületet foglal el. Az információt tartalmazó implementációs modul viszont a típus felhasználásakor még nem áll a fordítóprogram rendelkezésére (esetleg el sem készült). Ennek a szabálynak az alapján a fordítóprogram minden átlátszatlan típusú változónak annyi helyet foglal le, amennyi a `WORD` típushoz szükséges. (Pl. az `INTEGER` és a `POINTER` típus kompatibilis a `WORD` -del.)

Az átlátszatlan típus a definiálása után ugyanúgy használható, mint bármely predefinit típus, mint pl. az `INTEGER` vagy `REAL`. Átlátszatlan típus esetén a típus értékeire csak a típusssal egy modulban definiált típusműveletek, valamint az értékadás és az egyenlőség vizsgálata megengedett. Kimondottan hiba a megbízhatóság szempontjából, hogy MODULA-2 ez utóbbi kettőt is megengedi. Mivel ugyanis pointerrel való reprezentálás esetén ez kizárólag csak a pointert adja értékül a változónak ill. csak a pointerek egyezését vizsgálja, ezért félrevezető is (hiszen a predefinit típusoknál nem így van) és haszontalan is (hiszen erre a lehető legritkább esetben lehet szükség).

4.3. Példa

A példaként használt heap típus definíciós modulját tehát megadhatjuk még mielőtt a reprezentációról döntöttünk volna.

```
DEFINITION MODULE HeapType;
    EXPORT QUALIFIED heap, create, push, top, pop, empty;
    TYPE Heap;
    PROCEDURE create (VAR h : Heap);
    PROCEDURE push ( h : Heap; x : INTEGER);
    PROCEDURE top ( h : Heap) : INTEGER;
    PROCEDURE pop ( h : Heap);
    PROCEDURE empty ( h : Heap): BOOLEAN
END Heap Type.
```

Most egy olyan reprezentációt választunk, amelyben a legnagyobb elem könnyen elérhető, de egy elem behelyezése vagy törlése is kevés műveletet igényel. Egy heap-et egy `h[1..SIZE]` vektorral és egy `n` egésszel ábrázolunk. `n` a sorozat hossza és `n=0` jelenti az üres heap-et, a sorozatot pedig a `h[1], ... ,h[n]` elemekkel fejezzük ki. A vektorhoz hozzárendelt invariáns azonban erősebb, mint amit a 3. pontban használtunk [2]:

$$\forall i \in_{,} [1, n/2] : h[i] \geq \max(h[2*i], h[2*i+1])$$

Ez a vektorral való ábrázolás megfelel a heap egy kiegyensúlyozott rendezett bináris fával való ábrázolásának. A megfelelést az alábbi ábra szemlélteti:

$$h[1]\begin{cases} h[2] \begin{cases} h[4] \\ h[5] \end{cases} \\ h[3] \begin{cases} h[6] \\ h[7] \end{cases} \end{cases}$$

Az invariáns most azt jelenti, hogy minden részfában a gyökér a legnagyobb elem. Így a sorozat legnagyobb elemét mindig `h[1]` ábrázolja, annak kivételekor pedig `h[2]` vagy `h[3]`, ezek helyére pedig a „belőlük lelógó” elemek egyike, sit. „csúszik előre”. Egy elem behelyezésének vagy törlésének költsége egyaránt log n-nel arányos.

```
IMPLEMENTATION MODULE HeapType;
    FROM Storage IMPORT ALLOCATE;
    CONST SIZE = 100;
    TYPE HeapRep = RECORD
        n : CARDINAL;
        h : ARRAY [1..SIZE] OF INTEGER;
    END;
    TYPE Heap = POINTER TO HeapRep;
    PROCEDURE create(VAR h : Heap);
    BEGIN
        ALLOCATE(h, TSIZE(HeapRep));
        h↑.n := 0;
    END create;
    PROCEDURE push(h : Heap; x : INTEGER);
    VAR
        i,j : CARDINAL;
        swap : BOOLEAN;
    BEGIN
        IF h↑.n = SIZE THEN HALT; END;
        INC(h↑.n); j:= h↑.n; swap := TRUE;
        WHILE swap AND j >= 2 DO
            i := j; j := j DIV 2;
            IF x > h↑.h[j]
                THEN h↑.h[i] := h↑.h[j];
                ELSE swap := FALSE;
            END;
        END;
        IF swap
            THEN h↑.h[j] := x; ELSE h↑.h[i] := x;
        END;
    END push;
    PROCEDURE top(h : Heap) : INTEGER;
    BEGIN
        IF h↑.n = 0 THEN HALT; END;
        RETURN h↑.h[1];
    END top;
    PROCEDURE pop(h : Heap);
    VAR
        i.,j : CARDINAL;
        swap : BOOLEAN;
        tmp : INTEGER;
    BEGIN
        IF h↑.n = 0 THEN HALT; END;
        tmp := h↑.h[h↑.n]; DEC(h↑.n;)
        j := 1; swap := TRUE;
        WHILE swap AND 2*j <= h↑.n DO
```

```
                i := j; j := 2* j;
                IF j < h↑.n AND h↑.h[j+1] > h↑.h[j]
                    THEN INC(j);
                END;
                IF tmp > h↑.h[j]
                    THEN h↑.h[i] := h↑.h[j];
                    ELSE swap := FALSE
                END;
            END;
            IF swap
                THEN h↑.h[j] := tmp; ELSE h↑.h[i] := tmp;
            END;
        END pop;

        PROCEDURE empty(h : Heap) : BOOLEAN;
        BEGIN
            RETURN h↑.n = 0;
        END empty;
    END HeapType.
```

A Heap típus használata a MODULA-2 nyelvben nagyon természetesen, a többi típussal összhangban történik. A típusból felhasználni kívánt azonosítókat azonban előbb importálni kell:

```
FROM HeapType IMPORT Heap, create, push, pop, empty;
VAR h : Heap;
```

A típusműveletekben aktuális paraméterként kell megadni, hogy azok melyik heap-re vonatkoznak:

```
create (h);
push (h,0);
```

A MODULA-2 típusfogalma sokkal merevebb a SIMULA 67 nyelvénél. A MODULA-2 deklarációs részei fordítási időben értékelődnek ki, így a vektorok értelmezési tartományát csak fordítási időben kiértékelhető kifejezéssel lehet megadni. Mindezek miatt a heap típus nem paraméterezhető még számmal sem.

Bár az implementációs modul végén ebben a nyelvben is van egy inicializáló rész, az csak egyszer, a program elindításakor hajtódik végre. Így ez nem használható fel átlátszatlan típusú változók inicializálására. Egy `Heap` típusú változó definiálásakor annak értéke definiálatlan, s majd csak a `create` művelet végrehajtása után vesz fel meghatározott értéket. Az inicializálás (`create`) elmaradása programhibához vezet, de a nyelv nem ad eszközt ennek ellenőrzésére.

5. CLU

A CLU program modulok halmaza. A nyelv háromféle modult ismer, mindegyik egy-egy fajta absztrakció leírására szolgál. Az eljárás a funkcionális-, az iterátor a vezérlési-, a cluster a típusabsztrakció leírásának eszköze. [4,5]

Az eljárás valahány argumentumot felhasználva elvégzi a számítást, s annak eredményét változókban visszaadja a hívónak. A hívó és az eljárás között minden kommunikáció az argumentumokon és az eredményváltozókon keresztül történik (azaz globális változók nincsenek!). Az eljárás befejeződésekor az visszajelez egy, a befejeződés állapotára utaló feltételt is. Ez lehet normál, vagy lehet egy hibaállapot.

```
square_root = PROC (x:REAL) RETURNS (REAL)
                    SIGNALS(no_real_result)
    utasítások
END
```

Az iterátor az argumentumaiból értékek egy sorozatát számítja ki, mégpedig úgy, hogy az egymást követő meghívásakor a hívó megkapja a sorozat egy-egy újabb elemét. Az iterátort

```
iterator = ITER (t : típus) YIELDS (elem_típus)
    utasítások
END
```

formában definiálljuk, s egy FOR ciklus hívhatja.

```
FOR i : elem_típus IN iterator(x) DO
    ciklusmag
END
```

A cilkusváltozó a ciklusmag minden végrehajtása előtt felveszi az iterátor által szolgáltatott újabb értéket. A ciklus akkor ér véget, ha az iterátor terminál.

5.1 A cluster fogalma

A cluster szolgál a nyelvben új típus, adatabsztrakció létrehozására. A cluster objektumokat és ezeket létrehozó ill. ezeken manipuláló elemi műveleteket definiál.

```
set = CLUSTER IS create, insert, elements
    törzs
END
```

A cluster-en kívül a típus objektumai kizárólag a típusműveletek segítségével kezelhetők. A cluster belsejében kell megadni (más típusokból felépítve) a reprezentációt. A műveletek eljárások és iterátorok formájában implementálhatók.

A modulok statikus skalár értékekkel és típussal paraméterezhetők. Az utóbbi esetben megadhatók az aktuális típus paraméterre megszorítások is, amelyekben kiköthető, hogy annak milyen típusműveletekkel kell rendelkeznie.

```
set = CLUSTER [t : TYPE] IS create, insert, elements
        WHERE t HAS equal : PROCTYPE (t,t)
            RETURNS (BOOL)
    törzs
END
```

A paraméterezett cluster-t típusgenerátornak hívjuk.

5.2. Az ARRAY típusgenerátor

A CLU ARRAY típusa rendkívüli dinamizmusával különbözik a többi magasszintű nyelv tömbtípusától. Ebben a nyelvben a program végrehajtása közben megváltoztatható a tömb objektumok indextartománya és hossza is.

Az ARRAY típusgenerátor típusok egy végtelen osztályát definiálja. Bármely T típus esetén generálható az ARRAY[T] típus. Az ARRAY[T] típus egy egész típusú alsó indexhatárból és egy T típusú elemekből álló sorozatból áll. Az ARRAY típusgenerátorhoz több mint két tucat elemi művelet tartozik, ezekből azonban csak azokat soroljuk fel, amelyeket a példaprogramban felhasználtunk.

create: PROCTYPE (INT) RETURNS (ARRAY[T])
A create függvény ereménye egy olyan ARRAY[T] típusú érték, amelynek kezdőindexét a függvény argumentuma adja, a sorozat pedig nulla hosszúságú.

low: PROCTYPE (ARRAY[T]) RETURNS (INT)
A low függvény eredménye a tömb kezdőindexe.

size: PROCTYPE (ARRAY[T]) RETURNS (INT)
A size függvény eredménye az ARRAY[T] tömb elemszáma.

top: PROCTYPE (ARRAY[T]) RETURNS (T) SIGNALS (bounds)
A top függvény értéke a tömb utolsó (legnagyobb indexű) eleme.

fetch: PROCTYPE (ARRAY[T], INT) RETURNS (T) SIGNALS (bounds)
A fetch függvény értéke a tömbnek a második argumentumban megadott indexű eleme. A fetch függvény írható a szokásos formában is: x[i].

addh: PROCTYPE (ARRAY[T], T)
Az eljárás az argumentumában adott T típusú értéket hozzáfűzi a tömb elemeihez, megnövelve azzal a tömb hosszát.

remh: PROCTYPE (ARRAY[T]) RETURNS (T) SIGNALS (bounds)
Az eljárás elhagyja a tömb legnagyobb indexű elemét.

5.3. SEGNAL

A programok eljárásai nem minden esetben képesek elvégezni a rájuk bízott feladatot. A CLU lehetővé teszi az eljárás normális befejezése mellett annak olyan befejezését is, amikor a hívó információt kap az eljárás visszatérése után a rendellenességről. Ez a hibakezelési mechanizmus két részből áll: a névvel azonosítható hibaállapot beállításából (SIGNAL) és a hiba lekezeléséből (EXCEPT). A SIGNAL utasítást a hívott, az EXCEPT utasítást a hívó eljárás tartalmazza. Azt, hogy egy eljárás milyen hibákkal térhet vissza, a eljárásfej SIGNALS utasításában fel kell sorolni.

```
PROCTYPE (ARRAY[T], INT) RETURNS (T)
               SIGNALS (bounds)
PROCTYPE (STRING) RETURNS (INT)
               SIGNALS (invalid(CHAR))
```

SIGNAL utasítás bárhol előfordulhat egy eljárástörzsben. Hatására az eljárás végrehajtása megszakad, s a vezérlés visszatér a hívó utasításhoz csatolt hibakezelőbe (EXCEPT). A SIGNAL utasításban mód van értékek visszaadására is.

```
signaller :  PROCTYPE (i :  INT)
             SIGNALS (negative(INT), positive(INT), zero);
    IF i < 0 THEN SIGNAL negative (-i)
    ELSIF i > 0 THEN SIGNAL positive (i)
    ELSE SIGNAL zero
    END
END signaller
```

A SIGNAL utasítás által kiváltott hibákat a hívó utasításhoz csatolt EXCEPT utasításban lehet lekezelni. A hiba lekezelése után a program végrehajtása normál állapotban folytatódik.

```
signaller (x)
    EXCEPT WHEN negative (z :  INT) : utasítások
        WHEN positive (z :  INT) : utasítások
        WHEN zero :  utasítások
    END
```

5.4. Példa

Példánkban a heap típusnak ugyanolyan reprezentációját kódoljuk, mint a MODULA-2 nyelv esetében. Jól megfigyelhető, hogy a CLU eszközei (ARRAY típusgenerátor, SIGNAL) segítségével mennyivel rugalmasabb, használhatóbb modul definiálható.

```
heap_type = CLUSTER [item :  TYPE] IS
    create, push, top, pop, empty,
    WHERE item HAS lt :  PROCTYPE (item,item)
            RETURNS(BOOL);
REP = ARRAY [ item ];
create = PROC () RETURNS (CVT);
    RETURN ( REP$create(1) );
    END create;
push = PROC (h :  CVT, x :  INT);
    i,j :  INT;
    swap :  BOOL := TRUE;
    tmp :  item := x;
    REP$addh(h, x);j := REP$size(h);
    WHILE swap CAND j >= 2 DO
```

```
            i := j; j := j / 2;
            IF tmp > h[j]
                THEN h[i] := h[j];
                ELSE swap := FALSE;
            END;
        END;
        IF swap
            THEN h[j] := tmp; ELSE h[i] := tmp;
        END;
        END push;
    top = PROC (h :  CVT) RETURNS (item) SIGNALS (empty_heap);
        IF REP$size(h) = 0 THEN SIGNAL empty_heap; END:
        RETURN ( h[1] );
        END top;
    pop = PROC (h :  CVT) SIGNALS (empty_heap);
        i,j :  INT;
        swap :  BOOL := TRUE;
        tmp :  item;
        IF REP$size(h) = 0 THEN SIGNAL empty_heap; END;
        tmp := REP$top(h); REP$remh(h); j := 1;
        WHILE swap CAND 2*j <= h.n DO
            i := j; j:= 2* j;
            IF j < REP$size(h) CAND h[j+1] > h[j] THEN
                j := j + 1;
            END;
            IF tmp > h[j]
                THEN h[i] := h[j];
                ELSE swap := FALSE;
            END;
        END:
        IF swap
            THEN h[j] := tmp; ELSE h[i] := tmp;
        END;
        END pop;
    empty = PROC (h ; CVT) RETURNS (BOOL);
        RETURN ( REP$size(h) = 0);
        END empty;
    END heap_type;
```

6. Ada

Az Ada tervezésének fő szempontjai a program megbízhatósága, karbantarthatósága, emberközelisége (olvashatósága) és hatékonysága voltak. A nyelv alapja a Pascal, de néhány vonást átvett a következő nyelvekből is: Euclid, Lis, Mesa, Modula, Sue, Algol 68, Simula, Alphard, CLU. [6]

Az Ada program programegységek felsorolása. Ada programegység az alprogram, a package (ami alprogramok, típusok, konstansok és változók gyűjteménye), a task (ami párhuzamosan végrehajtható számításokat tartalmaz) és a generic (ami egy paraméterezett package vagy alprogram, ahol a paraméter típus is lehet).

A programegységek két részből állnak: specifikációs részből, ami a más egységekből látható információt tartalmazza, és a törzsből, ami az implementációt tartalmazza, és ami más egyégekből nem látható. A specifikációs rész és a törzs szétválasztása és a külön való fordítás lehetővé teszi nagymértékben független programkomponensek tervezését, írását és tesztelését.

Az Ada nyelvhez tartozik egy speciális könyvtár, amelyben a programegységek helyezkednek el. Egy Ada program kifejlesztése a könyvtári egységek adatbázisának felépítését jelenti. Kezdetben a könyvtár csak predefinit packageket tartalmaz, ezekhez illesztheti hozzá a felhasználó a maga könyvtári egységeit.

6.1. A package

Az Ada az általános célú package fogalmát felhasználva, azon belül ad eszközöket típus megvalósítására. A package a MODULA-2 modul fogalmához nagyban hasonlító programegység (és egyben fordítási egység). Szintén két részre, specifikációs részre és törzsre válik szét; a specifikációs rész felsorolja az exportálni kívánt objektumokat, típusokat és programegységeket, a törzs magába foglalja ezek implementációját.

```
package P is
    -- exportált, más egységből is látható nevek
end P;

package body P is
    -- az exportált egységek implementációja
end P;
```

Az Ada fordítási egységeinek fordítási sorrendjére vonatkozó szabály lényegét tekintve megegyezik a MODULA-2 ilyen szabályával.

6.2. A private típus

A megvalósítani kívánt típust a specifikációs részben ún. private típusként definiálhatjuk. A típus reprezentációját a specifikációs rész végén, az ún. private szakaszban kell megadni, ami nem kerül exportálásra.

```
package HEAP_TYPE is
    type HEAP is private;
    ...
private
    type HEAP is ...
end HEAP_TYPE;
```

A MODULA-2 -höz képest az a legnagyobb különbség, hogy így megszűnik a reprezentációra vonatkozó megkötés, hogy ti. az csak a `WORD` típussal kompatibilis típus lehet. Mivel viszont a reprezentáció a specifikációs részbe került, így annak módosítása a rá hivatkozó modulok újrafordítását vonja maga után.

Az Ada — a MODULA-2 -höz hasonlóan — a private típusra megenged két általános műveletet: az értékadást és az egyenlőség vizsgálatát. Mivel felléphetnek ezekkel a műveletekkel kapcsolatban bizonyos problémák, amelyeket a MODULA-2-nél már kifejtettünk, ezért az Adában a private típusokra ezeket a műveleteket meg lehet tiltani.

```
type HEAP is limited private;
```

A limited private típusú objektumhoz kozárólag az ugyanebben a package-ben definiált típusműveleteken keresztül lehet hozzáférni.

Az Ada lehetőséget ad private típushoz konstans definiálására. Vannak olyan típusok, amelyeknek nevezetes konstansai vannak. Ilyen pl. a `COMPLEX` típus, amit az alábbi módon definiálhatunk:

```
package COMPLEX_TYPE is
    type COMPLEX is private;
    I : constant COMPLEX;
    function "+" (A, B : COMPLEX) return COMPLEX;
    ...
private
    type COMPLEX is record
        RE, IM : REAL;
    end record;
    I : constant COMPLEX := (0.0, 1.0);
end COMPLEX_TYPE;
```

Ez a lehetőség nagyban javítja a private típus használhatóságát.

Az Adában a private típus ugyanúgy paraméterezhető objektummal, mint az egyéb típusok, tehát mint pl. a rekord. Az alábbi verem típus paramétere a verem maximális mérete.

```
package STACK_TYPE is
    type STACK(MAX:NATURAL := 100) is limited private;
    EMPTY_STACK, FULL_STACK : exception;
    procedure PUSH(S:in out STACK; X:in INTEGER);
    ...
private
    type STACK(MAX:NATURAL := 100) is record
        N : NATURAL := 0;
        S : array (1..MAX) of INTEGER;
    end record;
end STACK_TYPE;

with STACK_TYPE: use STACK_TYPE;
```

```
procedure P is
    S : STACK(64);
    T : STACK;
    ...
    PUSH(S, A);
end P;
```

A P eljárásban az S verem legfeljebb 64, a T verem 100 egész befogadására képes.

6.3. A generic

Azért, hogy a típus alaptípusát is megadhassuk paraméterként, ún. generic package programegységet kell használnunk. A generic package két fő dologban különbözik a package-től:

- értékkel, típussal és alprogrammal paraméterezhető, és
- mivel ez csak egy váz, a package-ként való használatához aktuális paraméterek megadásával generálni kell belőle egy (hagyományos) package-et.

```
generic
type ELEM is private;
    package STACK_TYPE is
type STACK(MAX:NATURAL := 100) is limited private;
    EMPTY_STACK, FULL_STACK : exception;
    procedure PUSH(S:in out STACK; X :in ELEM);
    ...
private
    type STACK(MAX:NATURAL := 100) is record
        N : NATURAL := 0;
        S : array (1..MAX) of ELEM;
    end record;
end STACK_TYPE:
```

Egy verem objektum definiálásakor most tehát a következőképpen kell eljárnunk:

```
package INTEGER_STACK is new STACK_TYPE(INTEGER):
...
S : INTEGER_STACK.STACK;
...
INTEGER_STACK.PUSH(S,0);
```

Itt az első deklarációs utasítás egy új verem típus bevezetése (instantiation), ahol megadjuk, hogy az új INTEGER_STACK package-be beágyazott STACK típus alaptípusa az INTEGER típus legyen. A következő sorokban a STACK típus és egy típusműveletének használatát mutatjuk be. Amennyiben az INTEGER_STACK package definiálása után a

```
use INTEGER_STACK;
```

utasítást is megadjuk, ezzel a package-ben definiált neveket „közvetlenül láthatóvá" tesszük. A korábbi sorok most így festenek:

```
S : STACK;
...
PUSH(S,0);
```

6.4. Az exception

A hibák és a kivételes helyzetek kezelésére az Ada hasonló mechanizmust alkalmaz, mint a CLU. A nyelv minden hibafajtát egy névvel jelöl meg; minden blokk tartalmazhat egy hibakezelőt, amelyben külön utasítások tartoznak a különböző hibákhoz.

```
begin
    A := B(I)/D;
exception
    when ZERO_DIVIDE =>
        A := 0;
    when CONSTRAINT_ERROR =>
        A := B(B'LAST);
end;
```

A blokkban fellépő valamely hiba esetén a blokk végrehajtása a hibakezelőben folytatódik, s ha itt nincs rendelkezés a hiba lekezelésére, akkor a hívó blokkban lép fel a hiba. Ez a keresés addig tart, amíg egy blokkban a hiba lekezelése meg nem történik, s ez a blokk már normál módon (hibaállapot nélkül) tér vissza az őt aktivizáló blokkba. Ez a rendszer bevált, és nagyon hasznos a specifikációjukat pontosan teljesítő, megbízható blokkok (eljárások, függvények), típusműveletek írásához.

Az Ada hibakezeléssel kapcsolatban is megemlíthető azért két hiányosság: az az eljárás, amelyben fellép egy hiba, nem adhat vissza információt a paramétereken keresztül, így a nyelv némely esetben kikényszeríti globális változók használatát; másodszor, ha a programozó nem ír a hívó blokkba hibakezelőt, a fellépő hiba tovaterjed, s esetleg majd egy olyan modulban lép fel, amelyben erre számítani nem lehet.

6.5. Példa

Az eddigi példákban választott reprezentációk — a CLU kivételével — olyanok voltak, ahol a HEAP elemeinek száma korlátozva volt. Újabb példánkban olyan reprezentációt választunk, amelyben ilyen elvi korlát nincs: a heap típust ezúttal vektorok kétirányú listájával ábrázoljuk. A heap fejrésze két pointert tartalmaz: F a lista első elemének, L az utolsó elemének pointere. Egy listaelemben (HEAP_REC) megtalálható egy vektor (D), a vektor hossza (M) és egy-egy a listában előre (FW) ill. vissza (BW) mutató pointer.

A listáról kikötjük az alábbi invariánst:

- mindegyik vektorra teljesül, hogy

$$\forall i \in [1, M/2] : D(i) \geq \max(D(2*i), D(2*i+1))$$

- a listában mindegyik vektor első eleme kisebb, mint az utána következő listaelem vektorának első eleme,
- az utolsó listaelemet kivéve a vektor mindegyikben megtelt.

```
generic
    type ITEM is private;
    with function ">" (A, B : ITEM) return BOOLEAN;
package HEAP_TYPE is
    type HEAP is limited private;
    HEAP_EMPTY : eception;
    procedure CREATE(H : in out HEAP);
    procedure PUSH (H : in out HEAP; X : in ITEM);
    function TOP (H : HEAP) return ITEM;
    procedure POP (H : in out HEAP);
    function EMPTY (H : HEAP) return BOOLEAN;
private
    SIZE : constant := 32;
    type HEAP_REC_ADDRESS;
    type HEAP is record
        F, L : HEAP_REC_ADDRESS := null;
    end record;
    type HEAP_REC is record
        FW, BW : HEAP_REC_ADDRESS := null;
        M : NATURAL := 0;
        D : array (1..SIZE) of ITEM;
    end record;
    type HEAP_REC_ADDRESS is access HEAP_REC;
end HEAP_TYPE;
package body HEAP_TYPE is
procedure CREATE(H : in out HEAP) is
begin
    H.F := new HEAP_REC; H.L := H.F;
end CREATE;
procedure PUSH(H : in out HEAP; X : in ITEM) is
    P : HEAP_REC_ADDRESS := H.L;
    TMP : ITEM := X;
    I, J : INTEGER;
    SWAP : BOOLEAN;
begin
    if P.M = SIZE then
        if P.FW = null then
            P.FW := new HEAP_REC;
            P.FW.BW := P;
        end if;
        P := P.FW;
```

```
    end if;
    - - p = H.L and H.L.M < SIZE
    P.M := P.M + 1; J := P.M; SWAP := TRUE;
    while SWAP and then J >= 2 loop
        I := J; J := J/2;
        if TMP > P.D(J)
            then P.D(I) := P.D(J);
            else SWAP := FALSE;
        end if:
    end loop;
    if SWAP
        then P.D(J) := TMP; else P.D(I) := TMP;
    end if;
    - - ∀ i ∈ [1,M/2] :  D(i) ≥ max (D(2*i),D(2*i+1))
    while P /= null and then P.D(1) > P.BW.D(1); loop
        TMP := P.D(1); P.D(1) ;= P.BV.D(1);
        P.BW.D(1) := TMP; P := P.BW;
    end loop;
end PUSH;
function TOP(H : HEAP) return ITEM is
begin
    if H.F.M = 0 then raise HEAP_EMPTY; end if;
    return H.F.D(1);
end TOP;
procedure POP(H : in out HEAP) is
    P : HEAP_REC_ADDRESS := H.F;
    I, J : INTEGER;
    SWAP : BOOLEAN;
    TMP : ITEM;
begin
    if H.F.M = 0 then raise HEAP_EMPTY; end if;
    while P /= H.L and then
        P.FW.D(1) > MAX(P.D(2),P.D(3)) loop
        P.D(1) := P.FW.D(1); P := P.FW;
    end loop;
    - - H.L.D(H.L.M) ≤ H.L.D(1) ⇒
    - - H.L.D(H.L.M) ≤ MAX(P.D(2),P.D(3))
    TMP := H.L.D(H.L.M); H.L.M.:= H.L.M - 1;
    J := 1; SWAP := TRUE;
    while SWAP and then 2* J <= P.M loop
        I := J; J := 2* J;
        if J < P.M and then P.D(J+1) > P.D(J) then
            J := J + 1;
        end if;
        if TMP > P.D(J)
            then P.D(I) := P.D(J);
            else SWAP := FALSE;
```

```
        end if:
    end loop;
    if SWAP
        then P.D(J) := TMP; else P.D(I) := TMP;
    end if;
    - - ∀i∈[1,M/2] :  D(I) ≥ max(D(2*i),D(2*i+1))
    if H.L.M = 0 and then H.L /= H.F then
        H.L := H.L.BW;
    end if;
    - - H.F /= H.L => H.L.M /= 0
end POP;

function EMPTY(H : HEAP) return BOOLEAN is
begin
    return H.F.M = 0;
end EMPTY;

end HEAP_TYPE;
```

Nagyon nehezen megoldható problémája a programozási nyelvnek a változók inicializálatlanságának a kérdése. Az Ada sem képes megbirkózni ezzel teljesen, viszont a rekord típusú változók esetében használható megoldást ad.

```
type HEAP_REC is record
    FW, BW : HEAP_REC_ADDRESS := null;
    M : NATURAL := 0;
    D : array (1..SIZE) of ITEM;
end record;
```

A mezőhöz írt kezdőértékadás azt jelenti, hogy minden változó definiálásakor az azonnal felveszi a típusdefinícióban megadott (és a típus definiálásakor meghatározott) értékeket. A nyelv tehát nem kényszeríti ki a programozóból az inicializálatlan változók elkerülését, de lehetőséget ad arra.

7. A legújabb nyelvek törekvései

A típusfogalom körüli kutatás egyáltalán nem esett vissza, sőt, az utóbbi években az objektum-orientált nyelvekkel kapcsolatban új lendületet vett. Olyan nyelvi eszközöket keresnek, amelyekkel már nem csak egy típus, hanem típusok egy rendszere, osztálya is leírható. Sőt, típusok osztályán értelmezett műveletek (pl. öröklés) nyelvi szinten való kezelésének lehetőségeit is tanulmányozzák. Ezektől a kutatásoktól nemcsak a típusfogalom általánosítása várható, hanem ezek valószínűleg nagy hatással lesznek a modulfogalomra és a párhuzamosságra is. Ezek az új próbálkozások, új nyelvek (pl. Owl, Oberon, Eiffel) azonban még csak kutatási stádiumban vannak, a gyakorlatban való kipróbálásukra kicsit még várni kell.

IRODALOM

[1] KNUTH, D.E., *The art of computer programming* (Addison-Wesley Publishing Company, 1968).

[2] DAHL, O.-J., MYHRHAUG, B., NYGAARD, K., *SIMULA 67 common base language* (Norwegian Computing Center, Oslo, 1970).

[3] WIRTH, N., *Programming in Modula-2* (Springer-Verlag, Berlin, Heidelberg, New York, 1983).

[4] LISKOV, B., MOSS, E., SCHAFFERT, C., SCHEIFLER, B., SNYDER, A., *CLU Reference Manual* (MIT Laboratory for Computer Science, 1978).

[5] LISKOV, B., SNYDER, A., ATKINSON, R., SCAFFERT, C., "Abstraction Mechanisms in CLU", *CACM* **20** (1977), 564–576.

[6] YOUNG, S.J., *An Introduction to Ada* (Ellis Horwood Limited,, 1983).

[7] VARGA, L., "Típusspecifikációk helyességének vizsgálata", *Alkalmazott Matematikai Lapok* **13** (1987–88), 57–68.

[8] DUPONCHEEL, L., HEYMAN, J., VAN POYMBROECK, W., "Algebraic Data Type Specifications Language, Method and Tools."

(Beérkezett: 1990. április 12.)

KOZICS SÁNDOR
ELTE TTK, ÁLTALÁNOS SZÁMÍTÁSTUDOMÁNYI TANSZÉK
H-117 BUDAPEST, BOGDÁNFY ÚT 10/B.

DEFINITION OF ABSTRACT DATA TYPES IN DIFFERENT PROGRAMMING LANGUAGES

S. KOZICS

An exploration of how the most important features of abstract data types, playing a central role in reliability of programs, developed in various programming languages from SIMULA 67 to Ada.

LOGIKAI FÜGGVÉNYEK PR—ZÁRT OSZTÁLYAINAK HÁLÓJA*

BAGYINSZKI JÁNOS

Gödöllő–Budapest

A logikai függvények halmaza hatványhalmazán bevezetünk egy *pr-lezárásnak* nevezett lezárási operációt. E lezárás szerint a logikai függvények egy halmaza *pr-zárt*, ha az a projekciókat tartalmazó olyan (POST definíciója szerint) zárt függvényosztály, amely zárt a primitív rekurzió képzésére. POST eredményeinek felhasználása nélkül meghatározzuk a *Post-diagram* analógiájára a *pr-zárt* osztályok hálóját. A „szigorúbb" lezárásból következően, a logikai függvények klónhálójánál egyszerűbb hálót kapunk eredményül, amely egy végtelen lánc. Megadjuk továbbá POST tételeinek megfelelőit. A vizsgálatokat a számítástudomány különböző részei közötti összefüggések keresése motiválta.

1. Bevezetés

E. L. POST 1920-ban írt, nem publikált dolgozatában (l. [3]) és a 21 évvel később kiadott [5] monográfiában meghatározta egy rögzített kételemű halmazon értelmezhető többváltozós függvények összes zárt osztályát, valamint az utóbbiban ezek tartalmazás szerinti diagramját. Az (orosz, német és magyar nyelvű) irodalomban félreérthető és hibás hivatkozások vannak POST dolgozatait és eredményeit illetően, részben azért, mert ezek nehezen [4] illetve egyáltalán nem érhetők el [5] (pl. hazánkban). Ezért [3] és [4] alapján megpróbálom röviden ismertetni a logikai függvények zárt osztályaira vonatkozó *Post-eredmények* tényleges előfordulását-bevezetésként. Az AMS (*Amerikai Matematikai Társaság*) egy ülésén, 1920-ban ismertették POST két dolgozatát (a nem-publikált [3] és a [4]). A [3] szerint 2. sorszám alatt bemutatott (s később megjelent [4]) dolgozatban WHITEHEAD és RUSSELL: *Principia Mathematica* (*Part I. Sect. A*) c. könyvében [10] leírt logikai deduktív rendszerből kiindulva, egységes módszert ad a posztulátumokból való származtathatóság eldöntésére — tetszőleges logikai függvényre. A módszer alapja logikai függvények igazság-táblázattal való ábrázolása. Az általánosítás egyik iránya az igazság-táblázat (a logikai függvények) értelmezése kettőnél több elemű halmazra (kapcsolódva SHEFFER és NIKOD munkájához). A második általánosítás új logikai rendszerek bevezetését jelenti. Ez utóbbi vezetett a 3. sorszám alatt bemutatott eredményekhez. E szerint POST primitív művelettáblák (azaz igazságfüggvények) kombinálásával generálható összes igazságfüggvény-rendszert tekinti. Megmutatja, hogy 65 olyan rendszer létezik, amely legfeljebb 3-változós (primitív) függvényekkel generálható és 8 végtelen családja van a generátor-halmazában legalább négyváltozós függvényt igénylő rendszereknek. A függvényeket formulával adja meg és bizonyítja, hogy a fentieken kívül nincs más zárt függvényrendszer. A [4] dolgozatból

*A dolgozat az OTKA T4295/92 számú téma keretében készült.

látható, hogy függvények kombinálásán összetett függvény képzését, változók permutálását és két változó azonosítását érti — véges sok lépésben. Más szavakkal: függvények egy halmazát zártnak nevezi POST, ha az összetett függvény képzése és a változók átjelölése nem vezet ki a függvényhalmazból.

A. I. MALCEV *iteratív osztály*nak nevez egy zárt osztályt (l. [8]), ha az zárt a fiktív változó hozzávételére és törlésére is: tehát egy függvénnyel együtt adottnak tekintjük az összes olyan függvényt, amely attól csak fiktív változókban különbözik. JABLONSZKIJ, GAVRILOV és KUDRJAVCEV a [11] könyvben átdolgozva POST monográfiáját (bizonyításokat egyszerűsítve) az összes iteratív osztály diagramját adják meg. (Például POSTnál az egyváltozós projekció és az összes projekció két külön osztályt jelent, iteratív osztályként viszont csak az utóbbi létezik). A [11] könyv magyar nyelvű ismertetése [2], további egyszerűsítéseket tartalmaz. DEMETROVICS J. dolgozatából átvett táblázatban (11. oldal) megtalálható a nyolcféle osztály definíciója ($O-$, $S-$, $P-$, $L-$, $D-$, $A-$, $C-$ és $F-$ osztályok, összesen 49 meghatározás, azonban nyolc, paraméterrel definiált F-osztály mindegyike megszámlálhatóan végtelen osztályt képvisel). Szerepel továbbá a táblázatban minden egyes osztály tipusának, rendjének és egy bázisának megadása.

A 70-es években megjelentek a $k(\geq 2)$ elemű halmazon értelmezett függvények zárt osztályait algebrai szempontból vizsgáló dolgozatok is.

Minthogy az algebrák függvényklónjai a projekció-tartalmazásra nézve zárt iteratív osztályok, ezért a *„Post-diagram”*-ból elhagyva a (projekciókat nem tartalmazó) O_2, O_3 és O_7 iteratív osztályokat, a részalgebra-hálónak megfelelő klónhálóhoz jutunk. BAKER, PIXLEY, MCKENZIE és mások univerzális-algebrai eredményeit felhasználva, RESCHKE és DENECKE [7] tisztán univerzális algebrai úton származtatják POST szóban forgó eredményeit. Ők is (akárcsak a [11] és [2] dolgozatok) megtartják POST jelöléseit a zárt osztályokra. A továbbiakban általunk megadott hálóval való összehasonlíthatóság céljából idézzük a [7] dolgozatból a hálót (l. még [6]): az 1. ábrán a *„Post-diagram”*, a 2. ábrán a *pr-zárt* osztályok hálója látható. Az előbbit kiegészítettük (folytonos vonal helyett pontsorral jelölve) az O_2, O_3 és O_7 osztályokkal, hogy az eredeti, *„Post-diagram”* is látható legyen. (Pontosabban, ez szerepel a [11] könyvben, POSTnál ezen kívül szerepelnek az S_1^*, P_1^*, A_4^* és O_i^* $(1 \leq i \leq 9)$ osztályok, ahol $O_i^* = O_i^{(1)}$ és S_1^*, P_1^*, A_4^* rendre az S_1, P_1, A_4 osztályok fiktív változót nem tartalmazó elemeiből áll. Erre azonban csak következtetni tudtam, mert [5] – az OSzk nyilvántartása és „magánnyomozásom” szerint – hazánkban nem hozzáférhető.)

Koncepciónkat egyrészt elsősorban a $k > 2$ eset (k-értékű logika függvényklónhálójának számossága nem megszámlálható) motiválta.

Másrészről az elméleti számítástudomány különböző ágai (így a k-értékű logika függvényklónjai és a rekurzív függvények illetve a formális nyelvek [1]) közötti kapcsolatok megvilágítása volt a cél. A koncepciót egy következő dolgozatban szeretnénk részletesebben kifejteni, amely a $k > 2$ esettel foglalkozik.

Az elképzelésünk lényege az, hogy a klónokhoz vezető, az algebrában természetes lezárásnál szigorúbb lezárási műveletet vezetünk be. Ez a számítástudományban természetes, rekurzióval való függvényképzést is tartalmazza: a (primitív) rekurzióra zárt klónok hálóját határozzuk meg (ebben a dolgozatban a *Boole-függvények* esetében). Meglepetéssel észleljük, hogy a *Post-diagramnak* megfelelő teljes háló lényegesen egyszerűbb az előbbinél — egy végtelen lánc adódik (2. ábra), amely a C_1 és C_3 osztályokon kívül csak az F_8^m ($m \geq 2$) és az F_8^∞ osztályokat tartalmazza. (Itt és a továbbiakban μ helyett m-et írunk). A *Post-háló* felhasználásával e lánc megadása csak annak bizonyítását igényelné, hogy

a) minden *pr-zárt* osztály tartalmazza az F_8^∞ elemeit,

b) F_8^∞, F_8^m ($m \geq 2$), C_3, C_1 a *pr-zárt* osztályok és

c) C_1-ben maximális C_3, C_3-ban maximális F_8^2, F_8^m-ben maximális F_8^{m+1} ($m \geq 2$), és $F_8^\infty = \bigcap\limits_{m\geq 2} F_8^m$.

Azonban éppen annak illusztrálására, hogy ez a szerkezet-leírás lényegesen egyszerűbb, nem használjuk fel POST eredményeit, hanem öntartalmazó módon építjük föl a hálót (ezért a reláció-terminológiát sem használjuk [8], [7]).

Megjegyezzük, hogy a *zárt függvényosztályok*, az *iteratív osztályok*, a *klónok* és a *pr-zárt* osztályok összessége egyaránt egy-egy (speciális) lezárási rendszert határoz meg. A lezárási rendszerek tulajdonságairól e lap hasábjain a [9] dolgozatból tájékozódhat az olvasó.

A 3.1. segédtétel és a 3.2. tétel lényegében ismert tételek, a 3.5. tétel és a 3.6. segédtétel állítását, továbbá a 3.9. tétel (2) és a 3.10. tétel (1)c. állítása bizonyításának alapötletét átvettük [11]-ből.

2. Definíciók és jelölések

2.1. Definíció. Jelölje $C_1^{(n)}$ a rögzített $\{0,1\}$ kételemű halmazon értelmezett *n-változós függvények* halmazát: $C_1^{(n)} := \{f : \{0,1\}^n \to \{0,1\}\}$, ahol $n \geq 1$ egész szám, és legyen $C_1 := \bigcup\limits_{n\geq 1} C_1^{(n)}$ az összes véges-változós függvény (logikai függvény) halmaza a ($\{0,1\}$ alaphalmazon). C_1 elemeit szokás *Boole-függvény*eknek is nevezni. (0-változós függvényeket nem definiálunk). $C \subset C_1$ esetén legyen $C^{(n)} := C \cap C_1^{(n)}$. A továbbiakban függvényen mindig *Boole-függvényt* értünk. Az egyszerűség kedvéért alkalmazzuk az $\tilde{x} := (x_1, \ldots, x_n) \in \{0,1\}^n$ és $(\tilde{x}, x_{n+1}) := (x_1, \ldots, x_n, x_{n+1})$ rövid jelöléseket, továbbá $+$ és $\cdot$ jelentse rendre a mod2 összeadást illetve mod2 szorzást, és legyen $a \vee b := a + b + ab$ ($a, b \in \{0,1\}$).

2.2. Definíció. c_0^n, c_1^n, e_i^n ($1 \leq i \leq n$) rendre az n-változós 0, 1 *konstans függvényt* illetve az *i-edik projekciót* jelöli: minden $\tilde{x} \in \{0,1\}^n$ esetén $c_0^n(\tilde{x}) = 0$, $c_1^n(\tilde{x}) = 1$, $e_i^n(\tilde{x}) = x_i$. Jelölje E a projekciók halmazát: $E := \{e_i^n \mid 1 \leq i \leq n,\ n \geq 1$ egész szám$\}$. Ugyancsak lekötjük többször előforduló függvények jelölésére a „h" betűt a következő (index szerinti rekurzív) definícióval: minden $(x_1, \ldots, x_{m+2}) \in \{0,1\}^{m+2}$ ($m \geq 2$, egész) esetén

$$h_{m+1}(x_1, \ldots, x_{m+2}) := x_1 \cdot \ldots \cdot x_{m+1} + (x_1 \cdot \ldots \cdot x_{m+2} + x_{m+2}) h_m(x_1, \ldots, x_{m+1})$$

és

$$h_2(x_1, x_2, x_3) := x_1 \cdot x_2 + x_2 \cdot x_3 + x_3 \cdot x_1.$$

Megjegyzés. Könnyen belátható, hogy $h_m(x_1, \dots, x_{m+1}) = 1$, ha $\{x_1, \dots, x_{m+1}\}$ elemei között legfeljebb egy 0 (a többi 1), egyébként $h_m(x_1, \dots, x_{m+1}) = 0$. (Bizonyítása m szerinti indukcióval). A negyedik 1-változós függvényt jelölje s; $s(x) := x + 1$, $x \in \{0, 1\}$.

2.3. Definíció. A $C \subseteq C_1$ függvényhalmazt *pr-zárt osztály*nak nevezzük, ha $E \subset C$, továbbá $g_1, \dots, g_n \in C$, $g_0 \in C^{(n)}$, $g \in C^{(n+2)}$ esetén $g_0(g_1, \dots, g_n) \in C$ és $f \in C^{(n+1)}$, ahol az f függvényt az

$$f(\tilde{x}, 0) := g_0(\tilde{x}),$$
$$f(\tilde{x}, 1) := g\big(\tilde{x}, f(\tilde{x}, 0), 0\big), \quad \tilde{x} \in \{0, 1\}^n$$

primitív rekurzió definiálja.

$C' \subseteq C_1$ függvényhalmaz *pr-lezártján* a C'-t tartalmazó legszűkebb C *pr-zárt* osztályt értjük. Jelölése: $C := [C']_{pr}$.

Könnyen látható, hogy C_1 részhalmazain a $C \rightarrow [C]_{pr}$ hozzárendelés lezárási operáció. Ha $[C]$, $[C]_{it}$, $[C]_{cl}$ jelöli rendre a POST által definiált lezárást, az iteratív illetve a klón-lezárást, akkor a definícióból következően $[C \cup E] = [C \cup E]_{it} = [C]_{cl}$, ezért igaz a következő

ÁLLÍTÁS. *Tetszőleges $C \subseteq C_1$ esetén $C \subseteq [C] \subseteq [C]_{it} \subseteq [C]_{cl} \subseteq [C]_{pr}$.* □

2.4. Definíció. Legyen $C \subseteq C_1$ egy *pr-zárt* osztály, azaz $[C]_{pr} = C$. $C' \subset C$ *maximális C-ben*, ha $[C']_{pr} \neq C$ és $f \in C \setminus C'$ esetén $[\{f\} \cup C']_{pr} = C$.

2.5. Definíció. A $C \subseteq C_1$ *pr-zárt* osztály *aritása* (vagy *rendje*) az a legkisebb pozitív egész n szám, amelyre $[C^{(n)}]_{pr} = C$.

Végül néhány függvényosztályt értelmezünk, amelyekről később megmutatjuk, hogy *pr-zártak*.

2.6. Definíció. $C_3 := \{f \mid f \in C_1$ és $f(0, \dots, 0) = 0\}$ halmaz a *0-őrző* függvények osztálya.

2.7. Definíció. Az $f \in C_1^{(n)}$ *kielégíti az $\langle A^m \rangle$ feltételt* ($m \geq 2$), ha az $A(f) := \{\tilde{a} \mid f(\tilde{a}) = 1\}$ halmaz bármely m-elemű $A_m := \{\tilde{a}_j \mid \tilde{a}_j = (a_{j1}, \dots, a_{jn}),\ 1 \leq j \leq m\}$ részhalmazához létezik olyan $1 \leq i \leq n$ index, hogy $a_{ji} = 1$ teljesül minden j esetén.

Ha f kielégíti az $\langle A^m \rangle$ feltételt $m = |A(f)|$ esetén is, akkor azt mondjuk, hogy *f eleget tesz az $\langle A^\infty \rangle$ feltételnek.*

$F^m := F_8^m := \{f \mid f \in C_1$ és eleget tesz az $\langle A^m \rangle$ feltételnek $\}$,
$F^\infty := F_8^\infty := \{f \mid f \in C_1$ és eleget tesz az $\langle A^\infty \rangle$ feltételnek $\}$.

Megjegyzés. A 2.6. és 2.7. definíciókból közvetlenül következik, hogy

$$F^{\infty} \subseteq F^{m+1} \subseteq F^{m} \subseteq C_3 \subseteq C_1, \quad \text{és} \quad F^{\infty(1)} = F^{m(1)} = C_3^{(1)} = \{c_0^1, e_1^1\}. \quad \square$$

3. Eredmények

Meghatározzuk a *pr-zárt* osztályokat és azok tartalmazás-szerinti hálóját, valamint a [11] dolgozatban bizonyított tételek *pr-lezárásra* vonatkozó megfelelőjét bizonyítjuk. A bizonyítások előkészítéséhez néhány segédtételt bocsátunk előre. A *Boole-függvények* elméletéből ismert kifejtési tételt a szokásostól eltérően, egyváltozóra kifejtett alakban és ÉS-NEM-VAGY reprezentáció (azaz teljes diszjunktív normálalak) helyett $GF(2)$ fölötti polinóm-reprezentációban adjuk meg. Ennek alapja az a jól ismert tény, hogy minden véges test fölött értelmezett függvény az illető test fölötti polinómfüggvényként írható.

3.1 SEGÉDTÉTEL (kifejtési tétel). *Minden Boole-függvényre érvényes a következő kifejtés* ($f \in C_1^{(n+1)}$):

$$f(\tilde{x}, x_{n+1}) = x_{n+1} \cdot f(\tilde{x}, 1) + (1 + x_{n+1}) f(\tilde{x}, 0).$$

Bizonyítás. $x_{n+1} = 0$ és $x_{n+1} = 1$ lehetséges helyettesítések azonosságokat eredményeznek. $\square$

Ismeretes, hogy a $GF(2)$ testben érvényesek az $a \cdot a = a$, $a + a = 0$ azonosságok. Nevezzünk egy $GF(2)$ fölötti ($n \geq 1$ határozatlanú) polinómot *Zsegalkin-polinómnak*, ha a szóban forgó azonosságokra nézve egyszerűsített alakban van megadva.

3.2 TÉTEL. *Minden Boole-függvény (a tagok sorrendjétől eltekintve) kölcsönösen egyértelműen reprezentálható ($GF(2)$-fölötti) Zsegalkin-polinómmal.*

Bizonyítás. $C_1^{(1)}$ elemeire az állítás nyilvánvaló: $c_0^1(x) = 0$, $c_1^1(x) = 1$, $e_1^1(x) = x$, $s(x) = x + 1$. Feltételezve, hogy $C_1^{(n)} (n \geq 1)$ elemei egyértelműen előállíthatók *Zsegalkin-polinómként*, az egyértelműség n-re vonatkozó teljes indukcióval adódik a 3.1. segédtétel alapján.

A kölcsönösség annak a ténynek a következménye, hogy az n-határozatlanú *Zsegalkin-polinómok* száma megegyezik a $C_1^{(n)}$ elemszámával. Ez utóbbi szám 2^b, ahol $b = 2^n$, a függvények értelmezési tartományának elemszáma, 2^b pedig a függvényekhez tartozó érték-sorozat lehetséges kitöltéseinek száma. Másrészről a *Zsegalkin-polinóm* lehetséges szorzat-tagjainak száma 2^n (bármelyik határozatlan vagy szerepel vagy nem), s egy polinómot egyértelműen meghatározzák a benne szereplő tagok. $\square$

A továbbiakban a függvényeket gyakran *Zsegalkin-polinómmal* reprezentáljuk. A következő állítás a primitív rekurzióra való zártság megállapítását segíti technikailag.

3.3 SEGÉDTÉTEL (*pr-lemma*). *Legyen* $C \subseteq C_1$ *egy pr-zárt osztály. Ha* $g_0 \in C^{(n)}$ *és* $g \in C^{(n+2)}$, *akkor*

$$f(\tilde{x}, z) := (1 + z) \cdot g_0(\tilde{x}) + z \cdot g(\tilde{x}, g_0(\tilde{x}), 0) \quad \left((\tilde{x}, z) \in \{0,1\}^{n+1}\right)$$

esetén $f \in C^{(n+1)}$.

Bizonyítás. Minthogy C *pr-zárt*, az $f(\tilde{x}, 0) := g_0(\tilde{x})$, $f(\tilde{x}, 1) := g(\tilde{x}, f(\tilde{x}, 0), 0) = g(\tilde{x}, g_0(\tilde{x}), 0)$ ($\tilde{x} \in \{0,1\}^n$) primitív rekurzióval definiált f függvényt is tartalmazza. A 3.1. segédtétel szerint $f(\tilde{x}, z) = (1 + z)f(\tilde{x}, 0) + zf(\tilde{x}, 1) = (1 + z)g_0(\tilde{x}) + z \cdot g(\tilde{x}, g_0(\tilde{x}), 0)$. □

Minden *pr-zárt* osztály definíció szerint tartalmazza a projekció-függvényeket. Azonban ennél több is igaz, mint a következő jól használható állítás mutatja.

3.4 SEGÉDTÉTEL. *Minden pr-zárt osztály részhalmazként tartalmazza a* $\{0, x, y, xy, xy + x\}$ *függvényhalmazt.*

Bizonyítás. A 3.3. segédtételből $g_0 := e_1^1(x)$, $g := e_3^3(x, y, z)$ választással nyerjük az $xz + x$ függvényt. Ebből $z = x$ azonosítás adja a c_0^1 állandó-függvényt és az $\left((x + 1)z\right) \circ \left((x + 1)z\right) = \left((x + 1)z + 1\right)z = (xz + z + 1)z = xz$ kompozíció eredményezi a szorzat-függvényt. □

Könnyen ellenőrizhető, hogy az előbbiekben definiált függvényosztályok a 3.4. segédtétel vonatkozásában úgy viselkednek, mint a *pr-zárt* osztályok. Ezen túlmenően, tetszőleges $f \in C_1$ esetén $e_i^n \cdot f \in C_3 \bigcap\limits_{m \geq 2} F^m \cap F^\infty$. (Később megmutatjuk, hogy a szóban forgó függvényosztályok is *pr-zárt* osztályok, sőt, az összes *pr-zárt* osztályt kimerítik.) A következő állítás az F^∞ halmaz egy jellemzését adja.

3.5 TÉTEL. $F^\infty = \left\{e_i^n f_1 \mid f_1 \in C_1^{(n-1)}, 1 \leq i \leq n, n > 1\right\}$.

Bizonyítás. Indirekte feltéve, hogy $f \in F^{\infty(n)}$ nem $e_i^n \cdot f_1$ ($f_1 \in C_1$) alakú, a kifejtési tétel szerint létezik $\tilde{a}_i = (a_1, \dots, a_i, \dots, a_n) \in \{0,1\}^n$ n-es, amelyre $a_i = 0$ és $f(\tilde{a}_i) = 1$. Ezt a feltételt $1 \leq i \leq n$ lehetséges esetekre felírva, az $A_n = \{\tilde{a}_1, \dots, \tilde{a}_n\} (\subseteq A(f))$ halmaz nem tesz eleget az $\langle A^n \rangle$-feltételnek, s így az $\langle A^\infty \rangle$-feltételnek sem, ezért $f \in F^\infty$ nem teljesülhet. □

3.6 SEGÉDTÉTEL. *Minden* $m \geq 2$ *egész szám esetén* $h_m \in F^m \setminus F^{m+1}$ *és* $h_m \notin F^\infty$.

Bizonyítás. A 2.2. definíciót követő megjegyzés szerint a $h_m \in F^m$ tartalmazás nyilvánvaló. Továbbá $h_m \notin F^{m+1}$, mert $A(h_m) \setminus \{(1, \dots, 1)\}$ a $\{0,1\}^{m+1}$ halmaz olyan $(m+1)$-elemű részhalmaza, amely rendezhető úgy, hogy az i-edik elemben az i-edik helyen 0 áll, az összes többi helyen 1, $i = 1, 2, \dots, m + 1$. Végül $h_m \notin F^{m+1}$ következtében $h_m \notin F^\infty$. □

3.7 TÉTEL. *Bármely Zsegalkin-polinóm előállítható az s függvényből pr-lezárással.*

Bizonyítás. A $p(x,0) = x$, $p(x,1) = s(e_2^3(x,p(x,0),0))$ primitív rekurzióval nyert $p(x,y) = x+y$ függvény kompozíció-iteráltjaként tetszőleges homogén-lineáris *Zs.-polinóm* előállítható: $p_{n+1}(\tilde{x}, x_{n+1}) := p(p_n(\tilde{x}), x_{n+1})$. Hasonlóan, a 3.4. segédtétel szerint felhasználható $q_2(x,y) := xy$ függvény kompozíció-iterálásával adódnak a $q_{n+1}(\tilde{x}, x_{n+1}) := q_2(q_n(\tilde{x}), x_{n+1}) = x_1 \cdot \ldots \cdot x_n \cdot x_{n+1}$ szorzatfüggvények. Világos, hogy minden *Zs.-polinóm* $f = p_n(q_{i_1}, \ldots, q_{i_n})$ vagy $s(f)$ alakú. □

A 3.9 tételben a $q_0(x,y) := q_2(x, s(y))$, $q_1(x,y) := q_2(s(x), y)$, $x, y \in \{0,1\}$ definíciókkal megadott q_0 és q_1 függvényeket is használjuk.

3.8 TÉTEL. *A* C_1, C_3, F^m $(m \geq 2)$ *és* F^∞ *függvényhalmazok pr-zárt osztályok.*

Bizonyítás. A C_1 osztály zártsága nem szorul bizonyításra. Az is világos F^∞ definíciójából, hogy $F^\infty = \bigcap_{m \geq 2} F^m$, és minthogy zárt halmazok metszete zárt (tetszőleges lezárási operáció esetén, l. [9]), csak C_3 és F^m zártságát kell bizonyítani. Nyilvánvaló az is, hogy $E \subset C_3$ és $E \subset F^m$. A függvény-kompozícióra és a primitív rekurzív függvények képzésére vonatkozó zártság bizonyításához legyen $g_1, \ldots, g_n \in C$, $g_0 \in C^{(n)}$ és $g \in C^{(n+2)}$.

$C = C_3$ eset. Itt $g_0(g_1(\tilde{0}), \ldots, g_n(\tilde{0})) = g_0(0, \ldots, 0) = 0$ és a (2.3. definíció szerint) primitív rekurzióval képezett f függvényre $f(\tilde{x}, 0) = g_0(\tilde{x})$, ezért $f(0, \ldots, 0, 0) = g_0(0, \ldots, 0) = 0$, tehát C_3 *pr-zárt* osztály.

$C = F^m$ eset. Jelölje az f, g, g_0, $g_1, \ldots, g_n$, $g_0(g_1, \ldots, g_n)$ függvényekhez tartozó A-halmazokat rendre $A^{(f)}$, $A^{(g)}$, $A^{(0)}$, $A^{(1)}, \ldots, A^{(n)}$, $A^{(c)}$.

(a) Ha $A_m \subseteq A^{(j)} \cap A^{(c)}$ valamely j $(1 \leq j \leq n)$ esetén, akkor $(g_j|_{A_m} = 1$, és) a g_j definíciója által biztosított i index egyidejűleg a $g_0(g_1, \ldots, g_n)$ összetett függvényre is a kívánt tulajdonságú.

(b) Ha viszont valamely $A_m \subseteq A^{(c)}$ esetén minden g_j $(1 \leq j \leq n)$ függvényhez létezik olyan $\tilde{a} \in A_m$, amelyre $g_j(\tilde{a}) = 0$, akkor a g_0 nem elégítheti ki az $\langle A^m \rangle$-feltételt, s ez ellentmondás. Ezért csak az (a) eset létezik.

(c) Végül, a g_0 és g függvényekből (2.3. definíció szerint) primitív rekurzióval képezett f függvény A-halmaza a *pr-lemmában* megadott, f-re vonatkozó kifejezésből leolvasható: $A^{(f)} = U \cup V \cup W$, ahol $U = \{(\tilde{x}, 0) \mid \tilde{x} \in A^{(0)}\}$, $V = \{(\tilde{x}, 1) \mid \tilde{x} \in A^{(0)}$ és $(\tilde{x}, 1, 0) \in A^{(g)}\}$, és $W = \{(\tilde{x}, 1) \mid \tilde{x} \notin A^{(0)}$ és $(\tilde{x}, 0, 0) \in A^{(g)}\}$. Azonban g_0, $g \in F^m$ következtében az összetett függvényre vonatkozó meggondolás (b) része szerint W üres halmaz. Minthogy $A_m \subseteq A^{(f)}$ minden eleme $(\tilde{x}, z)$ alakú, ahol $\tilde{x} \in A^{(0)}$, így $g_0 \in F^m$ következtében f is kielégíti az $\langle A^m \rangle$-feltételt. □

A következőkben megadjuk a felsorolt *pr-zárt* osztályok egymáshoz való viszonyát és megmutatjuk, hogy a felsoroltakon kívül más *pr-zárt* osztály nem létezik. E célból minden osztályhoz megadunk egy generáló rendszert, továbbá a 3.9. tételben felsorolt sorrendben minden szomszédos párhoz egy elemet (függvényt), amely a bővebb osztálynak eleme, de a szűkebbnek nem. A további bizonyítások szempontjából szerencsés egyszerűsítő körülmény, hogy mindegyik osztály generálható egyetlen elemmel. Eredményül kapjuk, hogy a „*Post-háló*"-nak megfelelő háló egy lánc, amelynek maximális eleme C_1, minimális eleme F^∞ (azaz, a végtelen sok osztály tartalmazás szerint teljesen rendezhető).

3.9 TÉTEL.

$$(1)\ F^\infty = \left[\{e_1^1\}\right]_{pr}; \qquad (2)\ F^m = \left[\{h_m\}\right]_{pr},\ m \geq 2;$$

$$(3)\ C_3 = \left[\{p\}\right]_{pr}; \qquad (4)\ C_1 = \left[\{c_1^1\}\right]_{pr}.$$

Bizonyítás. (1) A 3.4 segédtétel és a 3.8 tétel alapján elég megmutatni, hogy a $\{q_2, q_0\}$ halmaz generálja az F^∞ osztályt. A 3.5 tétel szerint viszont az $x \cdot f_1$, $f_1 \in C_1$ alakú függvények előállíthatóságát elegendő megmutatni (az előbb említett q_2 és q_0 függvényekből). Az $f(x,y,0) = q_2(x,y)$, $f(x,y,1) = e_3^4(x,y,q_0(x,y),0) = q_0(x,y)$ primitív rekurzió előállítja az $f(x,y,z) = q_2(x,p(y,z))$ függvényt. Most már a 3.2. tétel alapján a bizonyítás a 3.7. tétel bizonyításához hasonlóan fejezhető be, a p és q_2 függvények helyett rendre az f és q_3 függvényeket alkalmazva.

(2) A 3.6. segédtétel és a 3.8. tétel alapján $[\{h_m\}]_{pr} \subseteq F^m$, ezért csak azt kell belátni, hogy $F^m \subseteq [\{h_m\}]_{pr}$, azaz, hogy $F^m \subseteq [\{h_m\} \cup F^\infty]_{pr}$. Ha $f(\tilde{x}) \in F^m$ olyan függvény, amelyre $f(x,\dots,x) = 0$ ($x \in \{0,1\}$), azaz γ-függvény, akkor az $f^*(\tilde{x}, x_{n+1}) = q(\tilde{x}, x_{n+1}) + f(\tilde{x}) + q(\tilde{x}, x_{n+1}) \cdot f(\tilde{x}) \in F^m$ függvényből az $f^*(\tilde{x}, c_0^1)$ kompozícióval nyerhető, és $f^*(x,\dots,x) = x$ ($x \in \{0,1\}$). Ezért elegendő az $f(e_1^1,\dots,e_1^1) = e_1^1$ feltételnek eleget tevő függvények (α-függvények) generálhatóságát megmutatni.

A bizonyítás $|A(f)| := j$ szerinti teljes indukcióval történik. Először megmutatjuk, hogy $j \leq m+1$ esetén $f \in F^\infty$. Minthogy $f \in F^m$ és $f(1,\dots,1) = 1$, ezért az $A(f) \setminus \{(1,\dots,1)\}$ $(j-1)$-elemű halmaz bármely m-elemű részhalmazához van olyan i index, hogy mindegyik n-es i-edik komponense 1. Ezért $j-1 \leq m$ következtében az f függvény eleget tesz az $\langle A^\infty \rangle$-feltételnek. A továbbiakban feltehetjük, hogy $j > m+1$ (s ezért $n > 2$ is teljesül).

Legyen $A(f) = \{\tilde{a}_1, \dots, \tilde{a}_j\}$, $\tilde{a}_j = (1,\dots,1)$, és

$$f_i(\tilde{x}) = \begin{cases} 1, & \text{ha } \tilde{x} \in A(f) \setminus \{\tilde{a}_i\}, \\ 0, & \text{egyébként}, \end{cases} \qquad 1 \leq i \leq j-1.$$

Világos, hogy $f_i \in F^m$ és f_i is α-függvény, továbbá $|A(f_i)| = j-1$, ezért az indukciós feltevés szerint $f_i \in [\{h_m\} \cup F^\infty]_{pr}$. Végül, az indukciós lépés az $f(\tilde{x}) = h_m(f_1(\tilde{x}), \dots, f_{m+1}(\tilde{x}))$ azonosságból adódik.

(3) A 3.7. tétel bizonyításában megmutattuk, hogy a $\{p, q_2\}$ halmaz klónlezártja a homogén *Zsegalkin-polinómok* halmaza, azaz C_3.

(4) A 3.4. segédtétel szerint felhasználható q_0 függvény és c_1^1 kompozíciója, $q_0(c_1, y) = s(y)$, ezért a 3.7. tétel szerint adódik az állítás. □

A *pr-lezárás* definíciója és a 3.9. tétel (1) állításából közvetlenül adódik:

1. KÖVETKEZMÉNY. *Minden pr-zárt osztály tartalmazza az F^∞ osztályt.* □

2. KÖVETKEZMÉNY. *A tételben előforduló pr-zárt osztályok mindegyike egy elemmel generálható.* □

Az eddigiek alapján megadhatjuk a *Boole-függvények pr-zárt* osztályai halmazának szerkezetét leíró struktúra-tételt.

3.10 TÉTEL. *(1) a. A C_1 pr-zárt osztályban C_3 az egyetlen maximális osztály. b. A C_3 pr-zárt osztályban F^2 az egyetlen maximális osztály. c. Minden $m \geq 2$ egész szám esetén az F^m pr-zárt osztályban F^{m+1} az egyetlen maximális osztály.*

(2) Az F^∞, F^m ($m \geq 2$), C_3, C_1 halmazokon kívül nem létezik a C_1 osztálynak pr-zárt részosztálya.

(3) A pr-zárt osztályok egyikében sem maximális az F^∞ osztály.

Bizonyítás. A 2.6. definíciót követő megjegyzés szerint $F^\infty \subseteq F^{m+1} \subseteq F^m \subseteq C_3 \subseteq C_1$. A 3.6. segédtétel, a 3.8 és 3.9 tételek, valamint F_2, C_3 és C_1 definíciójából adódó $p \in C_3 \setminus F^2$ és $c_1^1 \in C_1 \setminus C_3$ relációk szerint a tartalmazások valódiak.

(1) a. $f \in C_1 \setminus C_3$ esetén $[\{f\} \cup C_3]_{pr} = C_1$ állítás igaz volta jelenti azt, hogy C_3 maximális a C_1 osztályban. Az $[\{f\}]_{pr} = C_1$ egyenlőséget bizonyítjuk, s ebből következik az a. állítás. Minthogy $f(c_0^1, \ldots, c_0^1) = c_1^1$, mert $f(0, \ldots, 0) = 1$, ezért a 3.9. tétel (4) állítása alapján az egyenlőség teljesül.

b. Azt mutatjuk meg, hogy tetszőleges $f \in C_3 \setminus F^2$ függvény generálja a p függvényt. Definíció szerint $f(0\ldots, 0) = 0$, továbbá létezik $\tilde{a} := (a_1, \ldots, a_n)$ és $\tilde{b} := (b_1, \ldots, b_n)$, amelyekre $f(\tilde{a}) = 1 = f(\tilde{b})$, minden $1 \leq i \leq n$ index esetén $a_i \cdot b_i = 0$ és valamely $1 \leq i < j \leq n$ indexpár esetén $a_i = b_j = 1$, $a_j = b_i = 0$. Képezzük az $r := f(g_1, \ldots, g_n)$ kompozíciót, ahol $g_\ell := c_0^2$, ha $a_\ell = b_\ell = 0$; $g_\ell := e_1^2$ ha $a_\ell = 1$ és $b_\ell = 0$; $g_\ell := e_2^2$ ha $a_\ell = 0$ és $b_\ell = 1$. Az így kapott $r \in C_3^{(2)}$ függvényre teljesülnek az $r(0,0) = f(0, \ldots, 0) = 0$, $r(1,0) = f(\tilde{a}) = 1 = f(\tilde{b}) = r(0,1)$ feltételek, ezért $r = p$ vagy $r = v$. Az $r(x, y) = v(x, y) = x + y + xy$ esetben p a $v(q_0, q_1)$ kompozícióval adódik: $v(q_0(x, y), q_1(x, y)) = x(y+1) + (x+1)y + x(y+1)(x+1)y = x + y$. Ezért a 3.9. tétel (3) állítása teljessé teszi a bizonyítást.

c. Azt mutatjuk meg, hogy tetszőleges $f \in (F^m \setminus F^{m+1})^{(n)}$, ($m \geq 2$) függvény generálja a h_m függvényt, ahol $n \geq 2$ a 2.7. definíciót követő megjegyzés szerint. A feltevés szerint létezik az $A(f)$ halmaznak olyan $\{\tilde{a}_1, \ldots, \tilde{a}_{m+1}\}$ részhalmaza, amelyre minden $1 \leq i \leq n$ esetén $a_{1i} \cdot \ldots \cdot a_{m+1,i} = 0$. Tekintsük az $(m+1)$-változós $g_1, \ldots, g_n$ következő függvényeket:

$$g_\ell(x_1, \ldots, x_{m+1}) = \begin{cases} 0, & \text{ha létezik } 1 \leq i < j \leq m+1, \text{ hogy } x_i = x_j = 0, \\ a_{j\ell}, & \text{ha } x_j = 0, \text{ és } x_i = 1, \text{ ha } i \neq j, 1 \leq i, j \leq m+1, \\ a_{1\ell}, & \text{ha } f(1, \ldots, 1) = 0 \text{ és } x_1 = \ldots = x_{m+1} = 1, \\ 1, & \text{ha } f(1, \ldots, 1) = 1 \text{ és } x_1 = \ldots = x_{m+1} = 1. \end{cases}$$

Világos, hogy a g_ℓ függvények F^∞ elemei, mert $A(g_\ell) \leq m+1$. Könnyen látható, hogy $f(g_1, \ldots, g_n) = h_{m+1}$. A 3.9. tétel (2) állítása teljessé teszi a bizonyítást.

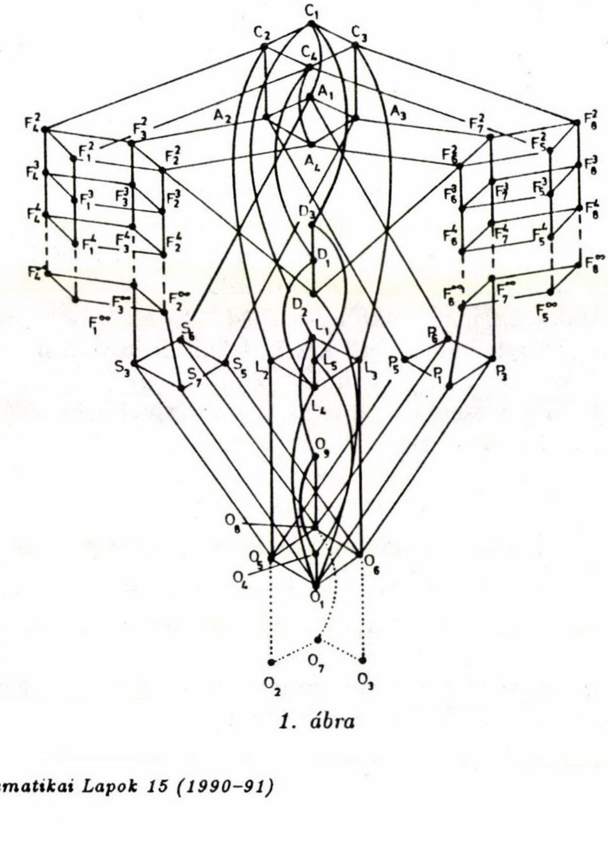
(2) Az (1) a., b. illetve c. állítások bizonyításából rendre adódik, hogy C_1 minden valódi *pr-zárt* részhalmaza C_3 részhalmaza, C_3 minden valódi *pr-zárt* részhalmaza F^2 részhalmaza, F^m minden valódi *pr-zárt* részhalmaza F^{m+1} részhalmaza, $m \geq 2$. Ezért, ha létezik egy F *pr-zárt* osztály, amelyre valamely $m \geq 2$ esetén $F^{m+1} \subseteq F \subset F^m$, akkor $F \subseteq F^{m+1}$ és így $F = F^{m+1}$. Ellenkező esetben minden $m \geq 2$ esetén F valódi részhalmaza az F^m halmaznak, ezért $F \subseteq \bigcap_{m \geq 2} F^m = F^\infty$. Másrészről, a 3.9. tétel 1. következménye szerint $F^\infty \subseteq F$. Következésképp, $F = F^\infty$.

(3) Az állítás közvetlen következménye az (1) és (2) állításoknak. □

A [11] és [2] dolgozatban felsorolt *Post-tételek* megfelelői közvetlenül vagy következményként kiolvashatók a 3.9. és 3.10. tételekből, ezért itt nem részletezzük — egyetlen kivétellel. Ez a 3.9. tétel egy élesítése, és a 3.10.1. állítás bizonyításában a szükségesnél többet, a következő állítást bizonyítottuk.

3.11 TÉTEL. *Legyen C egy pr-zárt osztály és tegyük fel, hogy C' maximális a C osztályban. Az $f \in C$ függvény pontosan akkor generálja a C osztályt, ha nem C'-beli.* □

Megjegyezzük, hogy az állítás az F^∞ osztályra is érvényes, ha C'-nek az üres halmazt tekintjük.



1. ábra

A C_1-ben levő zárt osztályok definíciója, típusa, bázisa, rendszáma [2]

Jelölés	Definíció	Típus	Bázis	Rendszám
O_1	Mindazon függvények halmaza, amelyek x-szel egyenlők;	$\langle\alpha\rangle$	$\{x\}$	1
O_2	Mindazon függvények halmaza, amelyek 1-gyel egyenlők;	$\langle\beta\rangle$	$\{1\}$	0
O_3	Mindazon függvények halmaza, amelyek 0-val egyenlők;	$\langle\gamma\rangle$	$\{0\}$	0
O_4	Mindazon függvények halmaza, amelyek egyenlők x-szel, ill. $\bar{x}$-sal;	$\langle\alpha,\delta\rangle$	$\{\bar{x}\}$	1
O_5	Mindazon függvények halmaza, amelyek egyenlők 1-gyel, ill. x-szel;	$\langle\alpha,\beta\rangle$	$\{x,1\}$	1
O_6	Mindazon függvények halmaza, amelyek egyenlők 0-val, ill. x-szel;	$\langle\alpha,\gamma\rangle$	$\{0,x\}$	1
O_7	Mindazon függvények halmaza, amelyek egyenlők 0-val, ill. 1-gyel;	$\langle\beta,\gamma\rangle$	$\{0,1\}$	0
O_8	Mindazon függvények halmaza, amelyek egyenlők 0-val, 1-gyel, x-szel;	$\langle\alpha,\beta,\gamma\rangle$	$\{x,0,1\}$	1
O_9	Mindazon függvények halmaza, amelyek egyenlők 0-val, 1-gyel, x-szel, ill. $\bar{x}$-sal;	$\langle\alpha,\beta,\gamma,\delta\rangle$	$\{\bar{x},0\}$	1
S_1	Logikai összeg függvények halmaza;	$\langle\alpha\rangle$	$\{x\vee y\}$	2
S_3	S_1-beli, ill. O_2-beli függvények halmaza;	$\langle\alpha,\beta\rangle$	$\{x\vee y,1\}$	2
S_5	S_1-beli, ill. O_3-beli függvények halmaza;	$\langle\alpha,\gamma\rangle$	$\{x\vee y,0\}$	2
S_6	S_1-beli, O_2-beli, ill. O_3-beli függvények halmaza;	$\langle\alpha,\beta,\gamma\rangle$	$\{x\vee y,0,1\}$	2
P_1	Logikai szorzat függvényének halmaza;	$\langle\alpha\rangle$	$\{xy\}$	2
P_3	P_1-beli, ill. O_3-beli függvények halmaza;	$\langle\alpha,\gamma\rangle$	$\{xy,0\}$	2
P_5	P_1-beli, ill. O_2-beli függvények halmaza;	$\langle\alpha,\beta\rangle$	$\{xy,1\}$	2
P_6	P_1-beli, O_2-beli, ill. O_3-beli függvények halmaza;	$\langle\alpha,\beta,\gamma\rangle$	$\{xy,0,1\}$	2
L_1	Lineáris függvények halmaza;	$\langle\alpha,\beta,\gamma,\delta\rangle$	$\{x+y,1\}$	2
L_2	Lineáris α és β függvények halmaza;	$\langle\alpha,\beta\rangle$	$\{x+y+1\}$	2
L_3	Lineáris α és γ függvények halmaza;	$\langle\alpha,\gamma\rangle$	$\{x+y\}$	2
L_4	Lineáris α függvények halmaza;	$\langle\alpha\rangle$	$\{x+y+z\}$	3
L_5	Lineáris, önduális függvények halmaza;	$\langle\alpha,\delta\rangle$	$\{x+y+z+1\}$	3
D_1	Önduális α függvények halmaza;	$\langle\alpha\rangle$	$\{xy\vee x\bar{z}\vee y\bar{z}\}$	3
D_2	Önduális monoton függvények halmaza;	$\langle\alpha\rangle$	$\{xy\vee xz\vee yz\}$	3
D_3	Önduális függvények halmaza;	$\langle\alpha,\delta\rangle$	$\{x\bar{y}\vee x\bar{z}\vee \bar{y}\bar{z}\}$	3
A_1	Monoton függvények halmaza;	$\langle\alpha,\beta,\gamma\rangle$	$\{xy,x\vee y,0,1\}$	2
A_2	Monoton α és β függvények halmaza;	$\langle\alpha,\beta\rangle$	$\{xy,x\vee y,1\}$	2
A_3	Monoton α és γ függvények halmaza;	$\langle\alpha,\gamma\rangle$	$\{xy,x\vee y,0\}$	2
A_4	Monoton α függvények halmaza;	$\langle\alpha\rangle$	$\{xy,x\vee y\}$	2

Jelölés	Definíció	Típus	Bázis	Rendszám
C_1	Kétértékű függvények halmaza;	$\langle\alpha,\beta,\gamma,\delta\rangle$	$\{x \mid y\}$	2
C_2	α és β függvények halmaza;	$\langle\alpha,\beta\rangle$	$\{x \vee y, x+y+1\}$	2
C_3	α és γ függvények halmaza;	$\langle\alpha,\gamma\rangle$	$\{xy, x+y\}$	2
C_4	α függvények halmaza;	$\langle\alpha\rangle$	$\{x \vee y, x(y+z+1)\}$	3
F_1^μ $(\mu \geq 2)$	Mindazon α függvények halmaza, amelyek kielégítik az $\langle a^\mu\rangle$ feltételt;	$\langle\alpha\rangle$	$\{x \vee y\bar{z},\ h_\mu^*(\bar{x})\}$	$\mu+1$
F_2^μ $(\mu \geq 2)$	Mindazon monoton α függvények halmaza, amelyek kielégítik az $\langle a^\mu\rangle$ feltételt;	$\langle\alpha\rangle$	$\{x \vee yz,\ h_2^*(\bar{x})\}$, ha $\mu = 2$, $\{h_\mu^*(\bar{x})\}$, ha $\mu \geq 3$	$\mu+1$
F_3^μ $(\mu \geq 2)$	Mindazon monoton függvények halmaza, amelyek kielégítik az $\langle a^\mu\rangle$ feltételt;	$\langle\alpha,\beta\rangle$	$\{1, h_\mu^*(\bar{x})\}$	$\mu+1$
F_4^μ $(\mu \geq 2)$	Mindazon függvények halmaza, amelyek kielégítik az $\langle a^\mu\rangle$ feltételt;	$\langle\alpha,\beta\rangle$	$\{x \vee \bar{y},\ h_\mu^*(\bar{x})\}$	$\mu+1$
F_5^μ $(\mu \geq 2)$	Mindazon α függvények halmaza, amelyek az $\langle A^\mu\rangle$ feltételt kielégítik;	$\langle\alpha\rangle$	$\{x(y \vee \bar{z}), h_\mu(\bar{x})\}$	$\mu+1$
F_6^μ $(\mu \geq 2)$	Mindazon monoton α függvények halmaza, amelyek az $\langle A^\mu\rangle$ feltételt kielégítik;	$\langle\alpha\rangle$	$\{x(y \vee z),\ h_2(\bar{x})\}$, ha $\mu = 2$, $\{h_\mu(\bar{x})\}$, ha $\mu \geq 3$;	$\mu+1$
F_7^μ $(\mu \geq 2)$	Mindazon monoton függvények halmaza, amelyek kielégítik az $\langle A^\mu\rangle$ feltételt;	$\langle\alpha,\gamma\rangle$	$\{0, h_\mu(\bar{x})\}$	$\mu+1$
F_8^μ $(\mu \geq 2)$	Mindazon függvények halmaza, amelyek kielégítik az $\langle A^\mu\rangle$ feltételt;	$\langle\alpha,\gamma\rangle$	$\{x\bar{y}, h_\mu(\bar{x})\}$	$\mu+1$
F_1^∞	Mindazon α függvények halmaza, amelyek kielégítik az $\langle a^\infty\rangle$ feltételt;	$\langle\alpha\rangle$	$\{x \vee y\bar{z}\}$	3
F_2^∞	Mindazon monoton α függvények halmaza, amelyek kielégítik az $\langle a^\infty\rangle$ feltételt;	$\langle\alpha\rangle$	$\{x \vee yz\}$	3
F_3^∞	Mindazon monoton függvények halmaza, amelyek kielégítik az $\langle a^\infty\rangle$ feltételt;	$\langle\alpha,\beta\rangle$	$\{1, x \vee yz\}$	3
F_4^∞	Mindazon függvények halmaza, amelyek az $\langle a^\infty\rangle$ feltételt kielégítik;	$\langle\alpha,\beta\rangle$	$\{x \vee \bar{y}\}$	2
F_5^∞	Mindazon α függvények halmaza, amelyek az $\langle A^\infty\rangle$ feltételt kielégítik;	$\langle\alpha\rangle$	$\{x(y \vee \bar{z})\}$	3
F_6^∞	Mindazon monoton α függvények halmaza, amelyek az $\langle A^\infty\rangle$ feltételt kielégítik;	$\langle\alpha\rangle$	$\{x(y \vee z)\}$	3
F_7^∞	Mindazon monoton függvények halmaza, amelyek az $\langle A^\infty\rangle$ feltételt kielégítik;	$\langle\alpha,\gamma\rangle$	$\{0, x(y \vee z)\}$	3
F_8^∞	Mindazon függvények halmaza, amelyek az $\langle A^\infty\rangle$ feltételt kielégítik;	$\langle\alpha,\gamma\rangle$	$\{x\bar{y}\}$	2

2. ábra

A $h^*_\mu(x_1,\dots,x_\mu,x_{\mu+1})$ függvény duális a $h_\mu(x_1,\dots,x_{\mu+1}) = \bigvee_{i=1}^{\mu+1} x_1\dots x_{i-1}x_{i+1}\dots x_{\mu+1}$ függvényhez.

IRODALOM

[1] BAGYINSZKI, J., "Connections between k-valued logics and formal languages", *Papers on Automata and Languages X, Budapest* DM 88–1 (1988), 83–91.
[2] DEMETROVICS, J., "A 2-értékű logika struktúrális vizsgálata", *Alkalmazott Matematikai Lapok* 1 (1975), 405–424.
[3] POST, E.L., "Determination of all closed systems of truth tables", *Bull. Amer. Math. Soc.* **26** (1920), 435 és 437.
[4] POST, E. L., "Introduction to a general theory of elementary propositions", *Am. J. Math.* **43** (1921), 163–185.
[5] POST, E.L., *"The two-valued iterative systems of mathematical logic"* (Annals of Math. Studies 5., Princeton Univ. Press, 1941).
[6] PÖSCHEL, R., KALUŽNYIN, L.A., *"Funktionen- und Relationen-algebren"* (Deutscher Verlag der Wissenschaften, 1979).
[7] RESCHKE, M., DENECKE, K., "Ein neuer Beweis für die Ergebnisse von E.L. Post über abgeschlossen Klassen Boolescher Funktionen", *J. Inf. Process. Cybern. EIK* **25**, 7 (1989), 361–380.
[8] ROSENBERG, I.G., *"Completeness properties of multiple-valued logic algebras", in: Computer Science and Multiple-valued Logic Ed. D.C. Rine* (North-Holland Publ. Co., 1977), pp. 144–186.
[9] SCHMIDT, E.T., "Megjegyzések a relációs adatbázis modellekben értelmezett függőségi relációkhoz", *Alk. Mat. Lapok* **8** (1982), 177–182.

[10] WHITEHEAD, A.N., RUSSELL, B., *"Principia Mathematica"* Vol 1 (Camb. Univ. Press, 1910).
[11] JABLONSZKIJ, SZ. V., GAVRILOV, G.P., KUDRJAVCEV, V.B., *"Funkcii algebrū logiki i klasszū Posta"* (Nauka, Moszkva, 1966).

(Beérkezett: 1991. március 6.)

BAGYINSZKI JÁNOS
AGRÁRTUDOMÁNYI EGYETEM
2103 GÖDÖLLŐ, PÁTER K. U. 1.

LATTICE OF PR-CLOSED CLASSES OF LOGICAL FUNCTIONS

J. BAGYINSZKI

On the power-set of the set of logical functions we define a closure operation, called pr-closure. By this closure, a set of logical functions is pr-closed if it is such a closed class of functions (in the sense of Post) containing the projections, that it is closed under primitive recursions. Without using Post's results we give the lattice of pr-closed classes, similarly to the Post-diagram. As the pr-closure is "stronger" than the clone-closure, the former yields a simpler lattice which is an infinite chain. Moreover we present theorems corresponding to Post's theorems. Investigations are motivated by research of connections between several parts of computer science.

Alkalmazott Matematikai Lapok 15 (1990–91) 303–307

REKURZIV ALGORITMUS ARMA FOLYAMATOK LIKELIHOOD FÜGGVÉNYÉNEK SZÁMOLÁSÁRA

HUHN EDIT

Szeged

A dolgozatban megadunk egy rekurziv algoritmust, amellyel *Gauss*–ARMA *folyamatok* esetén a likelihood függvény értéke tetszőleges helyen kiszámítható. Az algoritmus az ARMA folyamatok többdimenzós elemi *Gauss–folyamatokba* való beágyazásán és a Kálmán szűrésén alapul.

1. Bevezetés

Legyen $x(n)$, $E\{x(n)\} = 0$ $(n = 0, \pm 1, \ldots)$, valós értékű, reguláris *Gauss*–ARMA(p, q) *folyamat*, azaz $x(n)$ *Gauss-folyamat* és megoldása valamely

$$x(n) + a_1 x(n-1) + \cdots + a_p x(n-p) = e(n) + b_1 e(n-1) + \cdots + b_q e(n-q)$$

sztochasztikus differenciaegyenletnek, ahol $e(n)$, $E\{e(n)\} = 0$, $E\{e^2(n)\} = \sigma_e^2$ $(n = 0, \pm 1, \ldots)$ független, normális eloszlású valószínűségi változók sorozata és a

$$P(z) = \sum_{j=0}^{p} a_j z^{p-j} = 0 \quad (a_0 = 1)$$

egyenlet gyökei az egységkörön belül helyezkednek el [2].

Tegyük fel, hogy rendelkezésünkre áll a folyamat egy realizációjából származó $x(0), x(1), \ldots, x(N-1)$ N elemű statisztikai minta. [3]-ban az egzakt likelihood függvény meghatározására egy, a részlegesen megfigyelhető elemi *Gauss–folyamatok Kálmán–szűrésén* alapuló módszert ismertettünk. A fellépő szűrőegyenletek explicit megoldását azonban az általános esetben nem sikerült megadni. Az alkalmazott módszer viszont lehetőséget nyújt olyan algoritmus megadására, amellyel a likelihood függvény értéke tetszőleges helyen kiszámítható azokban az esetekben is, amikor az explicit módon nem ismert.

2. Az algoritmus megadása

Tekintsük a $\mathbf{z}(n)^T = (x_1(n), \ldots, x_p(n), t_1(n), \ldots, t_q(n)) = (\mathbf{x}(n)^T, \mathbf{t}(n)^T)$

$$\begin{aligned} x_1(n) &= x(n) \\ x_j(n) &= x_{j-1}(n+1), \quad j = 2, \ldots p \\ t_1(n) &= e(n+p-1) \\ t_j(n) &= t_{j-1}(n-1), \quad j = 2, \ldots, q \end{aligned}$$

folyamatot. $\mathbf{z}(n)$ elemi *Gauss-folyamat* [2], az $\mathbf{x}(n)$ megfigyelhető, a $\mathbf{t}(n)$ nem–megfigyelhető komponenseket tartalmaz. $x(n)$ és $\mathbf{t}(n)$ a

$$(2.1) \qquad \begin{cases} \mathbf{x}(n+1) = \mathbf{A}_0\mathbf{x}(n) + \mathbf{A}_1\mathbf{t}(n) + \mathbf{B}\mathbf{e}(n) \\ \mathbf{t}(n+1) = \qquad\qquad \mathbf{a}_1\mathbf{t}(n) + \mathbf{b}\mathbf{e}(n) \end{cases}^{2}$$

differenciaegyenlet–rendszernek az $(\mathbf{x}(0), \mathbf{t}(0))$ kezdeti értékekhez tartozó megoldása, ahol $(\mathbf{x}(0), \mathbf{t}(0))$ az $\mathbf{e}(n)$–től független $(n = 1, 2, \ldots)$ és

$$\mathbf{A}_0 = \begin{bmatrix} 0 & 1 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 1 & \ldots & 0 & 0 \\ \vdots & & & & & \\ 0 & 0 & 0 & \ldots & 0 & 1 \\ -a_p & & & & -a_2 & -a_1 \end{bmatrix}_{p\times p} \qquad \mathbf{A}_1 = \begin{bmatrix} 0 & \ldots & 0 \\ 0 & \ldots & 0 \\ \vdots & & \\ 0 & \ldots & 0 \\ b_1 & \ldots & b_q \end{bmatrix}_{p\times q}$$

$$\mathbf{B} = \begin{bmatrix} 0 & \ldots & 0 & 0 \\ \vdots & & & \\ 0 & \ldots & 0 & 0 \\ 0 & \ldots & 0 & \sigma_e \end{bmatrix}_{p\times p} \qquad \mathbf{a}_1 = \begin{bmatrix} 0 & 0 & \ldots & 0 & 0 \\ 1 & 0 & \ldots & 0 & 0 \\ \vdots & & & & \\ 0 & 0 & \ldots & 1 & 0 \end{bmatrix}_{q\times q}$$

$$\mathbf{b} = \begin{bmatrix} 0 & \ldots & 0 & \sigma_e \\ 0 & \ldots & 0 & 0 \\ \vdots & & & \\ 0 & \ldots & 0 & 0 \end{bmatrix}_{q\times p} \qquad \mathbf{e}(n) = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ e'(n+p) \end{bmatrix}_{p\times 1} \qquad e(n+p) = \sigma_e e'(n+p)$$

A (2.1) egyenleteket röviden

$$\mathbf{z}(n+1) = \mathbf{Q}\mathbf{z}(n) + \mathbf{B}_e$$

alakba írjuk.

A (2.1) első egyenlete $\mathbf{x}(n+1)$ p–edik komponensére nézve a következőt jelenti:

$$(2.2) \qquad x_p(n+1) \equiv x(n+p) = -\mathbf{a}^T\mathbf{x}(n) + \mathbf{b}^T\mathbf{t}(n) + e(n+p)\ ,$$

ahol

$$\mathbf{a}^T = (a_1, \ldots, a_p), \quad \mathbf{b}^T = (b_1, \ldots, b_q).$$

Jelölje F_n^x az $\mathbf{x}(0), \ldots, \mathbf{x}(n)$ változók által generált σ–algebrát, és legyen

$$\mathbf{m}(n) = E\left\{\mathbf{e}(n) \mid F_n^x\right\}, \quad \gamma(n) = E\left\{\left(\mathbf{e}(n) - \mathbf{m}(n)\right)\left(\mathbf{e}(n) - \mathbf{m}(n)\right)^T \mid F_n^x\right\}.$$

Mivel $x(0), \ldots, x(n+p-1), x(n+p)$ együttes eloszlása normális eloszlás, az $x(n+p)$-nek az $x(0), \ldots, x(n+p-1)$ változókra vonatkozó feltételes eloszlása is normális, amelynek paraméterei (2.2) alapján

$$m = E\left\{x_p(n+1) \mid F_n^x\right\} = -\mathbf{a}^T\mathbf{x}(n) + \mathbf{b}^T\mathbf{m}(n)$$

és

$$\sigma^2 = E\left\{\left(x_p(n+1) - m\right)^2 \mid F_n^x\right\} = \mathbf{b}^T\gamma(n)\mathbf{b} + \sigma_e^2\,.$$

Ismeretes [4], hogy $\mathbf{m}(n)$ illetve $\gamma(n)$ a

(2.3)

$$\begin{aligned}
\mathbf{m}(n+1) &= \mathbf{a}_1\mathbf{m}(n) + \\
&+ \frac{1}{\sigma_e^2 + \mathbf{b}^T\gamma(n)\mathbf{b}}\left(\mathbf{b}\mathbf{B}^T + \mathbf{a}_1\gamma(n)\mathbf{A}_1^T\right)\mathbf{P}\left(\mathbf{x}(n+1) - \mathbf{A}_0\mathbf{x}(n) - \mathbf{A}_1\mathbf{m}(n)\right)\\
\gamma(n+1) &= \mathbf{a}_1\gamma(n)\mathbf{a}_1^T + \mathbf{b}\mathbf{b}^T - \\
&- \frac{1}{\sigma_e^2 + \mathbf{b}^T\gamma(n)\mathbf{b}}\left(\mathbf{b}\mathbf{B}^T + \mathbf{a}_1\gamma(n)\mathbf{A}_1^T\right)\mathbf{P}\left(\mathbf{b}\mathbf{B}^T + \mathbf{a}_1\gamma(n)\mathbf{A}_1^T\right)^T
\end{aligned}$$

ú.n. szűrőegyenletek alapján számolhatók, ahol

$$\mathbf{P} = \{P_{ik}\}, \qquad P_{ik} = \begin{cases} 1, & i = k = q \\ 0, & \text{különben.}\end{cases}$$

A kezdeti értékek

$$\begin{aligned}
\mathbf{m}(0) &= E\{\mathbf{e}(0) \mid \mathbf{x}(0)\} = \mathbf{B}_{tx}\mathbf{B}_{xx}^{-1}\mathbf{x}(0)\,, \\
\gamma(0) &= E\left\{(\mathbf{e}(0) - \mathbf{m}(0))(\mathbf{e}(0) - \mathbf{m}(0))^T \mid \mathbf{x}(0)\right\} = \mathbf{B}_{tt} - \mathbf{B}_{tx}\mathbf{B}_{xx}^{-1}\mathbf{B}_{tx}^T\,,
\end{aligned} \tag{2.4}$$

és a

$$\mathbf{B}(0) = \begin{bmatrix} \mathbf{B}_{xx} & \mathbf{B}_{tx} \\ \mathbf{B}_{tx}^T & B_{tt} \end{bmatrix} = \operatorname{cov}\{\mathbf{z}(n), \mathbf{z}(n)\}$$

a

$$\mathbf{B}(0) = \mathbf{Q}\mathbf{B}(0)\mathbf{Q}^T + \mathbf{B}_e \tag{2.5}$$

egyenlet megoldása [1], ahol

$$\mathbf{Q} = \begin{bmatrix} \mathbf{A}_0 & \mathbf{A}_1 \\ 0 & \mathbf{a}_1 \end{bmatrix}, \quad \mathbf{B}_e = \begin{bmatrix} \mathbf{B}\mathbf{B}^T & \mathbf{B}\mathbf{b}^T \\ \mathbf{b}\mathbf{B}^T & \mathbf{b}\mathbf{b}^T \end{bmatrix}$$

Ezek után a következő módon történhet az $L\left(x(0),\ldots,x(N-1);\mathbf{a},\mathbf{b},\sigma_e^2\right)$ likelihood függvény kiszámítása:
1. Megoldjuk a (2.5) egyenletet;
2. A megoldást felhasználva kiszámítható

$$L\left(x(0),\ldots,x(p-1);\mathbf{a},\mathbf{b},\sigma_e^2\right)=\left((2\pi)^p|\mathbf{B}_{xx}|\right)^{-1/2}\exp\left\{-\frac{1}{2}\mathbf{x}(0)^T\mathbf{B}_{xx}^{-1}\mathbf{x}(0)\right\}$$

az $x(0),\ldots,x(p-1)$ változók likelihood függvénye;
3. Kiszámítjuk a (2.4) alatti $\mathbf{m}(0)$, $\gamma(0)$ kezdeti értékeket;
4. A következő rekurziv összefüggés alapján, valamint (2.3) ismeretében számolhatunk tovább:

$$\begin{aligned}&L\left(x(0),\ldots,x(n+p);\mathbf{a},\mathbf{b},\sigma_e^2\right)=\\&=L(x(n+p)\mid x(0),\ldots,x(n+p-1);\mathbf{a},\mathbf{b},\sigma_e^2)L(x(0),\ldots,x(n+p-1);\mathbf{a},\mathbf{b},\sigma_e^2)=\\&=\left\{2\pi(\mathbf{b}^t\gamma(n)\mathbf{b}+\sigma_e^2)\right\}^{-1/2}\exp\left\{-\frac{1}{2}(\mathbf{b}^T\gamma(n)\mathbf{b}+\sigma_e^2)^{-1}\left(x(n+p-1)+\mathbf{a}^T\mathbf{x}(n)-\right.\right.\\&\left.\left.-\mathbf{b}^T\mathbf{m}(n)\right)^2\right\}L(x(0),\ldots,x(n+p-1);\mathbf{a},\mathbf{b},\sigma_e^2);\end{aligned}$$

5. (2.3) alapján meghatározzuk $\mathbf{m}(n+1)$–et és $\gamma(n+1)$–et.
6. Végül a 4. és az 5. alattiakat ismételjük $n=0,\ldots,N+p-1$ esetén.

Az $x(0),\ldots,x(N-1)$ változók kovariancia–mátrixának explicit inverze csak speciális esetekben ismert [1] (autoregressziós illetve elsőrendű mozgóátlag folyamatok esetén). Ennek következtében eddig ezektől különböző folyamatok egzakt likelihood függvényét nem lehetett kiszámolni. A fenti módszer jelentősége abban van, hogy alkalmazásával viszont bármilyen ARMA(p,q) folyamat egzakt likelihood függvénye számolható.

Köszönetemet fejezem ki ARATÓ MÁTYÁSnak a kézirat átnézéséért és értékes tanácsaiért.

IRODALOM

[1] ARATÓ, M., "Linear Stochastic Systems with Constant Coefficients", *Lecture Notes in Control and Information Sciences*, vol. 45 (Springer Berlin–Heidelberg–New York, 1982).

[2] DOOB, J. L., *Stochastic Processes* (New York John Wiley & Sons, London–Chapman Hall, 1953).

[3] HUHN, E., "ARMA folyamatok egzakt sűrűségfüggvénye", *Alk. Mat. Lapok* 10 (1984), 297–304.

[4] Липцер, Р. Ш., Ширяев, А. Н., "Статистика случайных процессов", Наука (1974).

(Beérkezett: 1989. junius 16.)

(Átdolgozva beérkezett: 1990. március 6.)

HUHN EDIT
KÉE ÉLELMISZERIPARI FÖISKOLAI KAR
6724 SZEGED, MARX TÉR 7.

RECURSIVE ALGORITHM FOR EVALUATION OF THE LIKELIHOOD FUNCTION OF ARMA PROCESSES

E. HUHN

In the paper a recursive algorithm is given for the computation of the likelihood function of a gaussian ARMA process. The method can be effectively used to get the likelihood function at arbitrary values of the unknown parameters. The algorithm is based on the imbedding of the ARMA process into a multi dimensional elementary gaussian process containing non-observable components too, and on the Kalman filtering.

Alkalmazott Matematikai Lapok 15 (1990–91) 309–327

KVÁZIKONVEX ELSŐRENDŰ APPROXIMÁCIÓK

KOMLÓSI SÁNDOR*

Pécs

Az optimalizálás elmélet egyik központi problémája az optimális megoldások jellemzése, ún. optimalitási kritériumok kidolgozása. Szinte valamennyi, ezen a téren fellelhető eredmény valamilyen derivált fogalmat használ. A nemdifferenciálható optimalizálás elméletében széleskörűen alkalmaznak az immár klasszikusnak számító *gradiens, iránymenti derivált, iránymenti Dini-derivált* helyett, ezek „közelítése" gyanánt pozitív homogén (többnyire konvex) függvényeket, ún. elsőrendű approximációkat. Jelen dolgozatban kvázikonvex elsőrendű approximációkat vizsgálok és a szubdifferenciál fogalmát általánosító kváziszubdifferenciál fogalma segítségével a konvex esethez hasonló optimalitási feltételeket származtatok.

1. Bevezetés

A dolgozatban a

$$\min f(\mathbf{x}) \qquad \mathbf{x} \in C,\ C \subset R^n \tag{1.1}$$

optimalizálási feladatot vizsgálom. Legyen az a pont lehetséges megoldása a feladatnak, azaz legyen $\mathbf{a} \in C$. Tegyük fel, hogy a C halmaz lokálisan konvex az **a** pontban, azaz létezik a-nak olyan G környezete, hogy a $C \cap G$ halmaz konvex. Ebben az esetben minden $\mathbf{v} \in \mathrm{cone}(C \cap G - \mathbf{a})$ vektorra az $\{\mathbf{a} + t\,\mathbf{v},\ t \geq 0\}$ félegyenes kezdetben legalábbis C-ben halad. (cone M az M halmaz által generált kúpot jelöli.) A $\mathrm{cone}(C \cap G - \mathbf{a})$ kúp elemeit a-beli *lehetséges irányoknak* nevezzük. Mivel C lokálisan konvex **a**-ban, ezért az, hogy valamely **v** irány a-ban lehetséges irány-e csak a C halmaztól függ, a G környezettől nem. Jelölje a továbbiakban $C(\mathbf{a})$ az **a**-ban lehetséges irányok kúpját, azaz legyen

$$C(\mathbf{a}) = \mathrm{cone}(C \cap G - \mathbf{a}). \tag{1.2}$$

A

$$T_c(\mathbf{a}) = \mathrm{cl}\, C(\mathbf{a}). \tag{1.3}$$

zárt konvex kúpot a C halmaz a pontbeli *érintőkúpjának* nevezzük. (a cl jel a lezárást jelöli!)

***A dolgozat az OTKA 354. sz. pályázat keretében készült.**

Az optimalitás szükséges feltételei: Az nyilvánvaló, hogy az a pont csak abban az esetben lehet optimális megoldás, ha egyetlen a-ban lehetséges irány sem csökkenési iránya $f(\mathbf{x})$-nek a-ban. $f(\mathbf{x})$ a-beli csökkenési irányai kúpot alkotnak, jelölje ezt $A_f(\mathbf{a})$.

$$(1.4)\qquad A_f(\mathbf{a}) = \{\mathbf{v} \in R^n : \exists T > 0, \text{ hogy } f(\mathbf{a}+t\,\mathbf{v}) < f(\mathbf{a}), \text{ ha } 0 < t < T\}.$$

Az elmondottakból adódik a következő optimalitási kritérium. Ahhoz, hogy az $\mathbf{a} \in C$ pont lokális optimális megoldása legyen az (1.1) feladatnak, szükséges, hogy teljesüljön a következő feltétel:

$$(1.5)\qquad A_f(\mathbf{a}) \cap C(\mathbf{a}) = \emptyset.$$

Az optimalitásnak ez a viszonylag egyszerű szükséges feltétele azonban általában nem jól kezelhető, teljesülését szinte lehetetlen ellenőrízni. Bizonyos esetekben — pl. amikor a célfüggvénynek az a pont egy környezetében valamilyen speciális tulajdonsága van — az (1.5) feltétel „approximálható" könnyebben kezelhető feltételekkel. Ilyen approximációkhoz többnyire a célfüggvény approximációi által jutunk, melyek általában a differenciálhatóság különböző változataira ill. általánosításaira épülnek. Csak példaként említünk két „klasszikus" és egy „modern" derivált fogalmat.

$$(1.6)\qquad f^I(\mathbf{a};\mathbf{d}) = \lim_{t\downarrow 0}\frac{f(\mathbf{a}+t\,\mathbf{d}-f(\mathbf{a}),}{t}$$

$$(1.7)\qquad f^D(\mathbf{a};\mathbf{d}) = \limsup_{t\downarrow 0}\frac{f(\mathbf{a}+t\,\mathbf{d}-f(\mathbf{a})}{t}$$

$$(1.8)\qquad f^C(\mathbf{a};\mathbf{d}) = \limsup_{\substack{\mathbf{z}\to\mathbf{a}\\ t\downarrow 0}}\frac{f(\mathbf{z}+t\,\mathbf{d})-f(\mathbf{z})}{t}$$

$f^I(\mathbf{a};\mathbf{d})$ a jól ismert *iránymenti derivált*, $f^D(\mathbf{a};\mathbf{d})$ a *Dini-féle iránymenti derivált* és $f^C(\mathbf{a};\mathbf{d})$ pedig a *Clarke-féle derivált*. Számos további derivált fogalmat ill. változatot is említhetnénk, számunkra azonban a továbbiak szempontjából csupán néhány közös tulajdonságuk játszik központi szerepet.

Jelölje $f^K(\mathbf{a};\mathbf{d})$ az $f(\mathbf{x})$ függvény valamely általánosított iránymenti deriváltját az a pontban. Valamennyi általánosított deriváltfogalomra jellemző a következő két tulajdonság:

(Di) $f^K(\mathbf{a};\mathbf{d})$ $\mathbf{d}$ –nek pozitív homogén függvénye,

(Dii) $f^K(\mathbf{a};\mathbf{d}) < 0 \implies \mathbf{d} \in A_f(\mathbf{a})$.

Nevezzük az $f^K(\mathbf{a};\mathbf{d}) < 0$ feltételnek eleget tevő d irányokat a *K-deriválásra nézve csökkenési irányoknak*. Erre az elnevezésre a (Dii) tulajdonság ad alapot. Legyen:

$$A_f^K(\mathbf{a}) = \{\mathbf{d} \in R^n : f^K(\mathbf{a};\mathbf{d}) < 0\}.$$

A (Di) tulajdonság folytán az $A_f^K(\mathrm{a})$ halmaz kúp, (Dii) miatt pedig

$$A_f^K(\mathrm{a}) \subset A_f(\mathrm{a}). \tag{1.9}$$

Ebből a tartalmazásból az következik, hogy az (1.5) optimalitási feltétel teljesüléséhez szükséges az

$$A_f^K(\mathrm{a}) \cap C(\mathrm{a}) = \emptyset \tag{1.10}$$

feltétel teljesülése. Ahhoz tehát, hogy az $\mathrm{a} \in C$ pont lokális optimális megoldása legyen az (1.1) feladatnak szükséges, hogy teljesüljön az (1.10) feltétel. Mivel mindenféle derivált az adott függvény adott pontbeli approximációjának tekinthető, ezért vezessük be a következő fogalmat.

1.1. Definició. Az $f^K(\mathrm{a};\mathrm{d})$, $\mathrm{d} \in R^n$ függvényt az $f(\mathrm{x})$ függvény a pontbeli *elsőrendű approximációjának* nevezzük, ha rendelkezik a (Di) és (Dii) tulajdonságokkal.

Az elsőrendű approximáció fogalma — a mostanitól kissé eltérő módon — A.D. IOFFÉtől ered [6]. A dolgozat további fejezeteiben az (1.10) feltétel geometriai vonatkozásaira mutatunk rá speciális típusú elsőrendű approximációk esetén. A lineáris ill. konvex elsőrendű approximációkra vonatkozó eredmények jól ismertek. Áttekintésük inkább didaktikai indíttatású. Újak viszont a kvázikonvex elsőrendű approximációkra vonatkozó megállapítások.

2. Lineáris és konvex elsőrendű approximációk

Tegyük fel, hogy az $f(\mathrm{x})$ függvény a pontbeli $f^K(\mathrm{a},\mathrm{d})$ elsőrendű approximációja lineáris, azaz létezik egyetlen olyan $\mathrm{g} \in R^n$ vektor, hogy

$$f^K(\mathrm{a};\mathrm{d}) = (\mathrm{g},\mathrm{d}) \quad \text{minden} \quad \mathrm{d} \in R^n - \text{re}. \tag{2.1}$$

((g,d) a g és d vektorok belső szorzatát jelöli.) Jól ismert, hogy ha $f(\mathrm{x})$ *Frèchet-* ill. *Gateaux-differenciálható* az a pontban, akkor létezik $f(\mathrm{x})$-nek a-ban lineáris approximációja. Mivel a most vizsgált esetben $f^K(\mathrm{a};\mathrm{d})$ d-nek folytonos függvénye, ezért $A^K{}_f(\mathrm{a})$ nyílt halmaz. Ebből fakadóan az (1.10) feltétel ekvivalens a következővel:

$$A^K{}_f(\mathrm{a}) \cap T_c(\mathrm{a}) = \emptyset. \tag{2.2}$$

2.1. Megjegyzés. A (2.2) és az (1.10) feltételek minden olyan esetben ekvivalensek, amikor az $A^K{}_f(\mathrm{a})$ halmaz nyílt.

Az eléggé nyilvánvaló, hogy a (2.2) feltétel ekvivalens az

$$f^K(\mathrm{a};\mathrm{d}) \geq 0, \quad \text{minden} \quad \mathrm{d} \in T_c(\mathrm{a}) \quad \text{esetén} \tag{2.3}$$

feltétellel, ami esetünkben a következőt jelenti: $(-\mathbf{g}, \mathbf{d}) \leq 0$ minden $\mathbf{d} \in T_c(\mathbf{a})$ esetén. Ez a feltétel a poláris kúp fogalmának a segítségével a következő egyszerű alakot ölti:

$$-\mathbf{g} \in N_c(\mathbf{a}) := T_c(\mathbf{a})^0, \tag{2.4}$$

ahol

$$T_c(\mathbf{a})^0 := \{\mathbf{z} \in R^n : (\mathbf{z}, \mathbf{d}) \leq 0 \text{ minden } \mathbf{z} \in T_c(\mathbf{a}) \text{ esetén}\}.$$

Az $N_c(\mathbf{a})$ kúpot a C halmaz a pontbeli normális kúpjának nevezzük.

Lineáris esetben tehát az (1.10) ill. (2.2) optimalitási feltétele a $-\mathbf{g}$ „antigradiens" és az $N_c(\mathbf{a})$ kúp egy nagyon sajátos geometriai kapcsolatát fejezi ki.

2.2. TÉTEL. *Ha $f^K(\mathbf{a}; \mathbf{d})$ lineáris elsőrendű approximációja $f(\mathbf{x})$-nek $\mathbf{a}$-ban $-\mathbf{g}$ antigradienssel, akkor $A^K{}_f(\mathbf{a}) \cap T_c(\mathbf{a}) = \emptyset$ akkor és csak akkor teljesül, ha $-\mathbf{g} \in N_c(\mathbf{a})$.*

További érdekes — és az (1.1) feladat szempontjából is hasznos — megállapításokat tehetünk a $-\mathbf{g}$ antigradiens és az $N_c(\mathbf{a})$ normális kúp kapcsolatáról, ha az eddigi vizsgálatokat kiegészítjük az

$$A^K{}_f(\mathbf{a}) \cap T_c(\mathbf{a}) \neq \emptyset \tag{2.5}$$

feltétel vizsgálatával. Ez a feltétel a 2.2 tétel szerint ekvivalens a

$$-\mathbf{g} \notin N_c(\mathbf{a}) \tag{2.6}$$

feltétellel. Ennek a feltételnek a számunkra érdekes következményeihez a konvex analízis következő jól ismert tétele alapján juthatunk.

2.3. SEGÉDTÉTEL [8, Theorem 4.7]. *Legyen $M \subset R^n$ zárt konvex halmaz és legyen $\mathbf{b} \in R^n, \mathbf{b} \notin M$. Ekkor az M halmaznak van egyetlen a $\mathbf{b}$ vektorhoz legközelebb fekvő eleme, melyet jelöljön $\mathbf{u}$. Ekkor a $\mathbf{d} = \mathbf{b} - \mathbf{u}$ vektorra fennállnak a következő egyenlőtlenségek:*

$$(\mathbf{d}, \mathbf{z}) \leq (\mathbf{d}, \mathbf{b}) - (\mathbf{d}, \mathbf{d}) < (\mathbf{d}, \mathbf{b}) \quad \textit{minden} \quad \mathbf{z} \in M \quad \textit{vektorra,} \tag{2.7}$$

mely ekvivalens azzal, hogy a $(\mathbf{d}, \mathbf{x}) = \delta$ egyenletű hipersík minden

$$(\mathbf{d}, \mathbf{b}) - (\mathbf{d}, \mathbf{d}) \leq \delta \leq (\mathbf{d}, \mathbf{b})$$

feltételnek eleget tevő δ valós számra szeparálja az M halmazt és a $\mathbf{b}$ vektort. Ha az M halmaz kúp, akkor a (2.7) feltétel csak úgy állhat fenn, hogy $(\mathbf{d}, \mathbf{z}) \leq 0$ minden $\mathbf{z} \in M$ esetén, következésképpen $\mathbf{d} \in M^0$. Igaz továbbá, hogy $(\mathbf{d}, \mathbf{u}) = 0$ és $(\mathbf{d}, \mathbf{b}) = (\mathbf{d}, \mathbf{d}) > 0$.

Alkalmazzuk a 2.3. segédtétel megállapításait az $N_c(\mathbf{a})$ zárt konvex kúpra és a $-\mathbf{g}$ vektorra. Tegyük fel, hogy $-\mathbf{g} \notin N_c(\mathbf{a})$. Jelölje $\mathbf{u}$ az $N_c(\mathbf{a})$ halmaz $-\mathbf{g}$-hez legközelebbi elemét. Legyen:

$$\mathbf{s} = \frac{-\mathbf{g} - \mathbf{u}}{\|-\mathbf{g} - \mathbf{u}\|}. \tag{2.8}$$

2.4. TÉTEL. *Tegyük fel, hogy* $A^K{}_f(\mathbf{a}) \cap T_c(\mathbf{a}) \neq \emptyset$. *Ekkor*

$$\mathbf{d} \in A^K{}_f(\mathbf{a}) \cap T_c(\mathbf{a}) \tag{2.9}$$

akkor és csak akkor teljesül, ha a $(\mathbf{d}, \mathbf{x}) = 0$ *egyenletű hipersík oly módon szeparálja az* $N_c(\mathbf{a})$ *kúpot és az antigradienst, hogy*

$$(-\mathbf{g}, \mathbf{d}) > 0. \tag{2.10}$$

Igaz továbbá, hogy

$$\mathbf{s} \in A^K{}_f(\mathbf{a}) \cap T_c(\mathbf{a}) \cap S, \tag{2.11}$$

illetve, hogy $\mathbf{p} \in A^K{}_f(\mathbf{a}) \cap T_c(\mathbf{a}) \cap S$, $\mathbf{s} \neq \mathbf{p}$ *esetén*

$$-\| -\mathbf{g} - \mathbf{u} \| = f^K(\mathbf{a}; \mathbf{s}) < f^K(\mathbf{a}; \mathbf{p}), \tag{2.12}$$

ahol $S = \{\mathbf{d} \in R^n : \|\mathbf{d}\| = 1\}$.

Bizonyítás. A 2.3. segédtételből közvetlenül adódik, hogy

$$\mathbf{s} \in N_c(\mathbf{a})^0 \quad \text{és} \quad (\mathbf{s}, -\mathbf{g}) = -f^K(\mathbf{a}; \mathbf{s}) > 0.$$

Mivel zárt konvex kúp bipolárisa maga az adott kúp [11, Theorem 14.1], ezért $N_c(\mathbf{a})^0 = T_c(\mathbf{a})$, következésképpen (2.11) valóban fennáll. A 2.3. segédtételből következik továbbá, hogy $(\mathbf{s}, -\mathbf{g}) = \| -\mathbf{g} - \mathbf{u} \|$, következésképpen $f^K(\mathbf{a}; \mathbf{s}) = -\| -\mathbf{g} - \mathbf{u} \|$. Legyen most $\mathbf{p} \in A^K{}_f(\mathbf{a}) \cap T_c(\mathbf{a}) \cap S$ tetszőleges, de $\mathbf{s} \neq \mathbf{p}$. $\mathbf{p} \in A^K{}_f(\mathbf{a})$ miatt $(-\mathbf{g}, \mathbf{p}) = -f^K(\mathbf{a}; \mathbf{p}) > 0$. A $\mathbf{p} \in T_c(\mathbf{a})$ feltételből pedig az adódik, hogy $(\mathbf{p}, \mathbf{u}) \leq 0$, hiszen $\mathbf{u} \in N_c(\mathbf{a}) = T_c(\mathbf{a})^0$. Ebből nyílvánvaló, hogy

$$0 < -f^K(\mathbf{a}; \mathbf{p}) = (-\mathbf{g}, \mathbf{p}) \leq (-\mathbf{g} - \mathbf{u}, \mathbf{p}). \tag{2.13}$$

Mivel $\mathbf{s} \neq \mathbf{p}$, ezért a *Cauchy-Schwarz egyenlőtlenség* szerint

$$(-\mathbf{g} - \mathbf{u}, \mathbf{p}) < \| -\mathbf{g} - \mathbf{u} \| = -f^K(\mathbf{a}; \mathbf{s}).$$

Ezt az egyenlőtlenséget (2.13)-mal egybevetve kiadódik (2.12). Legyen $\mathbf{d} \in A^K{}_f(\mathbf{a}) \cap T_c(\mathbf{a})$. Ekkor $\mathbf{d} \in A^K{}_f(\mathbf{a})$ miatt $(-\mathbf{g}, \mathbf{d}) > 0$; $\mathbf{d} \in T_c(\mathbf{a})$ miatt pedig $(\mathbf{z}, \mathbf{d}) \leq 0$ minden $\mathbf{z} \in N_c(\mathbf{a}) = T_c(\mathbf{a})^0$ esetén. Ezek az összefüggések éppen azt fejezik ki, hogy a $(\mathbf{d}, \mathbf{x}) = 0$ egyenletű hipersík a kívánt módon szeparálja a $-\mathbf{g}$ antigradienst és az $N_c(\mathbf{a})$ kúpot.

Fordítva: tegyük fel, hogy a $(\mathbf{d}, \mathbf{x}) = 0$ hipersík úgy szeparálja $-\mathbf{g}$-t $N_c(\mathbf{a})$-tól, hogy $(-\mathbf{g}, \mathbf{d}) > 0$. Ekkor (2.1) miatt $\mathbf{d} \in A^K{}_f(\mathbf{a})$. A szeparálás miatt érvényes továbbá minden $\mathbf{z} \in N_c(\mathbf{a})$ vektorra a $(\mathbf{d}, \mathbf{z}) \leq 0$ egyenlőtlenség, amiből $\mathbf{d} \in N_c(\mathbf{a})^0 = T_c(\mathbf{a})$ következik. Ennélfogva az adott $\mathbf{d}$ vektorra (2.9) valóban teljesül. Q.E.D.

Az imént bizonyított tételben szereplő szeparálhatósági tulajdonság valójában a szigorú szeparálhatósággal azonos, amint ezt a következő definíció és segédtétel mutatja.

2.5. Definició. A C_1 és C_2 halmazokat *szigorúan szeparálhatóknak* nevezzük, ha létezik olyan $(\mathbf{d}, \mathbf{x}) = \delta$ egyenletű hipersík és olyan $\beta > 0$ szám, hogy a $C_1 + \beta B$ és a $C_2 + \beta B$ halmazok az adott hipersík által meghatározott különböző félterekben helyezkednek el. Itt B az R^n tér egységgömbjét jelöli, azaz $B = \{d \in R^n : \|\mathbf{d}\| = 1\}$.

2.6. SEGÉDTÉTEL [11, Theorem 11.1 és 11.4]. *A C_1 és C_2 nem üres halmazok akkor és csak akkor szigorúan szeparálhatók, ha van olyan $\mathbf{d} \in R^n$, hogy*

$$\inf\{(\mathbf{u}, \mathbf{d}) : \mathbf{u} \in C_1\} > \sup\{(\mathbf{z}, \mathbf{d}) : \mathbf{z} \in C_2\}. \tag{2.14}$$

A C_1 és C_2 nem üres konvex halmazok akkor és csak akkor szigorúan szeparálhatók, ha

$$0 < \inf\{\|\mathbf{u} - \mathbf{z}\| : \mathbf{u} \in C_1, \mathbf{z} \in C_2\}, \tag{2.15}$$

vagy, ami ezzel ekvivalens

$$0 \notin \mathrm{cl}(C_1 - C_2). \tag{2.16}$$

Vizsgáljuk most azt az esetet, amikor az $f^K(\mathbf{a}, \mathbf{d})$ elsőrendű approximáció $\mathbf{d}$-nek konvex függvénye. Jól ismert, hogy konvex, maximum-, ill. kvázidifferenciálható $f(\mathbf{x})$ függvény esetén az $f^I(\mathbf{a}, \mathbf{d})$ iránymenti derivált konvex elsőrendű approximáció [9]. Ismert továbbá, hogy *lokálisan Lipschitz függvényekre* az $f^C(\mathbf{a}; \mathbf{d})$ *Clarke derivált* ugyancsak konvex elsőrendű approximáció [1]. Nem derivált-típusú konvex elsőrendű approximációkra vonatkozó eredmények találhatók a [4, 5, 6, 10] dolgozatokban.

Mivel $f^K(\mathbf{a}, \mathbf{d})$ mindenütt értelmezett konvex függvény, ezért folytonos. Figyelembe véve még az $f^K(\mathbf{a}; \mathbf{d})$ függvény pozitív homogenitását, a konvex analízis egyik sarkalatos tétele szerint $f^K(\mathbf{a}; \mathbf{d})$ egy konvex kompakt halmaz tartó függvénye.

[11, Theorem 13.2] Jelölje $\partial^K f(\mathbf{a})$ azt a halmazt, melyet az $f(\mathbf{x})$ függvény a pontbeli *K-szubdifferenciáljának* nevezünk. A $\partial^K f(\mathbf{a})$ szubdifferenciál és az $f^K(\mathbf{a}; \mathbf{d})$ approximáció kapcsolatát a következő összefüggés fejezi ki.

$$f^K(\mathbf{a}; \mathbf{d}) = \max\{(\mathbf{d}, \mathbf{g}) : \mathbf{g} \in \partial^K f(\mathbf{a})\}. \tag{2.17}$$

Az irodalomból jól ismert, hogy a (2.5) feltétel ebben az esetben a $-\partial^K f(\mathbf{a})$ konvex kompakt halmaznak az $N_c(\mathbf{a})$ zárt konvex kúptól való szigorú szeparálhatóságával ekvivalens.

2.7. TÉTEL. *Legyen $f^K(\mathbf{a}; \mathbf{d})$ konvex elsőrendű approximációja az $f(\mathbf{x})$ függvénynek az $\mathbf{a}$ pontban. Ekkor a következő feltételek páronként ekvivalensek:*

(*i*) $\quad -\partial^K f(\mathbf{a}) \cap N_c(\mathbf{a}) = \emptyset$,

(*ii*) $\quad A^K{}_f(\mathbf{a}) \cap T_c(\mathbf{a}) \neq \emptyset$,

(*iii*) a $-\partial^K f(\mathbf{a})$ *és az* $N_c(\mathbf{a})$ *halmazok szigorúan szeparálhatóak,*

(*iv*) $\quad 0 < m = \inf\{\|\mathbf{z} - \mathbf{g}\| : \mathbf{z} \in N_c(\mathbf{a}),\ \mathbf{g} \in \partial^K f(\mathbf{a})\}$,

(*v*) $\quad 0 \notin N_c(\mathbf{a}) + \partial^K f(\mathbf{a})$.

Bizonyítás. A (iii), (iv) és (v) feltételek ekvivalenciája következik a 2.6. segédtételből, azzal a megjegyzéssel, hogy a $\partial^K f(\mathbf{a})$ szubdifferenciál korlátossága miatt az $N_c(\mathbf{a}) + \partial^K f(\mathbf{a})$ halmaz zárt halmaz, ezért nincs szükség a (2.16) formulában szereplő lezárás operációra. A $\partial^K f(\mathbf{a})$ kompaktsága miatt (i) és (iv) ugyancsak ekvivalensek.

Most megmutatjuk, hogy (ii) és (iii) is ekvivalensek. Legyen $\mathbf{d} \in A^K{}_f(\mathbf{a}) \cap T_c(\mathbf{a})$. Ekkor $\inf\{(\mathbf{d}, -\mathbf{g}) : -\mathbf{g} \in -\partial^K f(\mathbf{a})\} = -f^K(\mathbf{a}; \mathbf{d}) > 0 \geq \sup\{(\mathbf{d}, \mathbf{z}) : \mathbf{z} \in N_c(\mathbf{a})\}$. Ez az összefüggés azonban a 2.6. segédtétel szerint azt jelenti, hogy az $N_c(\mathbf{a})$ és a $-\partial^K f(\mathbf{a})$ halmazok szigorúan szeparálhatók, vagyis $(ii) \implies (iii)$. Tegyük most fel azt, hogy (iii) teljesül. A 2.6. segédtétel szerint ekkor van olyan $\mathbf{d}$ vektor, hogy

$$\begin{aligned}(2.18)\qquad \inf\{(\mathbf{d}, -\mathbf{g}) : \mathbf{g} \in \partial^K f(\mathbf{a})\} &= -f^K(\mathbf{a}; \mathbf{d}) > \mu = \\ &= \sup\{(\mathbf{d}, \mathbf{z}) : \mathbf{z} \in N_c(\mathbf{a})\}.\end{aligned}$$

Mivel $N_c(\mathbf{a})$ kúp, ezért (2.18) csak úgy teljesülhet, ha $\mu = 0$. Ebből viszont azonnal $f^K(\mathbf{a}; \mathbf{d}) < 0$ és $\mathbf{d} \in T_c(\mathbf{a})$ következik, vagyis $(iii) \implies (ii)$. Q.E.D.

A 2.3. segédtétel állítása kiterjeszthető tetszőleges M zárt konvex halmaz és B konvex kompakt halmaz esetére; nevezetesen az $N_c(\mathbf{a})$ zárt konvex halmaznak van a $\partial^K f(\mathbf{a})$ konvex kompakt halmazhoz legközelebb fekvő eleme, egészen pontosan léteznek olyan $\mathbf{u}_0 \in N_c(\mathbf{a})$ és $-\mathbf{g}_0 \in -\partial^K f(\mathbf{a})$ vektorok, hogy

$$(2.19)\qquad 0 < \|-\mathbf{g}_0 - \mathbf{u}_0\| < \|-\mathbf{g} - \mathbf{z}\|$$

bármely $-\mathbf{g} \in -\partial^K f(\mathbf{a})$ és $\mathbf{z} \in N_c(\mathbf{a})$ esetén, hacsak a $\mathbf{g}_0 \neq \mathbf{g}$, $\mathbf{u}_0 \neq \mathbf{z}$ feltételek közül akárcsak az egyik is teljesül. Legyen

$$(2.20)\qquad \mathbf{s} = \frac{-\mathbf{g}_0 - \mathbf{u}_0}{\|-\mathbf{g}_0 - \mathbf{u}_0\|}.$$

Ekkor érvényes a 2.4. tétel analogonja.

2.8. TÉTEL. $\mathbf{s} \in A^K{}_f(\mathbf{a}) \cap T_c(\mathbf{a}) \cap S$, *továbbá, ha* $\mathbf{p} \in A^K{}_f(a) \cap T_c(\mathbf{a}) \cap S$ *és* $\mathbf{p} \neq \mathbf{s}$, *akkor*

$$(2.21)\qquad f^K(\mathbf{a}; \mathbf{p}) > f^K(\mathbf{a}; \mathbf{s}) = -\|-\mathbf{g}_0 - \mathbf{u}_0\|.$$

A legmeredekebb csökkenés iránya: az $\mathbf{s}$ vektort a (2.21) összefüggésben megnyilvánuló szélsőérték tulajdonsága miatt az $f(\mathbf{x})$ függvény *legmeredekebb csökkenési irányának* nevezzük az $f^K(\mathbf{a}; \mathbf{d})$ approximációra nézve az $\mathbf{a}$ pontban.

Az optimalitás szükséges feltétele: A (2.2) feltétel és a 2.7. tétel alapján nyilvánvaló a következő állítás helyessége.

2.9. TÉTEL. *Legyen* $f^K(\mathbf{a}; \mathbf{d})$ *konvex elsőrendű approximációja az* $f(\mathbf{x})$ *függvénynek az* $\mathbf{a}$ *pontban. Ekkor ahhoz, hogy az* $\mathbf{a}$ *pont optimális megoldása legyen az (1.1) feladatnak szükséges, hogy teljesüljön a*

$$(2.22)\qquad 0 \in N_c(\mathbf{a}) + \partial^K f(\mathbf{a})$$

feltétel.

3. Kvázikonvex elsőrendű approximációk

Vizsgáljuk most azt az esetet, amikor az $f^K(\mathrm{a};\mathrm{d})$ elsőrendű approximáció d-nek kvázikonvex függvénye.

A $h(\mathrm{d})$, $\mathrm{d} \in R^n$ függvényt akkor nevezzük *kvázikonvexnek*, ha bármely valós c-re a $\{\mathrm{d} : h(\mathrm{d}) \leq c\}$ alsó nívóhalmaz konvex.

Ismert, hogy kvázikonvex függvény esetén az $f^D(\mathrm{a};\mathrm{d})$ *Dini-derivált* kvázikonvex elsőrendű approximáció [2, Proposition 16]. A kvázikonvexítás — mint az elnevezés is sugallja — a konvexítás függvénytulajdonság egyfajta általánosítása. Sajnos az általánosítás során a konvex függvények több jó tulajdonsága megszünik. Ilyen megszünő tulajdonság a folytonosság. Ebből adódóan egy kvázikonvex elsőrendű approximáció már nem szükségképpen folytonos. Az ebből származó hátrányokat mérséklendő, a továbbiakban csak olyan kvázikonvex approximációkat fogunk vizsgálni, melyek rendelkeznek a következő két tulajdonsággal:

3.1. Definició. Az $f^K(\mathrm{a};\mathrm{d})$ kvázikonvex approximációt *regulárisnak* nevezzük, ha rendelkezik a következő két tulajdonsággal:

(Ri) $f^K(\mathrm{a};\mathrm{d})$ d-nek alulról félig folytonos függvénye,

(Rii) az $A^K{}_f(\mathrm{a})$ kúp nyílt halmaz.

Hogy a most vizsgált esetben is lehetséges a (2.2) ill. a (2.5) feltételeknek a konvex esethez hasonló geometriai interpretációja, az a pozitív homogén kvázikonvex függvények *Crouzeix-féle reprezentációján* alapszik.

A CROUZEIX-FÉLE REPREZENTÁCIÓS TÉTEL: [2,7] Legyen $f^K(\mathrm{a};\mathrm{d})$ d-nek pozitív homogén alulról félig folytonos kvázikonvex függvénye. Ekkor léteznek olyan $\partial^K_- f(\mathrm{a})$ illetve $\partial^K_+ f(\mathrm{a})$ nem korlátos zárt ill. kompakt konvex halmazok, hogy

$$(3.1\mathrm{a})\qquad f^K(\mathrm{a};\mathrm{d}) = \begin{cases} \sup\{(\mathrm{d},\mathrm{g}) : \mathrm{g} \in \partial^K_- f(\mathrm{a})\}, & \text{ha } \mathrm{d} \in clA^K{}_f(\mathrm{a}) \\ \max\{(\mathrm{d},\mathrm{z}) : \mathrm{z} \in \partial^K_+ f(\mathrm{a})\}, & \text{ha } \mathrm{d} \notin clA^K{}_f(\mathrm{a}) \end{cases}$$

$$(3.1\mathrm{b})\qquad \sup\{(\mathrm{d},\mathrm{g}) : \mathrm{g} \in \partial^K_- f(\mathrm{a})\} = \infty, \quad \text{ha } \mathrm{d} \notin \mathrm{cl}\, A^K{}_f(\mathrm{a}),$$

$$(3.1\mathrm{c})\qquad \max\{(\mathrm{d},\mathrm{z}) : \mathrm{z} \in \partial^K_+ f(\mathrm{a})\} = 0, \quad \text{ha } \mathrm{d} \in \mathrm{cl}\, A^K{}_f(\mathrm{a}).$$

3.2. Definíció. A $\partial^K_- f(\mathrm{a})$ nemkorlátos zárt konvex halmazt az $f(\mathbf{x})$ függvény a-beli *kvázíszubdifferenciáljának* nevezzük.

A továbbiakban megmutatjuk, hogy a (2.5) feltétel geometriai háttere a most vizsgált esetben is az $N_c(\mathrm{a})$ kúp és a $-\partial^K_- f(\mathrm{a})$ *antikvázíszubdifferenciál* szigorú szeparálhatósága.

Mivel konvex approximáció egyúttal kvázikonvex approximáció is, ezért konvex approximációnak létezik a $\partial^K f(\mathrm{a})$ szubdifferenciálja is és a $\partial^K_- f(\mathrm{a})$ kvázíszubdifferenciálja is. A kettő azonban soha nem esik egybe. Az első ugyanis korlátos halmaz, a másik pedig nem korlátos. Ismert azonban a közöttük lévő két összefüggés

[3, Proposition 3.4]:

$$\partial_-^K f(\mathbf{a}) = \{\cup t\partial^K f(\mathbf{a}) : t \geq 1\}, \tag{3.2a}$$

$$\partial^K f(\mathbf{a}) = \partial_-^K f(\mathbf{a}) \cap \partial_+^K f(\mathbf{a}). \tag{3.2b}$$

A kváziszubdifferenciállal kapcsolatban további eredmények találhatók a [2,3,7] cikkekben.

3.3. TÉTEL. *Legyen $f^K(\mathbf{a};\mathbf{d})$ reguláris kvázikonvex approximációja az $f(\mathbf{x})$ függvénynek az $\mathbf{a}$ pontban. Ekkor a következő feltételek ekvivalensek:*

(*i*) $A^K{}_f(\mathbf{a}) \cap T_c(\mathbf{a}) \neq \emptyset$,

(*ii*) $0 \notin \mathrm{cl}\{N_c(\mathbf{a}) + \partial_-^K f(\mathbf{a})\}$,

(*iii*) $\mathbf{a}$ $-\partial_-^K f(\mathbf{a})$ *és az* $N_c(\mathbf{a})$ *halmazok szigorúan szeparálhatók,*

(*iv*) $m = \inf\{\|\mathbf{z} - \mathbf{g}\| : \mathbf{z} \in N_c(\mathbf{a}),\ \mathbf{g} \in \partial_-^K f(\mathbf{a})\} > 0.$

Bizonyítás. A 2.6. Segédtétel alapján nyilvánvaló a (*ii*), (*iii*) és (*iv*) feltételek ekvivalenciája. Most bebizonyítjuk az (*i*) $\Longleftrightarrow$ (*iv*) ekvivalenciát. Induljunk ki először az $A^K{}_f(\mathbf{a}) \cap T_c(\mathbf{a}) \neq \emptyset$ feltételből. Legyen $\mathbf{d} \in A^K{}_f(\mathbf{a}) \cap T_c(\mathbf{a})$ és legyenek $\mathbf{g} \in \partial_-^K f(\mathbf{a})$ ill. $\mathbf{u} \in N_c(\mathbf{a})$ tetszőlegesek. Mivel $A^K{}_f(\mathbf{a}) \cap T_c(\mathbf{a})$ kúp, ezért feltehetjük, hogy

$$\mathbf{d} \in A^K{}_f(\mathbf{a}) \cap T_c(\mathbf{a}) \cap S.$$

A Cauchy-Schwarz egyenlőtlenség, az $\mathbf{u} \in N_c(\mathbf{a})$ feltétel és (3.1a) alapján igazak a következő egyenlőtlenségek:

$$\|-\mathbf{g}-\mathbf{u}\| \geq (-\mathbf{g}-\mathbf{u},\mathbf{d}) = (-\mathbf{g},\mathbf{d}) - (\mathbf{u},\mathbf{d}) \geq -(\mathbf{g},\mathbf{d}) \geq -f^K(\mathbf{a};\mathbf{d}).$$

Ebből nyilvánvaló, hogy

$$m \geq -f^K(\mathbf{a};\mathbf{d}) > 0, \tag{3.3}$$

következésképpen igaz az (*i*) $\Longrightarrow$ (*iv*) implikáció.

Induljunk most ki az $m > 0$ feltételből. Legyenek $g_n \in \partial_-^K f(\mathbf{a})$ és $\mathbf{u}_n \in N_c(\mathbf{a})$, $n = 1,2,\ldots$ olyan végtelen sorozatok, hogy $\|-\mathbf{g}_n - \mathbf{u}_n\| \to m$. Mivel a $\{\mathbf{g}_n + \mathbf{u}_n\}$ vektorsorozat korlátos, ezért kiválasztható belőle konvergens részsorozat. Az egyszerűség kedvéért tegyük fel, — és ezzel az általánosságot sem sértjük — hogy $\mathbf{g}_n + \mathbf{u}_n \to \mathbf{w}$. Nyilvánvaló, hogy $\mathbf{w}$ a $\mathrm{cl}\{N_c(\mathbf{a}) + \partial_-^K f(\mathbf{a})\}$ zárt konvex halmaz minimál-normájú eleme és $\|\mathbf{w}\| = m$. Legyen

$$\mathbf{s} = -\frac{\mathbf{w}}{\|\mathbf{w}\|}. \tag{3.4}$$

Megmutatjuk, hogy $\mathbf{s} \in A_f^K(\mathbf{a}) \cap T_c(\mathbf{a})$. A $\mathbf{w}$ minimál-norma tulajdonságából következik, hogy minden $\mathbf{g} \in \partial_-^K f(\mathbf{a})$ esetén $(\mathbf{g}+\mathbf{u}-\mathbf{w},\mathbf{w}) \geq 0$, [8, Theorem 4.2.7 és 4.2.9], következésképpen $(\mathbf{g}+\mathbf{u},\mathbf{w}) \geq (\mathbf{w},\mathbf{w})$ teljesül. Ez utóbbi (3.4) folytán a következőképpen is írható:

$$(\mathbf{g}+\mathbf{u},\mathbf{s}) \geq -m,$$

illetve

$$(\mathbf{g},\mathbf{s}) \geq -m-(\mathbf{u},\mathbf{s}). \tag{3.5}$$

Mivel $N_c(\mathbf{a})$ kúp, ezért (3.5) ekvivalens a következővel:

$$(\mathbf{g},\mathbf{s}) \geq -m-c(\mathbf{u},\mathbf{s}) \tag{3.6}$$

minden $\mathbf{g} \in \partial_-^K f(\mathbf{a})$, $\mathbf{u} \in N_c(\mathbf{a})$ és $c \geq 0$ esetén. Ez viszont csak úgy lehet igaz, ha minden $\mathbf{u} \in N_c(\mathbf{a})$ és $\mathbf{g} \in \partial_-^K f(\mathbf{a})$ esetén

$$(\mathbf{u},\mathbf{s}) \leq 0 \quad \text{és} \quad (\mathbf{g},\mathbf{s}) \leq -m.$$

Ezekből nyilvánvaló, hogy $\mathbf{s} \in T_c(\mathbf{a})$, illetve (3.1a,b) miatt, hogy

$$f^K(\mathbf{a};\mathbf{s}) \leq -m < 0, \tag{3.7}$$

következésképpen $\mathbf{s} \in A^K{}_f(\mathbf{a})$, tehát igaz a $(iv) \implies (i)$ implikáció. Q.E.D.

A legmeredekebb csökkenés iránya. Az $\mathbf{s}$ vektornak ebben az esetben is megvan a (2.21) szélsőérték tulajdonsága.

3.4. TÉTEL. *Legyen $f^K(\mathbf{a};\mathbf{d})$ reguláris kvázikonvex approximációja $f(\mathbf{x})$-nek $\mathbf{a}$-ban. Legyen $\mathbf{s}$ a (3.4) képlettel definiált vektor. Ekkor*

$$\mathbf{s} \in A^K{}_f(\mathbf{a}) \cap T_c(\mathbf{a}) \cap S, \tag{3.8}$$

továbbá, ha $\mathbf{p} \in A^K{}_f(\mathbf{a}) \cap T_c(\mathbf{a}) \cap S$ és $\mathbf{p} \neq \mathbf{s}$, akkor

$$-m = f^K(\mathbf{a};\mathbf{s}) < f^K(\mathbf{a};\mathbf{p}). \tag{3.9}$$

Bizonyítás. Az előző tétel bizonyítása során (3.8) bizonyítást nyert. (3.3) és (3.7) együtt a $-m = f^K(\mathbf{a};\mathbf{s})$ egyenlőséget adja. Legyen $\mathbf{p} \in A^K{}_f(\mathbf{a}) \cap S$ tetszőleges, de $\mathbf{p} \neq \mathbf{s}$. Ekkor (3.1a) szerint minden $\mathbf{g} \in \partial_-^K f(\mathbf{a})$ vektorra $(\mathbf{g},\mathbf{p}) \leq f^K(\mathbf{a},\mathbf{p})$ teljesül. Ebből már egyszerűen adódik, hogy

$$(\mathbf{g}+\mathbf{u},\mathbf{p}) \leq f^K(\mathbf{a},\mathbf{p}) \tag{3.10}$$

minden $\mathbf{g} \in \partial_-^K f(\mathbf{a})$ és $\mathbf{u} \in N_c(\mathbf{a})$ esetén.

Tekintsük most a 3.3 tétel bizonyításában szereplő $\{\mathbf{g}_n\}$ és $\{\mathbf{u}_n\}$ vektorsorozatokat. Mivel $\mathbf{g}_n + \mathbf{u}_n \to \mathbf{w}$ és $\mathbf{w} = -m\,\mathbf{s}$, ezért a (3.10) egyenlőtlenséget a $\{\mathbf{g}_n + \mathbf{u}_n\}$ sorozatra alkalmazva és az $n \to \infty$ határátmenetet végrehajtva a

$$(\mathbf{w},\mathbf{p}) = -m(\mathbf{s},\mathbf{p}) \leq f^K(\mathbf{a},\mathbf{p}) \tag{3.11}$$

egyenlőtlenségre jutunk. Mivel a $\mathbf{p}$ és $\mathbf{s}$ vektorok egységnyi hosszúságúak és nem párhuzamosak, ezért a *Cauchy-Schwarz-egyenlőtlenség* szerint $-1 > (\mathbf{s},\mathbf{p})$. Ezt figyelembe véve a (3.11) egyenlőtlenségből a bizonyítandó (3.9) összefüggés adódik.

Q.E.D.

Az optimalitás szükséges feltétele: végezetül megfogalmazzuk a 2.9. tétel analogonját, mely közvetlen folyománya a 3.3. tételnek.

3.5. TÉTEL. *Legyen $f^K(\mathrm{a};\mathrm{d})$ reguláris kvázikonvex elsőrendű approximációja az $f(\mathrm{x})$ függvénynek az a pontban. Ekkor ahhoz, hogy az a pont lokális optimális megoldása legyen az (1.1) feladatnak, szükséges, hogy teljesüljön a*

$$0 \in \operatorname{cl}[N_c(\mathrm{a}) + \partial_-^K f(\mathrm{a})] \tag{3.12}$$

feltétel.

4. Az $N_c(\mathrm{a})$ kúp kapcsolata a feltételi függvények approximációival

Ebben a részben az (1.1) feladat következő speciális esetét vizsgáljuk

$$\begin{gathered}\min f(\mathrm{x}) \\ C = \{\mathrm{x} \in R^n : g_i(\mathrm{x}) \leq 0, \quad i = 1,2,\dots m\}.\end{gathered} \tag{4.1}$$

Vizsgálódásunk egy $\mathrm{a} \in C$ lehetséges megoldás egy környezetére korlátozódik. Továbbra is feltételezzük, hogy a C halmaz lokálisan konvex a-ban. Tegyük fel továbbá, hogy a $g_i(\mathrm{x})$ feltételi függvények valamennyien folytonosak az a pontban. Célunk az $N_c(\mathrm{a})$ kúp és a feltételi függvények kapcsolatának vizsgálata. „Aggregáljuk" a $g_i(\mathrm{x})$ feltételi függvényeket a következő módon; legyen

$$g(\mathrm{x}) = \max\{g_i(\mathrm{x}) : i = 1,2\dots m\}. \tag{4.2}$$

Nyilvánvaló, hogy

$$C = \{\mathrm{x} \in R^n : g(\mathrm{x}) \leq 0\}. \tag{4.2}$$

Első lépésként az $N_c(\mathrm{a})$ kúp és a $g(\mathrm{x})$ függvény kapcsolatát vizsgáljuk. Ha $g(\mathrm{a}) = \max g_i(\mathrm{a}) < 0$, akkor a $g_i(\mathrm{x})$ függvények folytonosságából adódóan $T_c(\mathrm{a}) = R^n$ és ennélfogva $N_c(\mathrm{a}) = \{0\}$. Vizsgáljuk a továbbiakban a $g(\mathrm{a}) = 0$ esetet.

4.1. TÉTEL. *Legyen $g^K(\mathrm{a};\mathrm{d})$ elsőrendű approximációja $g(\mathrm{x})$-nek az a pontban. Ekkor*

$$A^K{}_g(\mathrm{a}) \subset A_g(\mathrm{a}) \subset T_c(\mathrm{a}). \tag{4.4}$$

A tétel állítása a szóbanforgó kúpok definicióiból és a $g(\mathrm{a}) = 0$ feltételből közvetlenül következik.

4.2. Definició. Nevezzük a $g^K(\mathrm{a};\mathrm{d})$ elsőrendű approximációt ***limesztípusúnak***, ha rendelkezik a következő tulajdonsággal:
Ha van olyan $T > 0$, hogy $0 \leq t \leq T$ esetén $g(\mathrm{a}+t\,\mathrm{d}) - g(\mathrm{a}) \leq 0$, akkor $g^K(\mathrm{a};\mathrm{d}) \leq 0$.

4.3. TÉTEL. *Legyen $g^K(\mathrm{a};\mathrm{d})$ limesztípusú elsőrendű approximáció. Ekkor*

$$T_c(\mathrm{a}) \subset M^K{}_g(\mathrm{a}), \tag{4.5}$$

ahol $M^K{}_g(\mathrm{a}) = \{\mathrm{d} : g^K(\mathrm{a},\mathrm{d}) \leq 0\}$.

A tétel állítása a definiciók közvetlen következménye.

4.4. Definíció. A $g(\mathrm{x}) \leq 0$ feltételt *regulárisnak* mondjuk az a pontban a $g^K(\mathrm{a};\mathrm{d})$ elsőrendű approximációra nézve, ha

$$A^K{}_g(\mathrm{a}) \neq \emptyset \quad \text{és} \quad \mathrm{cl}\, A^K{}_g(\mathrm{a}) = M^K{}_g(\mathrm{a}). \tag{4.6}$$

Ha a $g^K(\mathrm{a};\mathrm{d})$ elsőrendű approximáció lineáris, akkor a $g(\mathrm{x}) \leq 0$ feltétel reguláris az a pontban $g^K(\mathrm{a};\mathrm{d})$-re nézve. Ha a $g^K(\mathrm{a};\mathrm{d})$ elsőrendű approximáció konvex és teljesül az $A^K{}_g(\mathrm{a}) \neq \emptyset$ *Slater-feltétel*, akkor a $g(\mathrm{x}) \leq 0$ feltétel reguláris az a pontban $g^K(\mathrm{a};\mathrm{d})$-re nézve.

4.5. TÉTEL. *Legyen a $g(\mathrm{x}) \leq 0$ feltétel reguláris az a pontban a $g^K(\mathrm{a};\mathrm{d})$ limesztípusú approximációra nézve, ahol $g(\mathrm{a}) = 0$. Ekkor*

$$\mathrm{cl}\, A^K{}_g(\mathrm{a}) = T_c(\mathrm{a}) = \{\mathrm{d} : g^K(\mathrm{a};\mathrm{d}) \leq 0\}. \tag{4.7}$$

Bizonyítás. (4.4), (4.5) és (4.6) miatt

$$\mathrm{cl}\, A^K{}_g(\mathrm{a}) = \mathrm{cl}\, A_g(\mathrm{a}) = T_c(\mathrm{a}) = M^K{}_g(\mathrm{a}),$$

ebből pedig (4.7) közvetlenül adódik.

4.6. TÉTEL. *Legyen $g^K(\mathrm{a};\mathrm{d})$ konvex elsőrendű approximáció $\partial^K g(\mathrm{a})$ szubdifferenciállal. Ekkor*

$$[\partial^K g(\mathrm{a})]^0 = M^K{}_g(\mathrm{a}). \tag{4.8}$$

Bizonyítás. Mivel

$$\mathrm{d} \in [\partial^K g(\mathrm{a})]^0 \iff (\mathrm{d},\mathrm{z}) \leq 0 \quad \forall\, \mathrm{z} \in \partial^K g(\mathrm{a})$$

és

$$g^K(\mathrm{a};\mathrm{d}) = \max\{(\mathrm{d},\mathrm{z}) : \mathrm{z} \in \partial^K g(\mathrm{a})\},$$

ezért (4.8) szükségképpen igaz.

4.7. TÉTEL. *Legyen a $g(\mathrm{x}) \leq 0$ feltétel reguláris az a pontban a $g^K(\mathrm{a};\mathrm{d})$ limesztípusú konvex elsőrendű approximációra nézve és legyen $g(\mathrm{a}) = 0$. Ekkor*

$$T_c(\mathrm{a}) = [\partial^K g(\mathrm{a})]^0. \tag{4.9}$$

és

$$N_c(\mathrm{a}) = \mathit{cone}\, \partial^K g(\mathrm{a}). \tag{4.10}$$

Bizonyítás. (4.9) közvetlenül következik (4.7)-ből, ill. (4.8)-ból. A (4.9) összefüggés az

$$N_c(\mathbf{a}) = T_c(\mathbf{a})^0 = [\partial^K g(\mathbf{a})]^{00} = \mathrm{cl\,cone}\,\partial^K g(\mathbf{a}) \tag{4.11}$$

összefüggésből adódik. A regularitás folytán $0 \notin \partial^K g(\mathbf{a})$ és ez biztosítja a cone $\partial^K g(\mathbf{a})$ halmaz zártságát, ennélfogva (4.10) és (4.11) jobb oldalai megegyeznek.

4.8. TÉTEL. *Legyen* $g^K(\mathbf{a};\mathbf{d})$ *reguláris kvázikonvex approximáció* $\partial_-^K g(\mathbf{a})$ *kváziszubdifferenciállal. Ekkor*

$$[\partial_-^K g(\mathbf{a})]^0 = \mathrm{cl}\, A^K{}_g(\mathbf{a}). \tag{4.12}$$

Bizonyítás. Mivel

$$\mathbf{d} \in [\partial_-^K g(\mathbf{a})]^0 \iff (\mathbf{d};\mathbf{z}) \le 0 \quad \forall \mathbf{z} \in \partial_-^K g(\mathbf{a})$$

és (3.1a,b) szerint

$$\sup\{(\mathbf{d},\mathbf{z}) : \mathbf{z} \in \partial_-^K g(\mathbf{a})\} \le 0, \text{ ha } \mathbf{d} \in \mathrm{cl}\, A^K{}_g(\mathbf{a}),$$

illetve

$$\sup\{(\mathbf{d},\mathbf{z}) : \mathbf{z} \in \partial_-^K g(\mathbf{a})\} = +\infty, \text{ ha } \mathbf{d} \notin \mathrm{cl}\, A^K{}_g(\mathbf{a}),$$

ezért (4.12) szükségképpen igaz.

4.9. TÉTEL. *Legyen a* $g(\mathbf{x}) \le 0$ *feltétel reguláris az a pontban a* $g^K(\mathbf{a};\mathbf{d})$ *limesztípusú reguláris kvázikonvex approximációra nézve és legyen* $g(\mathbf{a}) = 0$. *Ekkor*

$$T_c(\mathbf{a}) = [\partial_-^K g(\mathbf{a})]^0 \tag{4.13}$$

és

$$N_c(\mathbf{a}) = \mathrm{cl}\ \mathit{cone}\ \partial_-^K g(\mathbf{a}). \tag{4.14}$$

Bizonyítás. (4.13) közvetlenül következik (4.7)-ből és (4.12)-ből. (4.14) nem más, mint (4.13) duálisa.

Az optimalitás szükséges feltételei: a 2., ill. 3. fejezetekben tárgyalt optimalitási feltételek átfogalmazhatók jelen fejezet eredményeinek segítségével a

$$\begin{array}{c} \min f(\mathbf{x}) \\ g(\mathbf{x}) \le 0 \end{array} \tag{4.15}$$

feladat esetére.

4.10. TÉTEL. *Legyen* $f^K(\mathrm{a};\mathrm{d})$ *konvex elsőrendű approximációja az* $f(\mathrm{x})$ *függvénynek az* a *pontban, ahol* $g(\mathrm{a}) = 0$. *Legyen a* $g(\mathrm{x}) \leq 0$ *feltétel reguláris az* a *pontban a* $g^L(\mathrm{a};\mathrm{d})$ *limesztípusú konvex elsőrendű approximációra nézve. Ekkor ahhoz, hogy az* a *pont lokális optimális megoldása legyen a (4.15) feladatnak, szükséges, hogy teljesüljön a*

$$0 \in \partial^K f(\mathrm{a}) + \mathrm{cone}\ \partial^L g(\mathrm{a}) \tag{4.16}$$

feltétel.

Bizonyítás. A tétel állítása a 2.9. tétel és a 4.7. tétel közvetlen következménye.

4.11. TÉTEL. *Legyen* $f^K(\mathrm{a};\mathrm{d})$ *reguláris kvázikonvex elsőrendű approximációja az* $f(\mathrm{x})$ *függvénynek az* a *pontban, ahol* $g(\mathrm{a}) = 0$. *Legyen a* $g(\mathrm{x}) \leq 0$ *feltétel reguláris az* a *pontban a* $g^L(\mathrm{a};\mathrm{d})$ *limesztípusú konvex approximációra nézve. Ekkor ahhoz, hogy az* a *pont lokális optimális megoldása legyen a (4.15) feladatnak, szükséges, hogy teljesüljön a*

$$0 \in \mathrm{cl}[\partial^K_- f(\mathrm{a}) + \mathrm{cone}\ \partial^L_- g(\mathrm{a})] \tag{4.17}$$

feltétel.

Bizonyítás. A tétel állítása a 3.5. és a 4.9. tételek következménye.

4.12. TÉTEL. *Legyen* $f^K(\mathrm{a};\mathrm{d})$ *konvex elsőrendű approximációja az* $f(\mathrm{x})$ *függvénynek az* a *pontban, ahol* $g(\mathrm{a}) = 0$. *Legyen a* $g(\mathrm{x}) \leq 0$ *feltétel reguláris az* a *pontban a* $g^L(\mathrm{a};\mathrm{d})$ *limesztípusú reguláris kvázikonvex approximációra nézve. Ekkor ahhoz, hogy az* a *pont lokális optimális megoldása legyen a (4.15) feladatnak, szükséges, hogy teljesüljön a*

$$0 \in \partial^K f(\mathrm{a}) + \mathrm{cl\,cone}\ \partial^L_- g(\mathrm{a}) \tag{4.18}$$

feltétel.

Bizonyítás. A tétel állítása közvetlenül adódik a 2.9. és a 4.9. tételekből.

4.13. TÉTEL. *Legyen* $f^K(\mathrm{a};\mathrm{d})$ *reguláris kvázikonvex elsőrendű approximációja az* $f(\mathrm{x})$ *függvénynek az* a *pontban, ahol* $g(\mathrm{a}) = 0$. *Legyen a* $g(\mathrm{x}) \leq 0$ *feltétel reguláris az* a *pontban a* $g^L(\mathrm{a};\mathrm{d})$ *limesztípusú reguláris kvázikonvex approximációra nézve. Ekkor ahhoz, hogy az* a *pont lokális optimális megoldása legyen a (4.15) feladatnak, szükséges, hogy teljesüljön a*

$$0 \in \mathrm{cl}[\partial^K_- f(\mathrm{a}) + \mathrm{cone}\ \partial^L_- g(\mathrm{a})] \tag{4.19}$$

feltétel.

Bizonyítás. A tétel állítása a 3.5. és a 4.9. tételekből adódik.

5. Kuhn–Tucker típusú optimalitási feltételek

A továbbiakban visszatérünk a

$$\begin{aligned} &\min f(\mathbf{x}) \\ (5.1) \qquad &g_i(\mathbf{x}) \leq 0, \quad i = 1, 2, \dots m \\ &\mathbf{x} \in R^n \end{aligned}$$

feladat vizsgálatához. Legyen a lehetséges megoldása (5.1)-nek és legyenek a feltételi függvények folytonosak a-ban. Jelölje $J \subset \{1, 2, \dots, m\}$ az a-ban aktív feltételek ($g_j(\mathbf{a}) = 0$) indexhalmazát. Tegyük fel, hogy J nem üres. Legyenek $g_j^{L_j}(\mathbf{a}; \mathbf{d}), j \in J$ elsőrendű approximációi a $g_j(\mathbf{x})$, $j \in J$ függvényeknek az a pontban. Legyenek

$$g(\mathbf{x}) = \max\{g_j(\mathbf{x}) : j \in J\}$$

és

$$(5.2) \qquad g^L(\mathbf{a}; \mathbf{d}) = \max\ \{g_j^{L_i}(\mathbf{a}; \mathbf{d}) : j \in J\}.$$

5.1. TÉTEL. *A $g^L(\mathbf{a}; \mathbf{d})$ függvény elsőrendű approximációja $g(\mathbf{x})$-nek* **a***-ban.*

Bizonyítás. Egyszerűen belátható, hogy $g^L(\mathbf{a}; \mathbf{d})$ pozitív homogén függvénye a d iránynak. Legyen $\mathbf{d} \in A^L{}_g(\mathbf{a})$, azaz $g^L(\mathbf{a}; \mathbf{d}) < 0$. Ekkor $g_j^{L_j}(\mathbf{a}; \mathbf{d}) < 0$ minden $j \in J$ esetén. $i \in J$ esetén ebből, $i \notin J$ esetén pedig a $g_i(\mathbf{x})$ függvények a-beli folytonosságából következik, hogy létezik olyan $T > 0$ valós szám, hogy $0 \leq t \leq T$ esetén

$$g_i(\mathbf{a} + t\,\mathbf{d}) < 0 \quad \text{minden } i = 1, 2, \dots, m \text{ indexre.}$$

Ebből nyilvánvaló, hogy minden $0 \leq t \leq T$ esetén

$$g(\mathbf{a} + t\,\mathbf{d}) < 0 = g(\mathbf{a}).$$

Ez azt jelenti, hogy $\mathbf{d} \in A_g(\mathbf{a})$, következésképpen $A^L{}_g(\mathbf{a}) \subset A_g(\mathbf{a})$. Q.E.D.

5.2. TÉTEL. *Ha a $g_j^{L_j}(\mathbf{a}; \mathbf{d})$, $j \in J$ elsőrendű approximációk limesztípusúak, akkor $g^L(\mathbf{a}; \mathbf{d})$ is limesztípusú.*

Bizonyítás. Tegyük fel, hogy $g(\mathbf{a} + t\,\mathbf{d}) \leq g(\mathbf{a})$ teljesül valamely $0 \leq t \leq T$ intervallumon. Ekkor ugyanitt minden $j \in J$-re

$$g_j(\mathbf{a} + t\,\mathbf{d}) \leq g(\mathbf{a} + t\,\mathbf{d}) \leq g(\mathbf{a}) = 0 = g_j(\mathbf{a})$$

teljesül. Feltételezésünk szerint ekkor

$$g_j^{L_j}(\mathbf{a}; \mathbf{d}) \leq 0 \quad \text{minden} \quad j \in J - \text{re},$$

következésképpen $g^L(\mathbf{a}; \mathbf{d}) \leq 0$. Q.E.D.

5.3. TÉTEL. *Ha a $g_j^{L_j}(\mathbf{a};\mathbf{d})$ elsőrendű approximációk konvexek, akkor $g^L(\mathbf{a};\mathbf{d})$ is konvex és*

$$\partial^L g(\mathbf{a}) = \operatorname{conv}\{\cup\partial^{L_j} g_j(\mathbf{a}) : j \in J\}. \tag{5.3}$$

A tétel a konvex analízis egyik nevezetes tételét fogalmazza meg. [9, Theorem 3.6.].

5.4. TÉTEL. *Ha a $g_j^{L_j}(\mathbf{a};\mathbf{d})$, $j \in J$ elsőrendű approximációk kvázikonvexek és regulárisak, akkor $g^L(\mathbf{a};\mathbf{d})$ is kvázikonvex reguláris, továbbá*

$$\partial_-^L g(\mathbf{a}) = \operatorname{cl}\operatorname{conv}\{\cup\partial_-^{L_j} g_j(\mathbf{a}) : j \in J\}. \tag{5.4}$$

Bizonyítás. Mivel bármely c valós szám esetén

$$\{\mathbf{d} : g^L(\mathbf{a};\mathbf{d}) \le c\} = \bigcap_{j\in J}\{\mathbf{d} : g_j^{L_j}(\mathbf{a};\mathbf{d}) \le c\},$$

ezért a feltételekből következik, hogy bármely valós c esetén a $\{\mathbf{d} : g^L(\mathbf{a};\mathbf{d}) \le c\}$ alsó nívóhalmaz zárt és konvex, következésképpen a $g^L(\mathbf{a};\mathbf{d})$ approximáció kvázikonvex és alulról félig folytonos. Mivel

$$\{\mathbf{d} : g^L(\mathbf{a};\mathbf{d}) < 0\} = \bigcap_{j\in J}\{\mathbf{d} : g_j^{L_j}(\mathbf{a};\mathbf{d}) < 0\},$$

ezért a feltételekből következik, hogy $A^L{}_g(\mathbf{a})$ nyílt halmaz. A $g^L(\mathbf{a};\mathbf{d})$ kvázikonvex approximáció tehát reguláris is.

Most megmutatjuk az (5.4) összefüggés helyességét, melynek jobb oldalán szereplő zárt konvex halmazt jelöljük egyszerűen M-el. Először megmutatjuk, hogy minden $j \in J$-re

$$\partial_-^{L_j} g_j(\mathbf{a}) \subset \partial_-^L g(\mathbf{a}). \tag{5.5}$$

A *Crouzeix-féle repezentációs tétel* szerint

$$\sup\{(\mathbf{d},\mathbf{z}) : \mathbf{z} \in \partial_-^{L_j} g_j(\mathbf{a})\} = \begin{cases} g_j{}^{L_j}(\mathbf{a};\mathbf{d}) \le 0, & \text{ha } \mathbf{d} \in \operatorname{cl} A^{L_j} g_j(\mathbf{a}) \\ +\infty, & \text{egyébként,} \end{cases}$$

és

$$\sup\{(\mathbf{d},\mathbf{z}) : \mathbf{z} \in \partial_-^L g(\mathbf{a})\} = \begin{cases} g^L(\mathbf{a};\mathbf{d}) \le 0, & \text{ha } \mathbf{d} \in \operatorname{cl} A^L{}_g(\mathbf{a}) \\ +\infty, & \text{egyébként,} \end{cases}$$

Mivel $g_j{}^{L_j}(\mathbf{a};\mathbf{d}) \le g^L(\mathbf{a};\mathbf{d})$ minden d-re, ezért ezt a fenti reprezentációkkal összevetve kapjuk, hogy minden $\mathbf{d} \in R^n$-re

$$\sup\{(\mathbf{d},\mathbf{z}) : \mathbf{z} \in \partial_-^{L_j} g_j(\mathbf{a})\} \le \sup\{(\mathbf{d},\mathbf{z}) : \mathbf{z} \in \partial_-^L g(\mathbf{a})\}. \tag{5.6}$$

A konvex analízis szeparációs tételeiből következik, hogy az (5.5) tartalmazás teljesüléséhez szükséges és elégséges, hogy a szóbanforgó zárt konvex halmazok tartó függvényei között az (5.6) kapcsolat álljon fenn [10, Theorem 2.7]. Mivel az (5.6) összefüggés minden $j \in J$-re teljesül és $\partial_-^L g(\mathbf{a})$ zárt és konvex, ezért

$$M \subset \partial_-^L g(\mathbf{a}). \tag{5.7}$$

Az $M = \partial_-^L g(\mathbf{a})$ feltételt indirekt módon bizonyítjuk. Tegyük fel, hogy $\mathbf{z}_0 \in \partial_-^L g(\mathbf{a})$, de $\mathbf{z}_0 \notin M$. Mivel M zárt és konvex, ezért $\mathbf{z}_0$ szigorúan szeparálható M-től, azaz van olyan $\mathbf{d}_0 \in R^n$, hogy

$$\sup\{(\mathbf{d}_0, \mathbf{z}) : \mathbf{z} \in M\} < (\mathbf{d}_0, \mathbf{z}_0).$$

Mivel minden $j \in J$-re $\partial_-^{L_j} g_j(\mathbf{a}) \subset M$, ezért

$$\sup\{(\mathbf{d}_0, \mathbf{z}) : \mathbf{z} \in \partial_-^{L_j} g_j(\mathbf{a})\} \le \sup\{(\mathbf{d}_0, \mathbf{z}) : \mathbf{z} \in M\} < (\mathbf{d}_0, \mathbf{z}_0).$$

A *Crouzeix-féle reprezentációs képletek* szerint ekkor

$${g_j}^{L_j}(\mathbf{a}; \mathbf{d}) < (\mathbf{d}_0, \mathbf{z}_0),$$

amiből

$$g^L(\mathbf{a}; \mathbf{d}_0) < (\mathbf{d}_0, \mathbf{z}_0) \tag{5.9}$$

következik. Másrészt viszont, ugyancsak a *Crouzeix-féle képletek* miatt

$$g^L(\mathbf{a}; \mathbf{d}) = \sup\{(\mathbf{d}_0, \mathbf{z}) : \mathbf{z} \in \partial_-^L g(\mathbf{a})\} \ge (\mathbf{d}_0, \mathbf{z}_0).$$

Ez azonban ellentmond az (5.9) relációnak. Ez az ellentmondás igazolja az (5.4) egyenlőséget. Q.E.D.

Tegyük fel, hogy a $g(\mathbf{x}) = \max\{g_j(\mathbf{x}) : j \in J\} \le 0$ feltétel reguláris az $\mathbf{a}$ pontban a $g^L(\mathbf{a}; \mathbf{d}) = \max\{g^{L_j}(\mathbf{a}; \mathbf{d}) : j \in J\}$ elsőrendű approximációra nézve. Dolgozatunk hátralevő részében az (5.1) feladatra kirótt ezen regularítási feltevést külön említés nélkül is érvényesnek tekintjük.

5.5. TÉTEL. *Legyen $f^K(\mathbf{a}; \mathbf{d})$ konvex $\partial^K f(\mathbf{a})$ szubdifferenciállal és legyen ${g_j}^{L_j}(\mathbf{a}; \mathbf{d})$ minden $j \in J$-re limesztípusú és konvex $\partial^{L_j} g_j(\mathbf{a})$ szubdifferenciállal. Ekkor ahhoz, hogy az $\mathbf{a}$ pont lokális optimális megoldása legyen az (5.1) feladatnak, szükséges, hogy teljesüljön a*

$$0 \in \partial^K f(\mathbf{a}) + \text{cone conv } \{\cup\, \partial^{L}{}_j g_j(\mathbf{a}) : j \in J\}. \tag{5.10}$$

feltétel.

Ha $f^K(\mathbf{a}; \mathbf{d})$ reguláris kvázikonvex $\partial_-^K f(\mathbf{a})$ kvázisubdifferenciállal, akkor

$$0 \in \text{cl}[\partial_-^K f(\mathbf{a}) + \text{cone conv}\{\cup\, \partial^{L_j} g_j(\mathbf{a}) : j \in J\}] \tag{5.11}$$

feltétel szükséges feltétele a lokális optimalitásnak.

Bizonyítás. A tétel állítása közvetlenül következik a 4.10 és 5.3, ill. a 4.11 és 5.3 tételekből.

5.6. TÉTEL. *Legyen $f^K(\mathrm{a};\mathrm{d})$ konvex $\partial^K f(\mathrm{a})$ szubdifferenciállal és legyen ${g_j}^{L_j}(\mathrm{a};\mathrm{d})$ limesztípusú reguláris és kvázikonvex minden $j \in J$-re $\partial_-^{L_j} g_j(\mathrm{a})$ kvázi-szubdifferenciállal. Ekkor ahhoz, hogy az a pont lokális optimális megoldása legyen az (5.1) feladatnak, szükséges, hogy teljesüljön a*

$$0 \in \partial^K f(\mathrm{a}) + \mathrm{cl}\,\mathrm{cone}\;\mathrm{conv}\;\{\cup\, \partial_-^{L_j} g_j(\mathrm{a}) : j \in J\} \tag{5.12}$$

feltétel.

Ha $f^K(\mathrm{a};\mathrm{d})$ reguláris kvázikonvex $\partial_-^K f(\mathrm{a})$) kváziszubdifferenciállal, akkor a

$$0 \in \mathrm{cl}[\partial_-^K f(\mathrm{a}) + \mathrm{cone}\;\mathrm{conv}\;\{\cup\, \partial_-^{L_j} g_j(\mathrm{a}) : j \in J\}] \tag{5.13}$$

feltétel szükséges feltétele a lokális optimalitásnak.

Bizonyítás. A tétel állításai a 4.12 és 5.4, illetve a 4.13 és 5.4 tételek közvetlen következménye.

Az (5.10) feltétel a konvex programozásból jól ismert *Kuhn–Tucker feltétel* megfelelője. Ennélfogva az (5.11), (5.12) és (5.13) feltételeket joggal tekinthetjük úgy, mint *általánosított Kuhn–Tucker feltételeket.*

IRODALOM

[1] CLARKE, F.H., *Optimization and nonsmooth analysis* (John Wiley, New York, 1983).

[2] CROUZEIX, J.-P., "Continuity and differentiability properties of quasiconvex functions on R^n", in Generalized concavity in optimization and economics, Eds.: Schaible S., Ziemba W.T. (Academic Press, New York, 1981), pp. 109–130.

[3] CROUZEIX, J.-P., "About differentiability of order one of quasiconvex functions on R^n", *JOTA* **36** (1982), 367–385.

[4] ELSTER, K.-H. and THIERFELDER, J., "Abstract cone approximations and generalized differentiability in nonsmooth optimization", *Optimization* 19 (1988), 315–341.

[5] GIANNESSI, F., "Semidifferentiable functions and necessary optimality conditions", *JOTA* **60** (1989), 191–241.

[6] IOFFE, A.D., "Necessary and sufficient conditions for a local minimum: 1", *SIAM J. Control and Optimization* **17** (1979), 245–250.

[7] KOMLOSI, S., "On a possible generalization of Pshenichnyi's quasidifferentiability", *Optimization* **21** (1990), 3–11.

[8] MARTOS, B., *Nonlinear programming: theory and methods* (Akadémiai Kiadó, Budapest, 1975).

[9] PSHENICHNYI, B.T., *Necessary conditions for an extremum* (Nauka, Moszkva, 1969).

[10] PSHENICHNYI, B.T., *Convex analysis and extremum problems* (Nauka, Moszkva, 1980).

[11] ROCKAFELLAR, R.T., *Convex analysis* (Princeton University Press, Princeton, 1970).

(Beérkezett: 1990. február 8.)

KOMLÓSI SÁNDOR
JPTE KÖZGAZDASÁGTUDOMÁNYI KAR
GAZDASÁGMATEMATIKAI TANSZÉK
7601 PÉCS, RAKÓCZI U. 80.

QUASICONVEX FIRST ORDER APPROXIMATIONS

S. KOMLÓSI

The role of the concept of first order approximation introduced by A.D. IOFFE and weakened by the authoris analyzed in this paper. After recalling some known results concerning convex first order approximations the geometric background for quasiconvex approximations is investigated in order to obtain *Kuhn-Tucker type* necessary optimality conditions in terms of quasisub-diffferentials.

Alkalmazott Matematikai Lapok **15** (1990–91) 329–352

ADAPTÍV PARTÍCIÓS MÓDSZEREK GLOBÁLIS OPTIMALIZÁLÁSI FELADATOK MEGOLDÁSÁRA

PINTÉR JÁNOS

Budapest

A többextrémumú (globális) optimalizáció feladatkörének és legfontosabb módszertípusainak rövid áttekintését követően, *adaptív halmazfelbontáson és részhalmazonkénti célfüggvény-approximáción* alapuló megoldási stratégiák egy általános osztályát definiáljuk. Az elméleti eredmények áttekintésének keretében egy „prototípus algoritmust" adunk meg. Ez a megközelítés alkalmazható pl. számos, formailag különböző *egyváltozós* globális optimalizálási módszer egyszerű, egységes leírására; emellett lehetővé teszi az említett algoritmusok közvetlen *többváltozós* kiterjesztését és – megfelelő implementációkban – alkalmas pl. *folytonos* vagy *Lipschitz-folytonos* függvények globális extrémumának *intervallum (tégla-) tartományon, konvex halmazon, csillaghalmazon,* illetve *Lipschitz-folytonos függvények által definiált tartományon* történő meghatározására. A globális optimalizáció feladatának aktualitását három fontos problématípus bemutatásával illusztráljuk: nemlineáris egyenletrendszerek általános megoldása, nemlineáris (leíró) modellek kalibrációja és (valós vektor) adatok osztályozása. E feladattípusok megoldását numerikus példákkal is szemléltetjük.

1. Bevezetés

A *globális optimalizálási probléma* (GOP)

$$
(1.1)\qquad \begin{aligned} &\min f(\mathbf{x}) = f(\mathbf{x}^*) = z^* \qquad \mathbf{x} \in D \\ &D = \{\mathbf{x} \in R^n : g_i(\mathbf{x}) \le 0, \qquad i = 1, \ldots, m\} \end{aligned}
$$

— tehát egy megadott $f : D \to R^1$ függvény *abszolut* (globális) minimumának meghatározása a $D \subset R^n$ megengedett tartományon — az utóbbi években intenzív kutatások tárgyává vált. Ezt az érdeklődést a feladat elméleti érdekessége és a potenciális alkalmazások széles köre egyaránt motiválja. Közismert az a tény, hogy — bár a GOP optimális megoldásának létezése enyhe analitikus feltételek (pl. f folytonos, D kompakt volta) mellett biztosított — a „klasszikus" nemlineáris (konvex) programozás módszerei tipikusan csak (1.1) valamely *lokális* megoldásának

Köszönetnyilvánítás. A globális optimalizáció témakörével kapcsolatos munkámat elsősorban a ROMAN G. STRONGINnal és REINER HORSTtal folytatott eszmecserék támogatták, de más — e területen dolgozó — kutatók is segítették az idevonatkozó aktuális irodalomban való tájékozódásomat. Ezúton is megköszönöm ROGER COOKE, C. FODOR JÁNOS, PESTI GÉZA, SZABÓ JÁNOS és SOMLYÓDY LÁSZLÓ társszerzőként való közreműködését a dolgozat irodalomjegyzékében szereplő közös munkáink elkészítésében. A számítógépi programok és a numerikus eredmények egy része a *Delfti Műszaki Egyetem* vendégeként eltöltött idő (1987) alatt készült el; a gépi futtatások jelentős részét a *Vízgazdálkodási Tudományos Kutató Központ* (VITUKI) eszközeinek felhasználásával végeztem el.

meghatározására alkalmasak. Továbbmenve, a GOP *nem-polinomiális* (NP) *bonyolultságú* feladat és még annak bizonyítása is NP-nehézségű, hogy egy adott megoldás nem (csupán) lokális optimumhely (l. pl. MURTY és KABADI (1987) dolgozatát). Természetesen a globális megoldás és a lokális optimumok értékei között számottevő eltérések mutatkozhatnak: ilyen esetekben indokolt tehát a „lehető legjobb" megoldás keresése.

Alkalmazási szempontból igen jelentős feladattípusok vezethetnek GOP-hoz. Példaként a következő feladatosztályokat említjük meg: legjobb nemlineáris függvényapproximáció meghatározása, nemlineáris egyenletrendszerek, egyensúly-problémák megoldása, statisztikai leírás, ill. fizikai, kémiai, biológiai stb. modellek, műszerek és folyamatok kalibrációja, mechanikai, elektromos stb. berendezések tervezése, adatok osztályozása, kiszolgáló rendszerek tervezése stb. (illusztrációképpen lásd pl. GORDON (1981), RAO, TYAGI és MISHRA (1981), GROCH, VIDIGAL és DIRECTOR (1985), WOMERSLEY (1986), CSENDES (1988), PAPALAMBROS és WILDE (1988), COOKE és PINTÉR (1989) munkáit; később — néhány konkrét feladattípus kapcsán — további példákat is idézünk majd).

A GOP megoldására irányuló legfontosabb algoritmus-típusoknak csupán vázlatos áttekintésére vállalkozhatunk a jelen keretek között; az Olvasó bővebb tájékoztatást nyerhet pl. DIXON és SZEGŐ, szerk. (1975,1978), STRONGIN (1978), FEDOROV, szerk. (1985), ZILINSKAS (1986), PARDALOS és ROSEN (1987), CHEW és ZHENG (1988), HORST (1988b), RATSCHEK és ROKNE (1988), RINNOOY KAN és TIMMER (1988) munkái alapján.

Számos (*heurisztikus* elemeket is tartalmazó) *sztochasztikus* megoldási módszer — pl. az ismert adaptív véletlen keresési eljárások és különböző, numerikus szempontból előnyös kiegészítéseik (TÖRN (1978), DEVROYE (1978), SOLIS és WETS (1981), PINTÉR (1984), RINNOOY KAN és TIMMER (1987 a,b)) — biztosítja az 1 valószínűségű globális konvergenciát. Ezeket az eljárásokat célszerű és indokolt pl. az extremális eloszlások, ill. a *Bayes-típusú döntések* elméletén alapuló ***statisztikus megállási szabályokkal kiegészíteni*** (l. pl. a GUMBEL (1958), GALAMBOS (1978), BERGER (1980) munkák általános tárgyalását, ill. DE HAAN (1981), BOENDER (1984) vagy FODOR és PINTÉR (1988) vizsgálatait.

A *determinisztikus* algoritmusokra áttérve, néhány módszer a GOP ekvivalens transzformációja útján elvben biztosítja a globális optimumhely meghatározását — a transzformált feladat megoldására viszont számos esetben nem áll rendelkezésre univerzálisan alkalmazható, egzakt eljárás. Az említett transzformációk pl. nemlineáris egyenlet, egyenlet-, ill. egyenlőtlenségrendszer megoldására (GARCIA és ZANGWILL (1981), LEVY és MONTALVO (1985), GE RENPU (1983), PAPALAMBROS és WILDE (1988)), vagy pedig egy nemlineáris függvény integrálására (CHEW és ZHENG, 1988) vezetik vissza az (1.1) feladat megoldását. E módszerosztályhoz sorolható a HANSEN, JAUMARD és LU (1989) dolgozatában javasolt, a ***korlátozás és szétválasztás elvén*** alapuló általános stratégia is.

Az *egyváltozós* globális optimalizálási feladat megoldására jónéhány egzakt, implementálható algoritmust javasoltak (pl. KUSHNER (1964), ARCHETTI és BETRO (1979), DANILIN és PIYAVSKII (1967), SHUBERT (1972), MOCKUS, TIESIS és

Zilinskas (1978), Strongin (1978), Hansen (1979), Zilinskas (1981), Basso (1982), Schoen (1982), Boender (1984) : e módszerek korrekt, numerikusan jól kezelhető *többváltozós* kiterjesztései azonban egészen a legutóbbi évekig nem voltak ismeretesek.

Globális optimalizálási algoritmusok konvergenciájának általános vizsgálatával foglalkozik pl. Horst (1986), Pintér (1986a), Horst és Tuy (1987), Pintér (1988a): az e munkákban követett tárgyalásmód lehetővé teszi az egyváltozós algoritmusok többdimenziós esetre történő általánosítását, ill. azoknak a konvergenciát megtartó, numerikusan előnyös módosításait is (Pintér, 1986b, Horst, 1988a). Itt említjük meg, hogy a jelen dolgozatban nem tárgyaljuk a GOP számos, gyakorlati szempontból fontos speciális esetét (tehát az *általános* GOP vizsgálatára szorítkozunk): speciálisabb feladattípusokat — konkáv, fordított (*reverse*) konvex, differenciális konvex (d.c.), indefinit kvadratikus programozás — tárgyal pl. Forgó (1978) magyar nyelvű munkája vagy Horst (1984), ill. Pardalos és Rosen (1987).

A továbbiakban *determinisztikus, gradiensmentes (direkt) algoritmusok* egy általános osztályát vizsgáljuk. E módszerek a megengedett megoldások D halmazának *szekvenciális (adaptív) felbontásán és a célfüggvény ennek megfelelő approximációján* alapszanak. A javasolt axiomatikus leírásmód lehetővé teszi jónéhány egyváltozós algoritmus egységes tárgyalását, az említett algoritmusok közvetlen, többdimenziós általánosítását és azok különböző implementációit. Az elméleti eredmények tömör áttekintését néhány perspektivikus alkalmazási terület példák segítségével történő bemutatása követi.

2. Az elméleti eredmények összefoglalása

2.1. Alapvető fogalmak és jelölések.

Az (1.1) feladat szerkezetére vonatkozóan két (elsősorban numerikus szempontból lényeges) feltétellel fogunk élni:

2.1. Definíció. A megengedett megoldások D halmaza *testszerű (robust)*, vagyis egy (nem-üres, korlátos) nyílt halmaz lezárása R^n-ben.

2.2. Definíció. Az f célfüggvény *Lipschitz-folytonos*, tehát tetszőleges $\mathbf{x}, \mathbf{y} \in D$ pontpár esetén fennáll az

$$|f(\mathbf{x}) - f(\mathbf{y})| \leq L||\mathbf{x} - \mathbf{y}|| \tag{2.1}$$

egyenlőtlenség ($L = L(f, D) > 0$ az f függvénynek megfelelő *Lipschitz-konstans* a D halmazon, $||\ ||$ pedig az euklideszi normát jelöli).

A továbbiakban azt is feltesszük, hogy az (1.1) feladatnak csak *véges számú* globálisan optimális megoldása van, ezek halmazát X^* jelöli.

2.1. Megjegyzés. A (2.1) feltétel — egyebek között — teljesül, ha pl. f folytonosan differenciálható, vagy ha f előállítható egy konvex és egy konkáv függvény különbségeként. Ilymódon a (2.1) reláció fennállása természetesnek tekinthető sok gyakorlati feladat esetén, annak ellenére, hogy a (lehető legkisebb) *Lipschitz-konstans* értéke általában nem ismert. (Itt jegyezzük meg azt is, hogy bizonyos esetekben — amelyekre külön rámutatunk majd — az f függvény *folytonossága* is elegendő lesz algoritmusok megfelelő konvergencia-tulajdonságának fennállásához.)

Az (1.1) feladat megoldására egy általános *adaptív partíciós algoritmus sémát* (PAS) vezetünk be (PINTÉR, 1987); ennek érdekében először két további fogalmat definiálunk.

2.3. Definíció. Legyen $M \subset R^n$ testszerű és jelöljön I egy véges indexhalmazt. Az $\{M_i : i \in I\}$ halmazok összessége M egy *testszerű felbontását* adja, ha M_i, $i \in I$ *testszerű* halmazok, emellett $M = \bigcup_i M_i$ és $M_i \cap M_j = \delta(M_i) \cap \delta(M_j)$, $i, j \in I$, $i \neq j$; itt $\delta(M_i)$ az M_i halmaz (M-re vonatkoztatott) relatív határát jelöli.

2.4. Definíció. Legyen adott M egy testszerű felbontása. A $P\{M_i, f(M_i)\}$ *partíció operátor* egy olyan funkcionál, amelynek értékét az M_i halmazból vett *mintapontok* (*sample points*) és az ezekben felvett *függvényértékek* határozzák meg. Másszóval, ha $\mathbf{x}_i^{(j)}$ $j = 1, \dots, J_i$ a mintapontok és $z_i^{(j)} = f(\mathbf{x}_i^{(j)})$, akkor

$$P\{M_i, f(M_i)\} := P\{\mathbf{x}_i^{(j)}, z_i^{(j)},\ j = 1, \dots, J_i\} \qquad i \in I$$

2.2. Megjegyzés. A konkrét algoritmikus megvalósítás során mind M *szekvenciális* felbontásának módja, mind pedig a mintapontok megválasztása egyértelműen adott. Pl. ha M egy n-dimenziós intervallum, akkor eléggé természetes egy *intervallumokra* való felbontást megadni: itt a mintapontok pl. a keletkező *részintervallumok* kiválasztott csúcsai lehetnek (lásd pl. a PINTÉR (1986a,b) dolgozatokban leírt partíciós algoritmusokat). Így — rögzített algoritmustípus esetén — a $P\{M_i, f(M_i)\}$ „rövid" jelölés nem vezet félreértéshez.

Partíciós algoritmus séma

0. (Kezdőértékek beállítása) Legyen az iterációs index értéke $k := 0$; $I_0 = \{0\}$ és $\{M_0\} = \{D\}$ definiálják a kezdeti (triviális) partíciót.

k. fő iterációs ciklus ($k = 0, 1, 2, 3, \dots$)

k.1. (Mintavétel) Minden $i \in I_k$ esetén válasszunk ki egy véges $x_i^{(j)} \in M_i$ mimntapont-halmazt és határozzuk meg az e pontokban felvett $z_i^{(j)}$ $j = 1, \dots, J_i$ függvényértékeket.

k.2. (Aggregáció) Az $\{\mathbf{x}_i^{(j)}, z_i^{(j)}, j = 1, \dots, J_i\}$ információ alapján határozzuk meg a partíció operátor $P\{M_i, f(M_i)\}$ értékét, minden $i \in I_k$ esetén.

k.3. (Kiválasztás) Válasszunk ki egy tetszőleges t indexet, amelyre fennáll a

$$P\{M_t, f(M_t)\} = \max_{i \in I_k} P\{M_i, f(M_i)\} \tag{2.2}$$

reláció: ekkor az M_t halmaz további felbontása következik.

k.4. (Finomítás) Az M_t halmazt (nem-triviális) véges, testszerű felbontásával helyettesítjük:

$$M_t = \bigcup_{\ell \in L_k} M_{t_\ell}, \qquad M_{t_\ell} \subset M_t \quad \ell \in L_k \tag{2.3}$$

így a finomított felbontást az

$$\{M_i : i \in I_{k+1}\} \qquad I_{k+1} = I_k \setminus \{t\} \cup \{t_\ell : \ell \in L_k\}$$

halmazok rendszere definiálja. Térjünk vissza a következő fő iterációs ciklus elejére ($k := k + 1$).

2.3. Megjegyzés. A bevezetett PAS algoritmus-séma lényegileg leírja (speciális esetként tartalmazza) pl. DANILIN–PIYAVSKII (1967), SHUBERT (1972), FUJII–ICHIDA–OZASA (1986), GALPERIN (1988), HANSEN (1979, 1980), KUSHNER (1964), ARCHETTI–BETRÓ (1979), MEEWELLA–MAYNE (1988), MLADINEO (1986), MOCKUS–TIESIS–ZILINSKAS (1978), RATSCHEK (1985), SHEN–ZHU (1987), STRONGIN (1978), WOOD (1985) és ZILINSKAS (1981) globális optimalizálási eljárásait. A felsorolt algoritmusok részleteik tekintetében persze jelentős különbségeket mutatnak és konvergencia-tulajdonságaik sem egységesek.

2.4. Megjegyzés. Amint azt a (2.2) kiválasztási szabály alátámasztja, a ***partíciós*** operátor $P\{M_i, f(M_i)\}$ értéke informálisan az M_i részhalmaz „értékének optimista becslését" fejezi ki. Például megemlítjük, hogy $P\{M_i, f(M_i)\}$ megadható az M_i halmazon felvett minimális függvényérték egy alsó becslése által (ilyen becslés pl. f *Lipschitz-folytonossága* alapján vagy intervallum-aritmetikai módszerekkel számítható); egy másik — ettől lényegesen eltérő — megközelítésben $P\{M_i, f(M_i)\}$ értékét annak valószínűsége fejezi ki, hogy az M_i halmazon az aktuális optimumbecslésnél jobb eredményt adó pontot találunk. (A globális optimalizálási módszerek alapjául szolgáló célfüggvény-modellek és az ezeknek megfelelő kereső stratégiák leírását illetően pl. STRONGIN (1978), ARCHETTI és BETRÓ (1980), ZILINSKAS (1982), PINTÉR (1983), BOENDER (1984) munkáira, ill. a fentiekben felsorolt PAS típusú algoritmusokra utalunk.) Itt jegyezzük meg, hogy a PAS legtöbb ismert realizációjában elegendő a felbontás finomításaként keletkező újabb részhalmazokon elvégezni a mintapontok kiválasztását (megtartva tehát a már létező M_i halmazokat „jellemző" $P\{M_i, f(M_i)\}$ értékeket: e körülmény numerikus előnyei nyilvánvalóak (lásd a PAS k.1–k.3. lépéseit).

2.2. Konvergencia-tulajdonságok és realizációs kérdések

Feltéve, hogy a (2.1) feladatot egy PAS típusú eljárással kívánjuk megoldani, ***szükséges és elégséges*** globális konvergencia-feltételek adhatók meg, a partíciós algoritmus megkövetelt, ill. indukált tulajdonságainak függvényében. Az alábbiakban csak a fő eredmények összefoglalására szorítkozunk, a további részleteket illetően az alább idézett munkákra utalunk.

i) A felsorolt PAS típusú módszerek két csoportra oszthatók. Az első csoport algoritmusai (pl. KUSHNER (1964), ARCHETTI–BETRÓ (1979), MOCKUS–TIESIS–ZILINSKAS (1978) és ZILINSKAS (1981) egyváltozós módszerei) elméletileg egy *mindenütt sűrű* mintapont-sorozatot generálnak a D halmazon: ilymódon a minimumbecslések sorozata a z^* optimumhoz tart, a megfelelő minimalizáló pontsorozat torlódási helyei pedig az (1.1) feladat globálisan optimális megoldásai (tehát az X^* halmaz elemei). Megjegyezzük, hogy az említett típusú algoritmusok (már) a célfüggvény folytonossága esetén is rendelkeznek a fenti értelmű *elméleti* konvergenciával. (A további részleteket illetően lásd a PINTÉR (1987) dolgozatban tárgyaltakat.)

ii) A partíciós algoritmusok második csoportja (ide tartozik pl. DANILIN–PIYAVSKII (1967), SHUBERT (1972), STRONGIN (1978) és BOENDER (1984) egyváltozós módszere, valamint az intervallum-aritmetikai alapú eljárások) — *Lipschitz-folytonos* célfüggvény esetén — elméletileg olyan mintapont-sorozatot generál, amelynek torlódási pontjai *azonosak* az X^* halmaz elemeivel; speciálisan, ha $X^* = \{\mathbf{x}^*\}$, akkor a mintapontok sorozata az (1.1) feladat globális optimumhelyéhez konvergál. (A részleteket illetően lásd PINTÉR (1986a,b, 1988a), illetve HORST és TUY (1987) dolgozatait.)

iii) Az említett (és más PAS típusú, intervallum-tartományon működő) egyváltozós algoritmusok közvetlenül általánosíthatók *többváltozós* módszerré — a megfelelő konvergenciatípus megtartásával (PINTÉR, 1986a,b). Ezek az általánosítások korábban nem voltak ismertek, mivel az egyváltozós algoritmusokhoz vezető célfüggvénymodellek és döntési szabályok kiterjesztése mind elméleti, mind pedig numerikus szempontból kivihetetlennek tűnt.

iv) Az így nyert többváltozós módszerek (véges) *téglatartományon* közvetlenül alkalmazhatók (PINTÉR, 1986a,b). Egyszerű módosítás teszi lehetővé az algoritmusok *konvex* D halmazon, illetve *csillagtartományon* történő alkalmazását PINTÉR, 1986c). Általános (esetleg nemkonvex, nem összefüggő) *Lipschitz-folytonos* függvények által definiált megengedett halmaz esetén a PAS megfelelő módosítása szükséges (amely lehetővé teszi D nem megengedett, illetve szuboptimális részhalmazainak eliminációját, valamint a megmaradó partíciós részhalmazokon a célfüggvényértékbecslések finomítását (PINTÉR, 1988a).

v) A javasolt algoritmus-koncepció „gazdaságosan" valósítható meg: ezt illusztrálandó, az alábbiakban egy *intervallum-felbontási stratégiát* ismertetünk (tehát az (1.1) feladatot a $D = [a, b] = [\underline{\mathbf{x}}_1, \bar{\mathbf{x}}_1]$ tartományon kívánjuk megoldani). Az alábbi algoritmus a DANILIN–PIYAVSKII–SHUBERT (DPS) módszer ún. *diagonális típusú kiterjesztésén* alapszik (PINTÉR, 1986b).

Diagonális típusú kiterjesztett DPS algoritmus

0. (Kezdőértékek beállítása) Legyen $k := 1$ az iterációs index; $m := 1$ a további

bontható részintervallumok aktuális száma; $\ell := 2$ pedig a mintapontok aktuális száma. A kiinduló keresési információt a D intervallumot meghatározó „bal alsó" és „jobb felső" $\underline{\mathbf{x}}_1, \bar{\mathbf{x}}_1$ csúcsok és a $\underline{z}_1 = f(\underline{\mathbf{x}}_1)$, $\bar{z}_1 = f(\bar{x}_1)$ függvényértékek adják.

1. (Intervallum kiválasztás) Válasszunk ki egy olyan részintervallumot, amelyen az

$$L\|\bar{\mathbf{x}}_i - \underline{\mathbf{x}}_i\| - \frac{1}{2}(\bar{z}_i + \underline{z}_i), \qquad i = 1, \dots, m$$

ún. ***karakterisztika-függvény*** (STRONGIN, 1978) értéke maximális; legyen a szóbanforgó részintervallum indexe t. (A képletben szereplő $L = L(f, D)$ az f célfüggvény *Lipschitz-konstansának* egy érvényes felső becslése kell, hogy legyen).

2. (A globális kereső fázis megállási szabálya) Ha a kiválasztott részintervallum „elég kicsiny" (pl. $\|\bar{\mathbf{x}}_t - \underline{\mathbf{x}}_t\| < \delta$, ahol $\delta > 0$ az eljárás paramétere), akkor térjünk át az algoritmus 4. lépésére; az ellenkező esetben a 3. lépésre térünk át.

3. (Újabb mintapont kiválasztása és a felbontás finomítása) Válasszuk ki az $[\underline{\mathbf{x}}_t, \bar{\mathbf{x}}_t]$ intervallum

$$\mathbf{x}_t^{\text{new}} = \frac{1}{2}(\underline{\mathbf{x}}_t + \bar{\mathbf{x}}_t) - \frac{(z_t - \underline{z}_t)}{2L\|\bar{\mathbf{x}}_t - \underline{\mathbf{x}}_t\|}(\bar{\mathbf{x}}_t - \underline{\mathbf{x}}_t)$$

pontját ($\ell := \ell + 1$). Az $\mathbf{x}_t^{\text{new}}$ pont a t. részintervallum *kettéosztását* határozza meg: a „szétvágást" azzal a hipersíkkal adjuk meg, amelyik tartalmazza az $\mathbf{x}_t^{\text{new}}$ pontot és merőleges az $[\underline{\mathbf{x}}_t, \bar{\mathbf{x}}_t]$ intervallum (tetszőleges) leghosszabb élére. Ezt követően a t. részintervallumot felbontásával helyettesítjük ($m := m + 1$) és meghatározzuk a célfüggvény értékét az így nyert új (intervallum-csúcspont) mintapontokban is ($\ell := \ell + 2$). Innen visszatérünk a következő iterációs ciklus 1. lépésére ($k := k + 1$).

4. (A megoldás lokális kereséssel történő finomítása) A talált optimumbecslésből (a legkisebb célfüggvényértékhez tartozó argumentumból) kiindulva, azt alkalmas konvex programozási módszerrel pontosítjuk.

Megjegyezzük, hogy a fenti algoritmus-váz 1. és 3. lépése a PAS k.1 – k.3., illetve k.4. lépéseinek felel meg, míg a 2. és 4. lépések numerikus szempontból fontos kiegészítésnek tekinthetők. (A DPS algoritmus-váz numerikus hatékonysága természetesen számos további kiegészítéssel növelhető, az elméleti konvergencia fenntartása mellett.) Kiemeljük, hogy a fenti eljárás csak a generált részintervallumok főátlójának végpontjai által meghatározott információt használ fel. Így a tárolási és számítási igény (csak) *lineáris függvénye* mind a feladat dimenziójának (n), mind pedig az algoritmus által végrehajtott fő iterációk számának (k). Megjegyezzük még, hogy ***részintervallum-specifikus Lipschitz-konstans becslések*** is megadhatók (pl. további mintapontok felvétele útján): ez az opció viszonylag kis számítási ráfordítással jelentős hatékonyságnövekedéshez vezethet.

Mielőtt néhány példát mutatnánk a globális optimalizálási modellek és módszerek alkalmazására, megjegyezzük, hogy a vázolt DPS algoritmus-séma egy (konjugált irányokon alapuló lokális kereső eljárással és számos numerikus finomítással) kiegészített változatát FORTRAN nyelven programoztuk. Az alább bemutatott numerikus eredményeket a *Vízgazdálkodási Tudományos Kutató Központ* IBM

PC/XT, AT és MICROVAX-II kompatibilis számítógépeinek felhasználásával kaptuk.

3. Alkalmazási lehetőségek

Amint azt a bevezetésben már megjegyeztük, a tudományos modellalkotás, műszaki rendszertervezés, döntéselőkészítés fontos problématípusai vezetnek többextrémumú optimalizálási feladatokhoz: itt a globális megoldás előállítását egészen a legutóbbi évekig kilátástalan numerikus feladatnak tekintették. Bár e feladatok lényegüket tekintve (inherens értelemben) bonyolultak és megoldásuk számításigényes marad, a globális optimalizáció területén elért új eredmények lehetővé teszik *elméletileg korrekt és implementálható* algoritmusok alkalmazását. Az elmondottak illusztrálása céljából három gyakorlati szempontból jelentős feladattípust elemzünk az alábbiakban; az itt nem ismertetett további részletek tekintetében a PINTÉR (1986b, 1988b), PINTÉR, SZABÓ és SOMLYÓDY (1986), PINTÉR és PESTI (1988) dolgozatokra utalunk.

3.1. Nemlineáris egyenletrendszerek megoldása.

Ez a problémakör a numerikus analízis egyik alapvető klasszikus feladatának tekinthető. Tekintsük az

$$f_j(x) = 0, \quad j = 1, \dots, m, \quad x \in D \subset R^n. \tag{3.1}$$

egyenletrendszert, amelynek általános értelemben vett *legjobb (közelítő) megoldását* keressük. Megjegyezzük, hogy *nem* élünk sem a szokásos $m=n$ feltétellel, sem pedig a megoldás létezésére, illetve egyértelmű voltára vonatkozó feltevéssel: így pl. megoldással nem rendelkező, illetve több (véges számú) megoldással is bíró egyenletrendszerek vizsgálatát sem zárjuk ki.

A (3.1) feladat elemzésére irányuló, igen széleskörű irodalomból pl. ALLGOWER és GEORG (1983), DENNIS és SCHNABEL (1983), FORSTER, szerk. (1980), GARCIA és ZANGWILL (1981), ORTEGA és RHEINBOLDT (1970) vagy ROBINSON szerk. (1980) munkáit említjük. Az e munkákban tárgyalt eredmények alapján világos, hogy a „klasszikus" módszerek (standard egyváltozós algoritmusok, fixpont-eljárások, homotópia-módszerek, *Newton-típusú* algoritmusok és más konvex programozási technikák) a (3.1) feladat megoldását *általános* értelemben nem biztosítják. E negatív tény elsődleges oka a felsorolt megközelítési módok *lokális* szemlélete és alkalmazásuk lokális feltételei, amelyek teljesülése nem garantált. Pl. „megfelelő" kezdőpont híján (csupán) lokálisan legjobb közelítő megoldást kaphatunk, tehát a „megoldás" reziduális hibáinak összege pozitív. Emellett (3.1) szingularitása, rosszul kondicionált volta, analitikus (simasági stb.) tulajdonságok hiánya még elméletileg konvergens módszerek esetén is komoly numerikus nehézségekhez vezethet. Természetesen a (3.1) általános feladattípus nehezen kezelhető marad: ennek ellenére a globális optimalizáció szemlélete alapvetően redukálja, illetve megszünteti a felsorolt bonyodalmak jelentős részét. (Itt jegyezzük meg, hogy (3.1) típusú (egy-, kétváltozós) feladatok globális optimalizáció keretében történő megoldásával pl. STRONGIN (1978), VYSOTSKAYA és STRONGIN (1983), HORST és THOAI (1988) foglalkozott.)

A 2. szakaszban összefoglalt elméleti eredmények figyelembevételével feltesszük, hogy

i) A D halmaz kompakt testszerű;

ii) a (3.1) rendszert alkotó függvények *Lipschitz-folytonosak* a D halmazon;

iii) a (3.1) rendszer (legjobb közelítő) megoldásainak X^* halmaza véges.

(A ii) feltétel kapcsán felidézzük, hogy az teljesül pl. folytonosan differenciálható függvények egyenletrendszerére.)

Az alábbiakban a (3.1) feladatot egy *ekvivalens* globális optimalizálási problémává írjuk át. Ennek érdekében vezessük be a következő jelölést:

$$\mathbf{F}^T(\mathbf{x}) = (f_1(\mathbf{x}), \dots, f_m(\mathbf{x})) \quad \mathbf{F} : R^n \to R^m, \tag{3.2}$$

amelynek felhasználásával (3.1) az

$$\mathbf{F}(\mathbf{x}) = 0 \quad (0 \in R^m), \quad \mathbf{x} \in D \tag{3.3}$$

alakban írható fel. Természetes követelmény, hogy a (3.1) transzformációjával nyert probléma megoldáshalmaza azonos legyen az X^* halmazzal: az ilyen átalakításokat *megfelelőnek* nevezzük.

Egyszerű módszerrel kaphatunk *norma által indukált* megfelelő transzformációkat: legyen ui. N az R^m tér egy tetszőleges normája és definiáljuk az

$$f(\mathbf{x}) = N(\mathbf{F}(\mathbf{x})) = \|\mathbf{F}(\mathbf{x})\|_N, \quad \mathbf{x} \in D \tag{3.4}$$

függvényt, majd a

$$\min f(\mathbf{x}), \quad \mathbf{x} \in D \tag{3.5}$$

optimalizálási feladatot. (Megjegyezzük, hogy ha a (3.1) feladatnak létezik megoldása, akkor valamennyi — norma által indukált — (3.4)–(3.5) típusú transzformáció elméleti szempontból ekvivalens; ha viszont (3.1)-nek nincs megoldása, akkor az *általánosított megoldás* természetes módon definiálható, mint a (3.5) feladat *normafüggő* legjobb közelítő megoldása.) Elemi úton belátható, hogy az f függvény *Lipschitz-folytonos*: ennek következtében a PAS algoritmus-séma tetszőleges realizációja alkalmazható a (3.1) feladat megoldására.

Illusztrativ eredmények

A javasolt megközelítési mód tesztelése érdekében két *pszeudo-véletlen* feladatosztályt vizsgáltunk részletesebben. A véletlenszerűen generált feladatosztályokon végzett vizsgálatokat (egyedi tesztproblémák megoldásával szemben) azért tartjuk előnyösnek, mert i) — az adott függvényosztályra nézve — a szóbanforgó módszer statisztikailag megalapozott kiértékelését teszik lehetővé, enyhe valószínűségi feltevések mellett, emellet ii) a tesztproblémák egyszerűen generálhatók és reprodukálhatók, megkönnyítve különböző algoritmusok objektív összevetését. Az alábbiakban illusztratív eredményeket közlünk; a további részletek a PINTÉR (1988b) dolgozatban találhatók meg.

Trigonometrikus egyenletrendszerek.

Először a következő típusú egyenletrendszereket vizsgáltuk:

$$\text{(3.6)}\quad \begin{aligned} f_i(\mathbf{x}) &= \sum_{j=1}^{n} \{A_{ij}\sin^2[B_{ij}(x_j - x_j^*)] + C_{ij}\cos^2[D_{ij}(x_j - x_j^*) + \pi/2]\} \\ &\quad + (x_i - x_i^*)^2 = 0, \qquad i = 1,\dots,n \\ \mathbf{x} \in D &= \{\mathbf{x} = (x_1,\dots,x_n) : -\pi \le x_i \le \pi\} \end{aligned}$$

amelyek x_i^*, A_{ij}, B_{ij}, C_{ij}, D_{ij}, $i,j = 1,\dots,n$ paraméterei adott szekvenciális generálási szabályt követnek. (A generálási szabály biztosítja, hogy az egyetlen $\mathbf{x}^*$ globális optimumhely mellett még véletlen számú *általánosított lokális megoldás* is létezzék, minden tesztfeladat esetén.)

A kiterjesztett DPS algoritmus alkalmazásával néhány száz egy–három változós (3.6) típusú problémát oldottunk meg, az algoritmus különböző paraméterezései mellett. (A parametrizációk az alkalmazott megállási szabályt és a *Lipschitz-konstans* numerikus becslésének módját határozták meg.) A számítások során az

$$\text{(3.7)}\qquad f(\mathbf{x}) = \left|\sum_{i=1}^{n} f_i(\mathbf{x})\right|$$

célfüggvényalakot használtuk.

Illusztrációképpen egy futtatás eredményeit az alábbiakban foglaljuk össze: itt $n = 2$ (a változók száma); $\alpha = 1$ (globális megállási kritérium: lokális keresésre térünk át, ha a (3.7) célfüggvény legjobb talált értéke α alá csökken; $\delta = 0.6$ (globális megállási kritérium: lokális keresésre térünk át, ha a DPS séma 2. lépésében kiválasztott részintervallum ℓ_1 — normája δ alá csökken); $L = 20\sqrt{2}$ (a *Lipschitz-konstansra* nézve alkalmazott globális felső becslés); *maxfct* =200,300,500,2000 (*maxfct* a globális kereső fázisban végezhető célfüggvénykiértékelések maximális száma, amelynek bevezetésével az algoritmus megbízhatóságát és hatékonyságát kívántuk elemezni). Ötven véletlenszerűen generált tesztfuttatás eredményét az 1. *táblázatban* összegezzük: itt $E(nglob)$, $D(nglob)$, $E(nloc)$, $D(nloc)$ rendre a globális, illetve a lokális fázisban végzett kereső lépések számának várható értékét, illetve szórását jelöli. (Megjegyezzük, hogy az egyenletrendszerek megoldását az algoritmus az $\mathbf{x}^*$ argumentum szerint hét tizedes jegy pontossággal szolgáltatta.)

1. Táblázat

Trigomometrikus egyenletrendszerek ($n = 2$): a teszteredmények összegzése

maxfct	sikeres megoldások száma	$E(nglob)$	$D(nglob)$	$E(nloc)$	$D(nloc)$
		(kerekített értékek)			
200	32	189	35	17	7,1
300	36	263	71	17	7,2
500	44	335	177	15	6,5
2000	50	710	486	11	2,8

Amint azt a fenti táblázat tükrözi, a *maxfct* paraméter növelése az algoritmust „tökéletesen megbízhatóvá" teszi, de természetesen a számítási igény várható értéke és szórása is növekszik (a lokális jellemzők változása kismértékű). Figyelemreméltó, hogy a feladatok mintegy kétharmadát az algoritmus 200-nál kevesebb lépésben oldja meg.

Shekel-típusú egyenletrendszerek

A második itt vizsgált feladattípus a globális optimalizáció irodalmából (DIXON és SZEGŐ, szerk. 1978) ismeretes: keresendő az

$$
(3.8)\qquad \begin{aligned} & f_i(\mathbf{x}) = (x_i - x_i^*)^2 - \sum_{j=3(i-1)+1}^{3i} \frac{1}{\|\mathbf{x} - \mathbf{d}_j\| + c_j} + k_i = 0, \quad i = 1, \dots, n \\ & \mathbf{x} \in D = \{\mathbf{x} = (x_1, \dots, x_n) : \ 0 \le x_i \le 10\} \end{aligned}
$$

egyenletrendszer megoldása. A korábbiakhoz hasonlóan, a (3.8) függvények (tehát az x_i^*, k_i, c_j, d_j, $i = 1, \dots, n$, $j = 1, \dots, 3n$ értékek) generálása szekvenciálisan történik: a generálási szabályok itt is biztosítják az $\mathbf{x}^*$ globális optimumhely unicitását; emellett a (3.8) egyenletek mindegyikének három lokálisan legjobb közelítő megoldása van. Így az ℓ_2-norma felhasználásával nyert

$$
(3.9)\qquad \begin{aligned} & \min f(\mathbf{x}) \qquad \mathbf{x} \in D \\ & f(\mathbf{x}) = \sum_{i=1}^{n} f_i^2(\mathbf{x}) \end{aligned}
$$

optimalizálási feladatnak $3n$ számú általánosított lokális megoldása van.

A korábban leírt tesztfuttatásokhoz hasonlóan, az α és δ paraméterek által jellemzett megállási szabályokat alkalmaztuk; ezúttal azonban lokális (***részintervallum-specifikus***) ***Lipschitz-konstans*** becslést alkalmaztunk, a generált részintervallumokon kiválasztott (további) mintapontok felhasználásával. (Megjegyezzük, hogy ez a megközelítés elméleti szempontból korrekt és várhatóan számottevő numerikus előnyökhöz vezet, az „univerzális" — általában véve eléggé durva — *Lipschitz-konstans* becsléshez képest.)

Ötszáz tesztfeladat megoldását végeztük el (100–100 feladatot generálva $n = 1, \ldots, 5$ esetén); a numerikus kisérletek számítási igényét csökkentendő, csak a globális kereső fázist hajtottuk végre. Így a feladatot sikeresen megoldottnak tekintettük (a globális keresés szintjén), ha

i) a szóbanforgó egyenletrendszer összegzett négyzetes hibája kevesebb volt, mint $\alpha = 0,25n$; vagy

ii) a $\delta = 0,01n$ befejezési feltétel bekövetkezése esetén a globális optimumhely és annak számított közelítése között az ℓ_1-normában mért távolság $0,25n$ alatt maradt.

A részletes számítógépi eredmények tanusága szerint a legtöbb esetben a fenti megállási szabályok olyan közelítő megoldásokhoz vezettek, amelyek komponensenkénti hibája 0,01–0,05 körüli érték volt; még a kevésszámú „rossz" megoldás is legfeljebb 0,3–0,4 komponensenkénti hibával rendelkezett. Ilyen pontosság eléréséhez egyszerű rács-kereséssel (*grid search*) legalább 200^n függvénykiértékelésre volna szükség.) A teszteredményeket a 2. táblázatban foglaljuk össze.

2. táblázat

Shekel-típusú egyenletrendszerek: a teszteredmények összegzése

változók száma	sikeres megoldások száma (100 közül)	$E(nglob)$ (kerekített értékek)	$D(nglob)$ (kerekített értékek)
1	100	11	13
2	98	137	203
3	100	522	407
4	100	2138	820
5	99	10471	6499

Amint azt már megjegyeztük, még (három esetben 500 között) „sikertelen" megoldások is közel voltak a megfelelő $\mathbf{x}^*$ értékhez. A részletes futáseredmények azt mutatták, hogy néhány feladat igen nehéz volt (pl. az egyik 5-változós probléma megoldása mintegy 22000 lépést igényelt, bár a legtöbb hasonló feladat esetén kb. 1500 lépés elegendőnek bizonyult, sőt jónéhány feladat megoldása 200 alatti lépésszámmal elérhető volt.) Végül megjegyezzük, hogy az egy- és kétváltozós feladatok megoldása pár másodpercet vett igénybe, a három- és négyváltozós problémáké átlagosan kb. 0,4, illetve 2,5 percig tartott (IBM PC/AT kompatibilis gépen, aritmetikai koprocesszor nélkül); az ötváltozós problémák átlagos megoldási ideje pedig kb. 25 másodperc volt (egy MICROVAX-II kompatibilis számítógépen).

3.2. Modellkalibráció (Nemlineáris approximáció).

Fizikai, kémiai, biológiai, ökológiai stb. rendszerek, folyamatok és jelenségek modellezésének egyik kulcsfontosságú kérdése a (már identifikáltnak feltételezett) rendszerleírás *numerikus kalibrációja*, vagyis a modell olyan paraméterezése, amelyik a megfigyelési adatokkal, mérésekkel a leginkább összhangban van. Tekintsük pl. a (skalár) idősormegfigyelésen alapuló *nemlineáris* paraméterbecslési feladatot: itt az összegzett hibafüggvény

$$f(\mathbf{x}) = \sum_{t=t_{\min}}^{t_{\max}} [\tilde{y}_t(\mathbf{x}) - y_t]^2, \qquad \mathbf{x} \in D \tag{3.10}$$

(vagy más alkalmas, norma által indukált eltérési mérőszám) lehet a célfüggvény: y_t a megfigyelési értékeket, $\tilde{y}_t(\mathbf{x})$ pedig a modell output értékeket jelöli a $t = t_{\min}, \ldots, t_{\max}$ időpontokban. A modell output természetesen egy $\mathbf{x}$ paramétervektor függvénye, amely tehát a D tartományból választandó ki optimális módon.

A kalibrációs feladat (és általánosabban: a nemlineáris approximációs probléma) tipikusan többextrémumú voltát az alábbiakban illusztráljuk. Tegyük fel, hogy az $\tilde{y}$ modellkimenet az $\mathbf{x}$ vektor kétszer folytonosan differenciálható függvénye. Ekkor

$$\begin{aligned}
\nabla f(\mathbf{x}) &= 2 \sum_{t=t_{\min}}^{t_{\max}} [\tilde{y}_t(\mathbf{x}) - y_t] \nabla \tilde{y}_t(\mathbf{x}) \\
\nabla^2 f(\mathbf{x}) &= 2 \sum_{t=t_{\min}}^{t_{\max}} [\tilde{y}_t(\mathbf{x}) - y_t] \nabla^2 \tilde{y}_t(\mathbf{x}) + 2 \sum_{t=t_{\min}}^{t_{\max}} \nabla \tilde{y}_t(\mathbf{x}) \nabla \tilde{y}_t^T(\mathbf{x})
\end{aligned}$$

Itt a $\nabla^2 f(\mathbf{x})$ kifejezésben szereplő második összeg pozitív szemidefinit diagonális mátrixokból áll; az első összeg viszont tipikusan indefinit mátrixok összege (az $\mathbf{x}$ paraméter változásainak függvényében). Így a $\nabla^2 f(\mathbf{x})$ *Hesse-mátrix* gyakran ***indefinit*** az f függvény tehát lokálisan konvex, illetve konkáv lehet: másszóval, a (3.10) feladatnak több lokális extrémuma lehet. (Lásd ezzel kapcsolatban még pl. CSENDES (1988) dolgozatának 2. szakaszát, amely a (3.10) feladattal ekvivalens probléma struktúráját hasonló szempontból elemzi). Végül jegyezzük meg, hogy a nemlineáris egyenletrendszerek megoldása természetesen az alábbiakban tárgyalt feladattípus speciális esete.

A globális optimalizációval történő modell-illesztésre példaként egy víz-üledék kölcsönhatást leíró modell kalibrációját említjük meg. A szélkeltette felkeveredés jelensége sekély tavak vízminőségének (mesterséges eutrofizációjának) alakulásában jelentős szerepet játszhat (lásd pl. SOMLYÓDY és VAN STRATEN, szerk. (1986) vagy PINTÉR, SZABÓ és SOMLYÓDY (1986) munkáit). Ezért a lebegőanyag-koncentráció (C) dinamikájának meghatározása elsőrendűen fontos. Ha egy jól-kalibrált leíró modellel rendelkezünk, akkor „olcsó" szélsebesség (W) mérések alapján becsülhető C változása.

Tegyük fel, hogy C dinamikáját közelítőleg az alábbi közönséges differenciálegyenlet írja le:

$$\frac{dC}{dt} = -b_1 C + b_2 W^n \tag{3.11}$$

A (3.11) egyenletben W a szélsebesség abszolút értéke (10 méterrel a víztükör felett), C pedig a lebegőanyag-koncentráció (térfogatra átlagolt) értéke; végül b_1, b_2 és n meghatározandó modellparaméterek. (Megjegyezzük, hogy az egyenlet jobboldalának első tagja a kiülepedési fluxust közelíti, a második tag pedig a szélkeltette felkeveredést reprezentálja.)

A (3.11) differenciálegyenlet analitikus megoldását úgy közelítjük, hogy állandó W_t szélsebességet tételezünk fel az egyenlőközű $t_{\min} \leq t \leq t_{\max}$ időpontokban. Így a

$$\tilde{C}_t = \tilde{C}_t(b_1, b_2, n) = \tilde{C}_{t-\Delta t} \exp(-b_1 \Delta t) + \frac{b_2}{b_1} W^n_{t-\Delta t}[1 - \exp(-b_1 \Delta t)]$$

(Δt : az egymást követő szélmérések közötti időperiódus)

approximáció paramétereit kell megválasztani, egy (3.10) típusú célfüggvénynek megfelelően. Ez azt jelenti, hogy — tetszőleges modellparaméterezést rögzítve — egy differenciálegyenlet numerikus megoldását kell végrehajtani a célfüggvény kiértékeléséhez. (Ez a szituáció tipikusnak tekinthető számos kalibrációs feladat esetén, amikor tui. a vizsgált döntési variánsok számításigényes, algoritmikus meghatározása szükséges.)

A vázlatosan ismertetett fenti alkalmazás esetén a javasolt globális optimalizálási módszer a kalibráció minőségét (a (3.10) célfüggvény értékét) mintegy 20 százalékkal javította, a gyakran használt *Levenberg–Marquardt paraméter optimalizációs algoritmus* ismételt alkalmazásával nyert *legjobb* paraméterezéshez viszonyítva.

3.3. Adatosztályozás (cluster-analízis).

Egy véges elemszámú halmaz elemeinek megadott kritérium szerinti csoportosítása az operációkutatás és a matematikai statisztika számos feladatának lényeges része. Példaként kiszolgálóhelyek telepítése (*facility location*, RAO (1971) vagy MULVEY és COWDER (1979)), adatosztályozás (HARTIGAN (1975), GORDON (1981)) vagy bizonyos optimalizálási algoritmusok (TÖRN (1978), RINNOOY KAN és TIMMER (1987a,b)) említhetők.

A feladat megfogalmazása

A HANSEN és JAUMARD (1987) dolgozat jelöléseit követve, legyen

$O = \{O_k,\ k = 1, \dots, N\}$	az osztályozandó elemek halmaza
$DM = \{d_{k\ell},\ k, \ell = 1, \dots, N\}$	$d_{k\ell} = d(O_k, O_\ell)$ az elempárok különbségének mérőszáma
$P_M = \{C_1, \dots, C_M\}$	az O halmaz egy M clusterre történő felbontása (ahol tehát $C_j \neq \emptyset$, $C_i \cap C_j = \emptyset$, $\bigcup_j C_j = O$)
Π_M	a lehetséges P_M particiók halmaza

A bevezetett jelölésekkel az „optimális halmazfelbontási probléma" (OHP) a következőképpen fogalmazható meg:

Keresendő az a $P_M \in \Pi_M$ partíció, amelyik az O halmaz „leginkább homogén" (vagy „leginkább diszkriminatív") osztályozását biztosítja.

A fenti verbális feladat-megfogalmazás kapcsán megjegyezzük, hogy pl. orvosi, biológiai, környezeti adatok osztályozása azok „hasonlóságát" kell, hogy tükrözze. Kiszolgálóhelyek telepítése esetén azok a „fogyasztók" kerülnek azonos osztályba, amelyeket ugyanaz a kiszolgálóegység lát el. E modell-típusok esetén tehát a d különbözőségi mérőszámot pl. egy, az O halmazt tartalmazó térben definiált alkalmas távolságfogalom adhatja meg. (Megjegyezzük, hogy bár a $d_{kk} = 0$, $d_{k\ell} \geq 0$ és $d_{k\ell} = d_{\ell k}$ relácók teljesülését általában előírják, a „háromszög-egyenlőtlenség" megkövetelése nem minden esetben indokolt.)

Természetesen az (OHP) kvantitatív megfogalmazása sokféle, nem-ekvivalens módon történhet. Ezt illusztrálandó, definiáljuk az alábbi fogalmakat:

$$d(C_j) = \max_{O_k, O_\ell \in C_j} d_{k\ell} \qquad \text{a } C_j \text{ cluster } \textit{átmérője}$$

$$r(C_i, C_j) = \min_{O_k \in C_i,\ O_\ell \in C_j} d_{k\ell} \qquad \text{a } C_i \text{ és } C_j \text{ közti } \textit{diszkrimináció}$$

$$s(C_j) = \min_{O_k \in C_j,\ O_\ell \notin C_j} d_{k\ell} \qquad \text{a } C_j \text{ cluster } \textit{felbontása (split)}$$

Bevezetésükkel (OHP) pl. a következő módokon fogalmazható meg optimalizálási feladatként:

$$\begin{aligned}
&\min_{P_M \in \Pi_M} \| \{d(C_1), \dots, d(C_M)\} \|_p \\
(3.12) \qquad &\max_{P_M \in \Pi_M} \| \{r(C_i, C_j)\},\ 1 \leq i, j \leq M\} \|_p \\
&\max_{P_M \in \Pi_M} \| \{s(C_1), \dots, s(C_M)\} \|_p,
\end{aligned}$$

ahol $\| \ \|_p$ egy alkalmas ℓ_p-norma és a $\{ \ \}$ zárójel a megfelelő homogenitási/inhomogenitási mérőszámok vektorát foglalja össze. (Konkrét példák tekintetében a *cluster-analízis* már idézett irodalmára utalunk.)

A numerikus megoldás kérdéseire áttérve, először is felidézzük azt a tényt, hogy az *egzakt* halmazfelbontási módszerek kombinatorikus jellegük következtében már igen kis méretű feladatok esetén is rendkívül számításigényesek. Pl. a kiszolgálórendszer tervezésének feladata nem-polinomiális (NP) bonyolultságú. BRUCKER (1978) és WELCH (1983) megmutatták, hogy (3.12) néhány speciális esete szintén NP-nehézségű $M \geq 3$ számú cluster esetén; továbbá, még az $M = 2$ esetben is a legjobb ismert algoritmusok műveletigénye $O(N^3 \log N)$, illetve $O(N^5)$, ahol N az osztályozni kívánt halmaz elemszáma (lásd HANSEN és JAUMARD, 1987).

Az egzakt módszerekkel szemben „realizálható" *közelítő* osztályozások is ismertek: ezek bizonyos heurisztikus elvekre is építenek és csupán egy „lokálisan legjobb" osztályozás numerikus approximációját biztosítják.

Az elmondottak alapján indokolt olyan algoritmikus osztályozási eljárások keresése, amelyek nem csupán (elméletileg korrekt módon) a globálisan legjobb csoportosítás kiválasztására irányulnak, hanem numerikus szempontból is elfogadhatóan realizálhatók. Az alábbiakban a PINTÉR és PESTI (1988) dolgozatban javasolt módszert ismertetjük. Az eljárás alkalmazása során feltesszük, hogy az osztályozandó objektumok valós vektorokkal adhatók meg: amint azt a felsorolt alkalmazási példák mutatják, ez az eset eléggé gyakran fordul elő.

Magpontok által indukált osztályozási eljárás-séma

Rendeljünk hozzá minden C_j clusterhez egy s_j *magpontot* (*seed point*); nem követeljük meg, hogy s_j az O halmazhoz tartozzék, de legyen eleme az O halmazt tartalmazó S vektortérnek. A *magpontok*

$$SP_M = \{s_1, \dots, s_M\}$$

halmazának bevezetése lehetővé teszi ***szekvenciális osztályozási stratégiák*** általános definiálását. Pontosabban, tegyük fel, hogy tetszőleges, a magpontok által indukált részleges $\{C'_j\}$ osztályozás (amelyre nézve fennáll $s_j \in C'_j$, $\bigcup_{j=1}^{M} C'_j \subseteq O \cup SP_M$) minősége egy $g(C'_1, \dots, C'_M)$ funkcionál értéke alapján megadható. (Természetesen a g függvény értéke általában a d távolságfogalom valamely implicit függvénye, ezt azonban nem jelöljük külön.) Ekkor — tetszőlegesen választott SP_M halmaz esetén — egy általános algoritmus-váz adható meg.

Magpontok szerinti osztályozás

0. (A kezdőértékek beállítása) Adott O halmaz, DM mátrix és SP_M halmaz esetén legyen k:=1 (iterációs index), $O^{(k)}:=\emptyset$ (az osztályozott, O-hoz tartozó pontok aktuális halmaza), $C_j^{(k)} := \{s_j\}$, $j = 1, \dots, M$ (a kiinduló, magpontok által indukált clusterek).

1. (Szekvenciális osztályozás) A következő O_k elemet a $j = j(k)$ indexű clusterhez soroljuk, ha

$$g\left(C_1^{(k)},\dots,C_j^{(k)}\cup\{O_k\},\dots,C_M^{(k)}\right) = \operatorname*{ext}_i g\left(C_1^{(k)},\dots,C_i^{(k)}\cup\{O_k\},\dots,C_M^{(k)}\right) \tag{3.13}$$

A (3.13) képletben szereplő „ext" a g kritériumfüggvény megfelelő extrémumát jelöli. Másszóval: a következő O_k elemet ahhoz a clusterhez soroljuk, amelyre nézve a *feladat-függő* alakú g függvény értéke a „legjobb". (Egyenlőség esetén a (3.13) relációt kielégítő minimális indexű C_j halmazt választjuk ki.)

2. (Módosítás) Legyen $O^{(k+1)} = O^{(k)} \cup \{O_k\}$; $C_j^{(k+1)} = C_j^{(k)} \cup \{O_k\}$; $C_i^{(k+1)} = C_i^{(k)}$, $i \neq j$. A $k := k+1$ értékadással az 1. lépésre térünk vissza, amig $k \leq N$; az ellenkező esetben az osztályozási eljárás végetért.

A leírt algoritmus-váz kapcsán megjegyezzük, hogy annak filozófiája hasonló a *cluster-medoid (-medián)* fogalmán alapuló módszerekéhez (lásd pl. KAUFMAN and ROUSSEEUW (1987) dolgozatát). A *medoid-fogalmon* alapuló és a jelen megközelítés közötti lényegi különbséget az jelenti, hogy itt nem követeljük meg, hogy a magpontok ténylegesen létező (megfigyelt) objektumokat írjanak le (így a magpontok „ideális medoidokként" is interpretálhatók). Bár bizonyos problémák esetén ez a feltevés nem jogos, elfogadása mind koncepcionális, mind pedig numerikus előnyökkel bír. Először is, nem evidens, hogy a rendelkezésre álló megfigyelési adatok *tökéletesen* reprezentálják az osztályozandó adathalmazt (gondoljunk pl. diagnosztikai adatok csoportosítására). Másodszor, ha a magpontok *folytonos* szerkezetű halmazon kereshetők, akkor a már említett kombinatorikus nehézségek numerikus következményei elkerülhetők (helyettesíthetők). Ezzel kapcsolatban megjegyezzük, hogy — tetszőleges magpont-konfiguráció esetén — legfeljebb $O(MN)$ az algoritmus műveletigénye. Ezért, ha SP_M optimális megválasztása hatékonyan közelíthető, akkor — a konkrét feladat méretétől függő — numerikus előnyök várhatók a globális optimalizálási megközelítéstől. Megjegyezzük még, hogy a (3.13) feltétel csak a magpontok függvénye, akkor a javasolt algoritmus-séma *független* az O halmaz elemeinek aktuális sorrendjétől: ez a módszer bizonyos *objektivitását* tükrözi (szemben jónéhány ismert heurisztikus eljárással).

A konkretizáció kedvéért az algoritmus-váz 1. lépésében tekintsük a következő („legközelebbi magpont") osztályozási szabályt: az O_k elemet a C_j clusterba soroljuk, ha fennáll

$$d(s_j, O_k) = \min_i d(s_i, O_k) \quad (d \text{ a távolságfüggvény}) \tag{3.14}$$

Az SP_M halmaz meghatározása globális optimalizációval

Ismeretes, hogy a *„medoid-típusú"* klasszifikációs eljárásoknak is számos lokálisan legjobb megoldása létezhet, tehát ezek általában többextrémumú optimalizálási feladathoz vezetnek. Ennek megfelelően a „standard” (konvex) optimalizálási eljárások nem alkalmasak a legjobb magpont-rendszer kiválasztására, indokolt tehát globális kereső módszert választani.

Az egyszerűség kedvéért tegyük fel, hogy a magpontokat egy m-dimenziós intervallumon kell kiválasztani: ilyen tartomány megadása nem nehéz, hiszen ha az O_k elemek $O_{k,i}$, $i = 1, \dots, m$ koordinátái ismertek, akkor a magpontok $s_{\ell,i}$ komponenseire nézve általában előírhatók a

$$\min_k O_{k,i} \leq s_{\ell,i} \leq \max_i O_{k,i} \quad \ell = 1, \dots, M \quad i = 1, \dots, m \tag{3.15}$$

feltételek. A (3.15) relációk által meghatározott tartományt D-vel jelölve, az osztályozási feladatot globális optimalizálási problémaként fogalmazhatjuk meg:

$$\min f(\mathbf{x}), \qquad \mathbf{x} \in D \tag{3.16}$$

Itt $\mathbf{x}$ a teljes magpont-konfigurációt jelöli:

$$\mathbf{x} = \{s_{1,1}, \dots, s_{1,m}, s_{2,1}, \dots, s_{2,m}, \dots, s_{M,1}, \dots, s_{M,m}\} \tag{3.17}$$

($\mathbf{x}$ tehát $M \cdot m$-komponensű vektor), az f függvény pedig a magpont-konfiguráció „jóságát” méri (lásd pl. a (3.12) formulák által kifejezett kritériumokat).

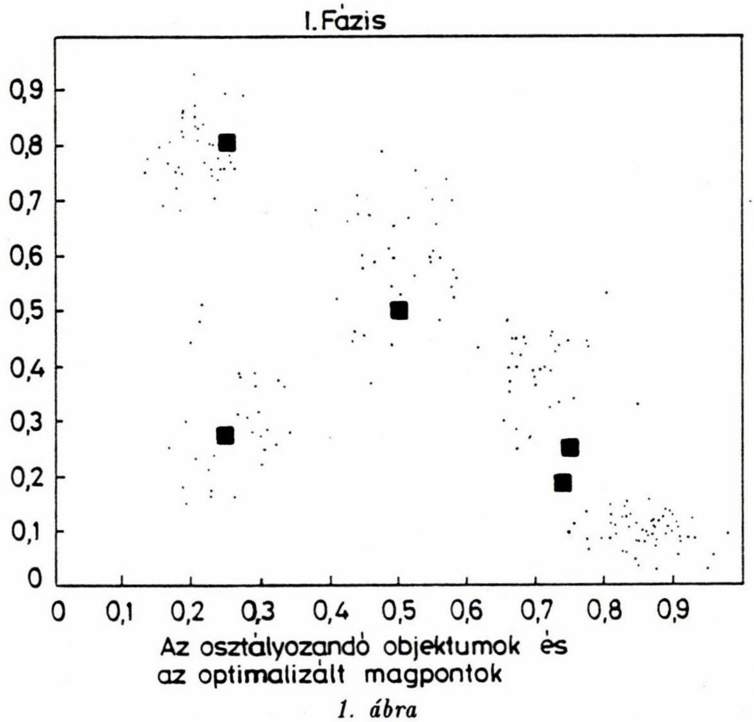
A korábban összefoglalt elméleti eredményeknek megfelelően, a megoldás módszereként pl. a D tartomány adaptívan generált részintervallumokra történő felbontása alkalmazható.

Numerikus eredmények

A javasolt megoldási módszer alkalmazhatóságát illusztrálandó, egy tesztfeladat megoldását foglaljuk össze az alábbiakban (a további részleteket illetően lásd a Pintér és Pesti (1988) dolgozat numerikus eredményeit). A feladatot véges számú, pszeudo-véletlen objektum (esetünkben: kétváltozós, az egység-intervallumba eső valós vektor) osztályozása jelenti (ezek pl. a „fogyasztók” halmazának elemeit jelölhetik egy kiszolgálóhely-telepítési feladatban). Az objektumok megadott kétváltozós normális eloszlások valamelyikéhez tartozó realizációk, megfelelő paraméterezés esetén tehát (sok valóságos feladattal ellentétben) az osztályozás minősége vizuálisan jól követhető.

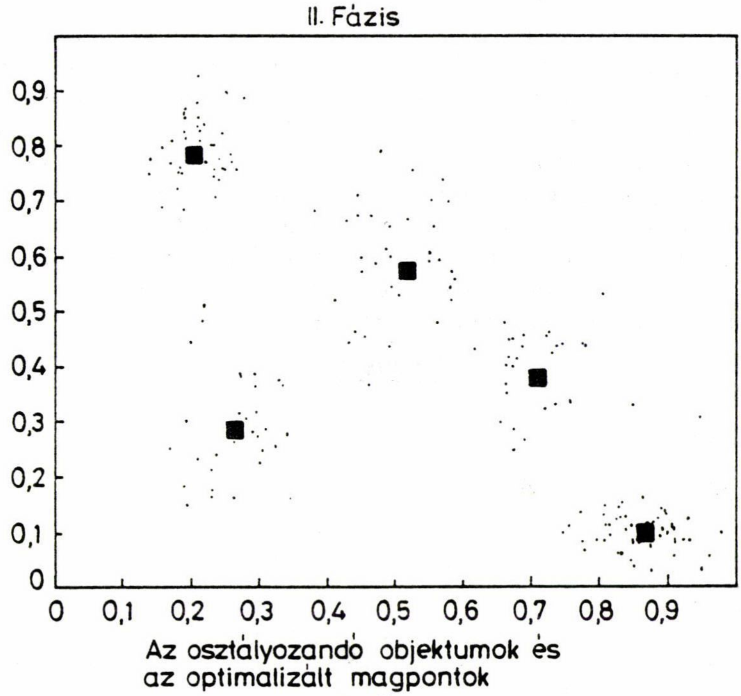
Az említett tesztfeladatban 200 objektum öt csoportra osztását kellett megvalósítani. Itt jegyezzük meg, hogy bár a clusterek számának megállapítása sem mindig triviális probléma, ezzel a kérdéssel a jelen keretek között nem foglalkoztunk (lásd pl. Rousseeuw (1984) e kérdéssel foglalkozó munkáját). A tesztfeladat viszonylagos egyszerűsége mellett eléggé nagyméretű ahhoz, hogy egzakt (kombinatorikus) megoldásának előállítása komoly numerikus nehézséget jelentsen.

A leírt osztályozási feladatot tehát a cluster-magpontok legjobb konfigurációjának keresésére vezettük vissza: ez tízváltozós globális optimalizálási problémához vezetett. (Megjegyezzük, hogy ez a viszonylag alacsony dimenziószám is már nemtriviális numerikus feladatot indukálhat.) Mivel a magpont-konfiguráció struktúrája nyilván *permutációinvariáns*, ezért a globális optimalizálási feladatnak nagyszámú ekvivalens megoldása volna. Ezt elkerülendő, *kétfázisú hierarchikus optimalizációt* hajtottunk végre: az első fázisban egy (összességében) közelítő optimális megoldást generáltunk, amelyet azután az *egyes magpontok* helyzetének (a neki megfelelő clusterhez viszonyított) pontosabb kijelölésére irányuló második fázis követett. Ez az egyszerű fogás nem csak a vázolt permutációs probléma elkerülését segítette, hanem a kiinduló $n = M \cdot m$-változós feladat nagypontosságú megoldását is (clusterenkénti m-változós) *dekompozicióval* lehetett végrehajtani.



1. ábra

Az elmondottakat illusztrálandó, tekintsük az 1. és 2. ábrát, amelyek az első, illetve a második optimalizációs fázis eredményét mutatják (a sötét négyzetek a magpontok becslését, az apró pontok pedig a véletlenszerűen generált, osztályozandó objektumokat jelölik). Ez a 200-elemű, 5-clusteres feladat az első fázisban mintegy ezer iterációs lépés végrehajtását igényelte (a tízváltozós döntési térben); a második fázisban a közelítő megoldást az egyes kétdimenziós alterekben kb. 300–300 lépésben finomítottuk. (Megjegyezzük, hogy a második fázisban a részmegoldásokat már a szóbanforgó magponthoz rendelt cluster alapján redukált intervallumon kereshettük.) Idézzük még fel, hogy az említett kereső lépések a kiválasztott (3.14) kritériumfüggvény *gradiensmentes* kiértékelését jelentették az egyes magpont-rendszerek



2. ábra

esetén: így ez az értékelés viszonylag gyorsan végrehjtható volt. (A számítógépi időigény — koprocesszorral rendelkező IBM PC/AT kompatibilis gép felhasználásával — 6 perc 55 másodperc volt az első fázisban, beleértve az 1. ábra és az eredményfájlok elkészítését; a teljes második fázis 50 másodpercet vett igénybe, ismét beleértve a 2. ábra és az eredmények rögzítését.)

Amint azt a 2. ábra mutatja, az elért megoldás jól közelíti az osztályozandó halmazok centrumait (a megfelelő várható érték vektorokat). Ennek kapcsán érdemes megjegyezni, hogy „jólkondicionált" osztályozási feladatok eredménye várhatóan nem túlságosan érzékeny a magpontok kisebb ingadozásaira: így azok *globálisan legjobb* helyzetének közelítése gyakorlati szempontból elfogadható.

A bemutatott numerikus példa tanusága szerint viszonylag nagyobb feladatok is megoldhatók (még személyi számítógépen is) a javasolt optimalizálási módszertípus alkalmazásával. Hangsúlyozzuk, hogy a PC-k használata elsősorban tárolókapacitás szempontjából (tehát nem a műveletigény következtében) jelent korlátozást — a feladat dimenziójának függvényében. Az említett numerikus nehézségek pl. a feladat dekompozíciója, ismételt futtatások és más heurisztikus eljárások segítségével bizonyos mértékig feloldhatók.

Végezetül megemlítjük a globális optimalizálási megközelítés egy további előnyét, amely annak *flexibilitásából* adódik. Amint azt már korábban említettük, számos *nem-ekvivalens* módon kifejezhető az „optimális osztályozás" kritériuma: mindezek a modellformák egy-egy megfelelő kombinatorikus módszer alkalmazását igényelnék. Ezzel ellentétben, az itt javasolt *folytonos* szemléletű megoldási stratégia lényegét tekintve nem változik, csak a célfüggvény problémafüggő megválasztására van szükség. Példaként tekintsük a kiszolgálórendszer telepítési feladatnak a „vásárlók" igényeivel súlyozott változatát (*weighted facility location problem*): ennek megoldására a javasolt eljárás lényegileg változatlan formában alkalmazható, tetszőleges távolságfogalom bevezetése mellett. Hasonló elvek alapján, számos egyéb „konfiguráció-tervezési feladat" (amely néhány objektum pozíciójának kijelölésétől tetszőlegesen bonyolult, de algoritmikusan leírható módon függ) fogalmazható és oldható meg globális optimalizálási feladatként (lásd pl. PINTÉR és COOKE (1987) tanulmányát, amelyben „szakértői vélemények" összegzésére irányuló különböző modelleket ismertetünk; az itt szereplő modellek egyike „nem-kivánatos" (veszélyes, zajos stb.) anyagok, berendezések helyének kijelölésére is alkalmazható.

A dolgozatunkban áttekintett elméleti eredmények arra utalnak, hogy bár a globális optimalizálási problémakör inherens nehézségei nem oldhatók fel, léteznek matematikailag korrekt megoldási módszerosztályok: itt ezek egyikét tárgyaltuk. Az illusztratív numerikus eredmények pedig azt mutatják, hogy legalábbis az alacsony dimenziójú, de valóban többextrémumú feladatok megoldása sikeresen hajtható végre az említett szabatos algoritmusok felhasználásával.

IRODALOM

[1] ALLGOWER, E.L. and GEORG, K., "Predictor-corrector and simplicial methods for approximating fixed points and zero points of nonlinear mappings", in Mathematical Programming, The State of the Art. Bachem, A., Grötschel, M. and Korts, B. eds. (Springer–Verlag, Berlin Heidelberg New York, 1983), pp. 15–56.

[2] ARCHETTI, F. and BETRO, B., "A probabilistic algorithm for global optimization", *Calcolo* **16** (1979), 335–343.

[3] ARCHETTI, F. and BETRO, B., "Stochastic models and global optimization", *Bolletino della Unione Matematica Italiana* **5** *17-A* (1980), 295–301.

[4] BASSO, P., "Iterative methods for the localization of the global maximum", *SIAM Journal on Numerical Analysis* **19** (1982), 781–792.

[5] BERGER, J. O., *Statistical Decision Theory* (Springer-Verlag, Berlin Heidelberg New York, 1980).

[6] BOENDER, C.G.E., "The Generalized Multinomial Distribution", A Bayesian Analysis and Applications Ph.D. Thesis, *Erasmus University, Rotterdam* (1984).

[7] BRUCKER, P., "On the complexity of clustering problems", in Optimization and Operations Research, Lecture Notes in Economics and Mathematical Systems **157** Beckmann, M. and Künzi, H.P., eds. (Springer–Verlag, Berlin Heidelberg New York, 1978), pp. 45–54.

[8] CHEW, S.H. and ZHENG, Q., *Integral global optimization. Lecture Notes in Economics and Mathematical Systems* **298** (Springer–Verlag, Berlin Heidelberg New York, 1988).

[9] COOKE, R. and PINTÉR, J., *Optimization in risk management* (Civil Engineering Systems, 1989), (megjelenés alatt).

[10] CSENDES, T., "Nonlinear parameter estimation by global optimization — efficiency and reliability", *Acta Cybernetica* **8** (1988), 361–370.

[11] DANILIN, YU.M. and PIYAVSKII, S.A., "An algorithm for finding the absolute minimum", in Theory of Optimal Decisions, Vol. 2. (Institute of Cybernetics of the Ukrainian Academy of Sciences, Kiev, 1967), pp. 25–37, (In Russian).

[12] DENNIS, J.E. and SCHNABEL, R.B., *Numerical Methods for Nonlinear Equations and Unconstrained Optimization* (Prentice-Hall, Englewood Cliffs, New Jersey, 1983).

[13] DEVROYE, L.P., "Progressive global random search of continuous functions", *Mathematical Programming* **15** (1978), 330–342.

[14] DIXON, L.C.W. and SZEGŐ G.P., *eds. Towards Global Optimisation*, Vols 1–2 (North-Holland, Amsterdam, 1975, 1978).

[15] FEDOROV, V.V., *ed. Problems of Cybernetics. Models and Methods in Global Optimization* (USSR Academy of Sciences, Moscow, 1985), (In Russian).

[16] FODOR, J. and PINTÉR J., "Extreme order statistics applied for optimum estimation in "hard" MP problems", in Transactions of the Tenth Prague Conference on Information Theory, Statistical Decision Functions, Random Processes (Academia, Prague, 1988), pp. 321–328.

[17] FORGÓ, F., *Nemkonvex és diszkrét programozás* (Közgazdasági és Jogi Könyvkiadó, Budapest, 1978).

[18] FORSTER, W., *ed. Numerical Solution of Highly Nonlinear Problems* (North-Holland, Amsterdam, 1980).

[19] FUJII, V., ICHIDA, K. and OZASA, M., "Maximization of multivariate functions using interval analysis", in Interval Mathematics 1985. Nickel, K.L.E., ed. (Springer–Verlag, Berlin Heidelberg New York, 1986), pp. 37–56.

[20] GALAMBOS, J., *The asymptotic Theory of Extreme Order Statistics* (Wiley, New York, 1978).

[21] GALPERIN, E.A., "Precision, complexity and computational schemes of the cubic algorithm", *Journal of Optimization Theory and Applications* **57** (1988), 223–238.

[22] GARCIA, C.B. and ZANGWILL, W.I., *Pathways to Solutions, Fixed Points and Equilibria* (Prentice-Hall, Englewood Cliffs, New Jersey, 1981).

[23] GE RENPU, *A filled function method for finding a global minimizer of a function of several variables*, Paper presented at the Dundee Biennial Conference on Numerical Analysis (1983).

[24] GORDON, A.D., *Classification: Methods for the Exploratory Analysis of Multivariate Data* (Chapman and Hall, New York, 1981).

[25] GROCH, A., VIDIGAL, L.M. and DIRECTOR, S.W., "A new global optimization method for electronic circuit design", *IEEE Transactions on Circuits and Systems* **CAS–32** (1985), 160–170.

[26] GUMBEL, E.J., *Statistics of Extremes* (Columbia University Press, New York, 1958).

[27] DE HAAN, L., "Estimation of the minimum of a function using order statistics", *Journal of the American Statistical Association* **76** (1981), 467–469.

[28] HANSEN, E.R., "Global optimization using interval analysis — the one-dimensional case", *Journal of Optimization Theory and Applications* **29** (1979), 331–344.

[29] HANSEN, E.R., "Global optimization using interval analysis — the multidimensional case", *Numerische Mathematik* **34** (1980), 247–270.

[30] HANSEN, P. and JAUMARD, B., "Minimum sum of diameters clustering", *Journal of Classification* **4** (1987), 215–226.

[31] HANSEN, P., JAUMARD, B. and LU, S.H., *An analytical approach to global optimization*, Research Report G–89–01, HEC-GERAD (University of Montreal, 1989).

[32] HARTIGAN, J.A., *Clustering Algorithms* (John Wiley and Sons, New York, 1975).

[33] HORST, R., "On the global minimization of concave functions", Introduction and survey, *Operations Research Spektrum* **6** (1984), 195–205.

[34] HORST, R., "A general class of branch-and-bound methods in global optimization with some new approaches for concave minimization", *Journal of Optimization Theory and Applications* **51** (1986), 271–291.

[35] HORST, R., "Deterministic global optimization with partition sets whose feasibility is not known: application to concave minimization, reverse convex constraints, d.c. programming, and Lipschitzian optimization", *Journal of Optimization Theory and Applications* **58** (1988a), 11–37.

[36] HORST, R., *Deterministic methods in constrained global optimization: some recent advances and new fields of application*, Research Report 19, Department of Mathematics IV (University of Trier, 1988b).

[37] HORST, R. and THOAI, NG.V., "Branch-and-bound methods for solving systems of Lipschitzian equations and inequalities", *Journal of Optimization Theory and Applications* **58** (1988), 139–146.

[38] HORST, R. and TUY, H., "On the convergence of global methods in multiextremal optimization", *Journal of Optimization Theory and Applications* **54** (1987), 253–271.

[39] KAUFMAN, L. and ROUSSEEUW, P., *Clustering by means of methods*, Research Report 87–03 (Department of Mathematics and Informatics, Delft University of Technology, 1987).

[40] KUSHNER, H.J., "A new method of locating the maximum of an arbitrary multipeak curve in the presence of noise", Transactions of ASME, Series D, Journal of Basic Engineering.

[41] LEVY, A.V. and MONTALVO, A., "The tunneling algorithm for the global optimization of functions", *SIAM Journal on Scientific and Statistical Computing* 6 (1985), 15–29.

[42] MEEWELLA, C.C. and MAYNE, D.Q., "An algorithm for global optimization of Lipschitz-functions", *Journal of Optimization Theory and Applications* 57 (1988), 307–323.

[43] MLADINEO, R.H., "An algorithm for finding the global maximum of a multimodal, multivariate function", *Mathematical Programming* 34 (1986), 188–200.

[44] MOCKUS, J., TIESIS, V. and ZILINSKAS, A., "The application of Bayesian methods for seeking the extremum", in Dixon and Szegő, eds. (1978), pp. 117–129.

[45] MULVEY, J. and COWDER, H., "Cluster analysis: An application of Lagrangian relaxation", *Management Science* 25 (1979), 329–340.

[46] MURTY, K.G. and KABADI, S.N., "Some NP-complete problems in quadratic and nonlinear programming", *Mathematical Programming* 39 (1987), 117–130.

[47] ORTEGA, J. and RHEINBOLDT, W.C., *Iterative Solution of Nonlinear Equations in Several Variables* (Academic Press, New York, 1970).

[48] PAPALAMBROS, P.Y. and WILDE, D.J., *Principles of Optimal Design: Modeling and Computation* (Cambridge University Press, Cambridge, 1988).

[49] PARDALOS, P.M. and ROSEN J.B., "Constrained Global Optimization: Algorithms and Applications", *Lecture Notes in Computer Science* 268 (Springer-Verlag, Berlin Heidelberg, 1987).

[50] PINTÉR, J., *A unified approach to globally convergent one-dimensional optimization algorithms*, Research Report IAMI 83–5 (Institute of Applied Mathematics and Informatics CNR, Milano, 1983).

[51] PINTÉR, J., "Convergence properties of stochastic optimization procedures", *Optimization* 15 (1984), 405–427.

[52] PINTÉR, J., "Globally convergent methods for n-dimensional multiextremal optimization", *Optimization* 17 (1986a), 187–202.

[53] PINTÉR, J., "Extended univariate algorithms for n-dimensional global optimization", *Computing* 36 (1986b), 91–103.

[54] PINTÉR, J., "Global optimization on convex sets", *Operations Research Spektrum* 8 (1986c), 197–202.

[55] PINTÉR, J., *Convergence qualification of partition algorithms in global optimization*, Research Report 87–61 (Department of Mathematics and Informatics, Delft University of Technology, 1987).

[56] PINTÉR, J., "Branch-and-bound methods for solving global optimization problems with Lipschitzian structure", *Optimization* 19 (1988a), 101–110.

[57] PINTÉR, J., *Solving nonlinear equation systems via global partition and search: Some experimental results*, Research Report, VITUKI (1988b), (Submitted for publication).

[58] PINTÉR, J. and COOKE, R., *Combining expert opinions: An optimization approach*, Research Report 87–84 (Department of Mathematics and Informatics, Delft University of Technology, 1987).

[59] PINTÉR, J. and PESTI, G., *Set partition by globally optimized cluster seed points*, Research Report, VITUKI (1988), (Submitted for publication).

[60] PINTÉR, J., SZABÓ, J. and SOMLYÓDY L., "Multiextremal optimization for calibrating water resources models", *Environmental Software* 1 (1986), 98–105.

[61] RAO, M.R., "Cluster analysis and mathematical programming", *Journal of the American Statistical Association* 66 (1971), 622–626.

[62] RAO, G.S., TYAGI, R.S. and MISHRA R.K., "Calculation of the minimum energy conformation of biomolecules using a global optimization technique", I. Methodology and application to a molecular fragment (normal pentane)., *Journal of Theoretical Biology* 90 (1981), 377–389.

[63] RATSCHEK, H., "Inclusion functions and global optimization", *Mathematical Programming* 33 (1985), 300–317.

[64] RATSCHEK, H.R. and ROKNE, J., *New Computer Methods for Global Optimization* (Ellis Horwood, New York, 1988).

[65] RINNOOY KAN, A.H.G. and TIMMER, G.T., "Stochastic global optimization methods. Part I: Clustering methods", *Mathematical Programming* 39 (1987a), 27–56.

[66] RINNOOY KAN, A.H.G. and TIMMER, S.T., "Stochastic global optimization methods. Part II: Multi-level methods", *Mathematical Programming* **39** (1987b), 57–78.
[67] RINNOOY KAN, A.H.G. and TIMMER, G.T., "Global Optimization: A Survey. To appear", in Handbooks of Operations Research, Vol. 1. (1988).
[68] ROBINSON, S.M. ED., *Analysis and Computation of Fixed Points* (Academic Press, New York, 1980).
[69] ROUSSEEUW, P., *Silhouettes: A graphical aid to the interpretation and validation of cluster analysis*, Research Report 84–41 (Department of Mathematics and Informatics, Delft University of Technology, 1984).
[70] SCHOEN, F., "On a sequential search strategy in global optimization problems", *Calcolo* **19** (1982), 321–334.
[71] SHEN, Z. and ZHU, Y., "An interval version of Shubert's iterative method for the localization of the global maximum", *Computing* **38** (1987), 275–280.
[72] SHUBERT, B.O., "A sequential method seeking the global maximum of a function", *SIAM Journal on Numerical Analysis* **9** (1972), 379–388.
[73] SOLIS, F.J. and WETS, R.J.-B., "Minimization by random search techniques", *Mathematics of Operations Research* **6** (1981), 19–30.
[74] SOMLYÓDY, L. and VAN STRATEN, G., EDS., *Modeling and Managing Shallow Lake Eutrophication* (Springer-Verlag, Berlin Heidelberg New York, 1986).
[75] STRONGIN, R., *Numerical Methods for Multiextremal Problems* (Nauka, Moscow, 1978), (In Russian).
[76] TÕRN, A.A., "A search clustering approach to global optimization", in Dixon and Szegő, eds. (1978), pp. 49–62.
[77] VYSOTSKAYA, I.N. and STRONGIN, R., "A method for solving nonlinear equations using prior probabilistic root estimates", *Journal of Computational Mathematics and Mathematical Physics* **23** (1983), 3–12, (In Russian).
[78] WELCH, J.W., "Algorithmic complexity: Three NP-hard problems in computational statistics", *Journal of Statistical Computation and Simulation* **15** (1983), 17–25.
[79] WOMERSLEY, R.S., "Censored discrete linear ℓ_1-approximation", *SIAM Journal on Scientific and Statistical Computing* **7** (1986), 105–122.
[80] WOOD, G.R., *Multidimensional bisection and globál minimization*, Research Report (University of Canterbury, New Zealand, 1985).
[81] ZILINSKAS, A., "Two algorithms for one-dimensional multimodal minimization", *Optimization* **12** (1981), 53–63.
[82] ZILINSKAS, A., "Axiomatic approach to statistical models and their use in multimodel optimization theory", *Mathematical Programming* **22** (1982), 104–116.
[83] ZILINSKAS, A., "Global Optimization.", *Axiomatics of Statistical Models, Algorithms and Their Applications* (Mokslas, Vilnius, 1986).

(Beérkezett: 1989. június 19.)

PINTÉR JÁNOS
KÖZLEKEDÉSTUDOMÁNYI INTÉZET
1119 BUDAPEST, THÁN KÁROLY U. 3–5.

ADAPTIVE PARTITION STRATEGIES FOR SOLVING MULTIEXTREMAL OPTIMIZATION PROBLEMS

J. PINTÉR

Following a brief introduction to the subject of multiextremal (global) optimization, a general class of direct (gradient-free) adaptive partition and search type solution strategies is considered. In an outline of the underlying main theoretical results, a prototype algorithmic framework is presented. This unifying approach subsumes e.g. a number of known one-dimensional global optimization methods; moreover, it has natural multivariate extensions and is implementable (in „tailored" realizations) for handling continuous or Lipschitz-continuous objective functions defined on rectangular, convex, star-shaped or Lipschitzian nonconvex feasible sets. As some prospective applications of multiextremal optimization, three important problem-types are investigated: the general solution of nonlinear equation systems, nonlinear model calibration and data classification (cluster analysis): numerical results are also presented.

Alkalmazott Matematikai Lapok 15 (1990–91) 353–364

NEMLINEÁRIS EGYENLETRENDSZEREK MEGOLDÁSÁRA SZOLGÁLÓ ABS-MÓDSZEREK NUMERIKUS VIZSGÁLATA

JENEY ANDRÁS

Miskolc

Az ABS-módszerek egy lehetséges diszkretizációját [7]-ben vezettük be. Jelen dolgozatban 5 különböző paraméter választással adódó konkrét diszkrét algoritmust vizsgáltunk numerikusan. A vizsgálatba bevontuk a Brown- és a Brent módszert, valamint 4 folytonos ABS-algoritmust is. Mind a 11 módszerrel megkíséreltük a [11]-ben közölt 100 egyenletrendszer megoldását, különböző kezdeti értékek mellett. Mintegy 3300 számítógépi futás eredményeit több szempont szerinti csoportosításban diagramokon ábrázoltuk. Ezek alapján megállapítható, hogy a bevezetett diszkrét módszerek általában ajánlhatók a nemlineáris egyenletrendszerek megoldására.

1. Bevezetés

Az eredetileg lineáris egyenletrendszerekre javasolt ABS-*módszereket* ABAFFY, GALÁNTAI és SPEDICATO [5], valamint ABAFFY és GALÁNTAI [4] kiterjesztették az

$$(1) \qquad F(x) = 0; \quad x \in \mathbb{R}^n; \quad F : D \subset \mathbb{R}^n \to \mathbb{R}^n$$

(vagy skalárisan: $f_i(x) = 0$; $x \in \mathbb{R}^n$; $f_i : D \subset \mathbb{R}^n \to \mathbb{R}$; $i = 1, 2, \ldots, n$) alakú, n ismeretlenes nemlineáris egyenletrendszerek megoldására is.

A paraméterek különböző választásával kapott algoritmusokat numerikusan tesztelte ABAFFY, BROYDEN és SPEDICATO [3], valamint SPEDICATO és BODON [10].

Említett dolgozatokhoz csatlakozva, [7]-ben az ABS-*módszerek* egy hatékony diszkrét részosztályát adtuk meg. Ugyanott rámutattunk ezen algoritmusok, valamint a SCHMIDT és HOYER által [8]-ban általánosított *Brown-Brent-módszerek* kapcsolatára.

Ahhoz, hogy a módszerekről átfogó véleményt alkothassunk, a konvergencia-viszonyok, valamint a hatékonysági mutató elméleti tisztázása mellett numerikus kísérletek is szükségesek. A dolgozatban 5 ismert és 6 általunk bevezetett módszer numerikus vizsgálatát végeztük el. Valamennyi módszerrel, számítógéppel, 100 egyenletrendszert oldottunk meg, egyenként 3-3 kezdő vektor mellett. A mintegy 3300 futtatás alapján végzett összehasonlítások eredményeit diagramokban dolgoztuk fel.

2. A numerikusan vizsgált algoritmusok

A közölt algoritmusokban az e_j , ill. I a megfelelő méretű, mindenkori j-edik egységvektor, ill. egységmátrix. A vektorok skalár komponenseit zárójeles felső indexelés mutatja.

A bázis ABS-*algoritmus* [5] alapján a következő (az $a_k^T(y_k)$ a *Jacobi-mátrix* k-adik sorvektora):

I. Algoritmus (bázis ABS)

Az x^* megoldáshoz elég közeli $x_0 \in \mathbf{R}^n$-ből kiindulva határozzuk meg az $\mathbf{R}^n$-beli $x_1, x_2, \ldots$ sorozat elemeit az alábbiak szerint ($i = 0, 1, \ldots$):

1.lépés. Legyen $k := 1$; $y_1 := x_i$; $H_1 := I$

2.lépés. Válasszunk egy z_k paraméter vektort és számítsuk a p_k vektort:

$$p_k := H_k^T z_k$$

A z_k paramétert úgy válasszuk, hogy $p_k^T a_k(y_k) \neq 0$ teljesüljön.

3.lépés. Legyen $y_{k+1} := y_k - \dfrac{f_k(y_k)p_k}{p_k^T a_k(y_k)}$

4.lépés. Válasszuk meg a w_k paramétert úgy, hogy $w_k^T H_k a_k(y_k) \neq 0$ teljesüljön, majd ezzel legyen

$$H_{k+1} := H_k - \frac{H_k a_k(y_k) w_k^T H_k}{w_k^T H_k a_k(y_k)}$$

5.lépés. Ha $k < n$, akkor legyen $k := k + 1$ és ismétlés a 2.lépéstől, egyébként legyen $x_{i+1} := y_{n+1}$; $i := i + 1$ és ismétlés az 1.lépéstől. □

(Megjegyezzük, hogy egész precíz írásmód esetén a k index mellett a főiteráció i indexét is fel kellene mindenütt tüntetni. Ezt a jobb áttekinthetőség érdekében hagytuk el itt és a később tárgyalt algoritmusnál is.)

A z_k és a w_k megválasztásától függően más és más konkrét módszer adódik. Ezek közül az alábbiakat vizsgáltuk:

a) (ABSLU) $z_k = w_k = e_k$. Implicit LU felbontás. Elméletileg ekvivalens a *folytonos Brown-módszerrel.*

b) (ABS1) $z_k = w_k = a_k(y_k)$. *Huang-algoritmus.*

c) (ABS4) $z_k = w_k = H_k a_k(y_k)$. *Módosított Huang-algoritmus.* Ez az algoritmus elméletileg egyenértékű az előzővel, mert a H_k projektor ($H_k H_k = H_k$).

d) (ABSX) $z_k^{(j)} = w_k^{(j)} = \operatorname{sgn} a_k^{(j)}(y_k)$ $(j = 1, 2, \ldots n)$. Ez a választás a b) szerinti választás olyan célzatú módosítása, hogy a képletek nevezőjébe ne kerüljön nagyságrenddel kisebb érték, ha az $a_k(y_k)$ komponensei abszolutértékben kicsik.

Az I.algoritmus-család egy részosztályának (amelyekben a z_k és w_k paraméterek első $k - 1$ komponense zérus) diszkrét változata [7] szerint:

II. Algoritmus (diszkrét ABS)

Az x^* megoldáshoz elég közeli $x_0 \in \mathbf{R}^n$-ből kiindulva határozzuk meg az $\mathbf{R}^n$-beli $x_1, x_2, \ldots$ sorozat elemeit az alábbiak szerint ($i = 0, 1, \ldots$):

1.lépés. Legyen $k := 1$; $y_1 := x_i$; $H_1 := I$; $h := h_i \neq 0$

2.lépés. Határozzuk meg az $\bar{a}_k \in \mathbf{R}^n$ vektort a következőképpen:

$$\bar{a}_k := \frac{1}{h}\begin{bmatrix} 0 \\ \vdots \\ 0 \\ f_k(y_k + hH_k^T e_k) - f_k(y_k) \\ f_k(y_k + hH_k^T e_{k+1}) - f_k(y_k) \\ \vdots \\ f_k(y_k + hH_k^T e_n) - f_k(y_k) \end{bmatrix}$$

3.lépés. Válasszunk egy z_k paraméter vektort, amelynek első $k-1$ komponense zérus és amelyre $z_k^T \bar{a}_k \neq 0$ teljesül. Legyen $p_k := H_k^T z_k$

4.lépés. Legyen $y_{k+1} := y_k - \dfrac{f_k(y_k)p_k}{z_k^T \bar{a}_k}$

5.lépés. Válasszuk meg a w_k paramétert úgy, hogy első $k-1$ komponense zérus legyen és $w_k^T \bar{a}_k \neq 0$ teljesüljön, majd ezzel legyen

$$H_{k+1} := H_k - \frac{\bar{a}_k w_k^T H_k}{w_k^T \bar{a}_k}$$

6.lépés. Ha $k < n$, akkor legyen $k := k + 1$ és ismétlés a 2.lépéstől, egyébként legyen $x_{i+1} := y_{n+1}$; $i := i + 1$ és ismétlés az 1.lépéstől. □

Ezt az algoritmust a következő paraméterek mellett vizsgáltuk:

a) (DABSLU) $z_k = w_k = e_k$. Implicit LU felbontás. Elméletileg ekvivalens a *főelem kiválasztás nélküli Brown-módszerrel.*

b) (DABS4) $w_k = z_k = \bar{a}_k$. *Módosított Huang-szerű algoritmus.* Nem tekinthetjük a folytonos módosított *Huang-eljárás* direkt diszkretizációjának, mert az $\bar{a}_k$-nak csak az utolsó $n - k + 1$ komponense approximálja a $H_k a_k(y_k)$ vektor megfelelő komponenseit.

c) (DABSINV)

$$(2) \qquad w_k = \bar{a}_k\,; \quad z_k^{(j)} = \begin{cases} -w_k^{(k)} \dfrac{\bar{a}_k^{(k)}}{w_k^T \bar{a}_k} & \text{, ha } j = k \\ -w_k^{(j)} \left(\dfrac{\bar{a}_k^{(k)}}{w_k^T \bar{a}_k} + \dfrac{1}{w_k^{(k)}} \right) & \text{, ha } j = k+1, \ldots, n \end{cases}$$

d) (DABSINV1) $w_k^{(j)} = \operatorname{sgn} \bar{a}_k^{(j)}$ $(j = 1, 2, \ldots, n)$; z_k a (2) szerint.

e) (DABSX) $w_k^{(j)} = z_k^{(j)} = \operatorname{sgn} \bar{a}_k^{(j)} \quad (j = 1, 2, \ldots, n)$

Utóbbi kettőnél a w_k, ill. z_k választása ugyanazon meggondolás alapján történik, mint a folytonos eset d) választásánál. A c) és d)-vel az ekvivalens általánosított Brown-Brent-módszernél [7] involutórius mátrix adódik.

A vizsgált eljárásokat az I. táblázatban foglaltuk össze:

I. táblázat

Sorszám	Algoritmus	Paraméterek	A használt elnevezés
1.	I.	$z_k = w_k = a_k(y_k)$	ABS1
2.	I.	$z_k = w_k = H_k a_k(y_k)$	ABS4
3.	I.	$z_k = w_k = e_k$	ABSLU
4.	I.	$z_k^{(j)} = w_k^{(j)} = \operatorname{sgn} a_k^{(j)}(y_k)$	ABSX
5.	II.	$z_k = w_k = \bar{a}_k$	DABS4
6.	II.	$z_k = w_k = e_k$	DABSLU
7.	II.	$w_k = \bar{a}_k$; z_k a (2) szerint	DABSINV
8.	II.	$w_k^{(j)} = \operatorname{sgn} \bar{a}_k^{(j)}$; z_k a (2) szerint	DABSINV1
9.	II.	$z_k^{(j)} = w_k^{(j)} = \operatorname{sgn} \bar{a}_k^{(j)}$	DABSX
10.	Brent	–	BRENT
11.	Brown	–	BROWN

Az 1–4. sorszámú algoritmusok (az „ABS"-sel kezdődő nevűek) folytonos, az 5.–9. sorszámúak (a „D"-vel kezdődőek) diszkrét ABS-*módszerek*. Referencia-eljárásként bevontuk a [8]-ban közölt *diszkrét Brown- és Brent-módszert* is. A diszkrét módszerekben a h_i lépésközt az i-edik főiterációs lépésben egységesen a

$$h_i = \max(10^{-20}, \|x_i - x_{i-1}\|)$$

szerint választottuk ($h_0 = 0.01$).

3. A numerikus vizsgálat körülményei

Az I. táblázatban felsorolt módszerek tesztelésére 100 egyenletrendszert alkalmaztunk: az Estonian Software and Computer Service Company által összeállított 3–90-es verziószámú sorozatot [11]. A sorozat tartalmazza a 14 ARGONNE tesztegyenletrendszert, amelyeket elfogadottan használnak iterációs eljárások minősítésére. Az egyenletrendszerek egy részében a *Jacobi-mátrix* szinguláris a megoldás helyén.

A számítógépi programokat Turbo Pascal (5.0 verzió) nyelven írtuk. A programokban 11-12 értékes decimális jegynek megfelelő lebegőpontos mantisszaábrázolást alkalmaztunk. A futtatásokat IBM PC 286 személyi számítógépen végeztük.

Minden egyenletrendszernél, minden módszerrel, 3-3 kezdővektor mellett kerestünk megoldást: a [11]-ben közölt standard kezdőértékekkel, azok 10-szeresével, valamint 100-szorosával. Az összehasonlíthatóság érdekében valamennyi futtatásnál ugyanazt a leállási feltételt írtuk elő, mégpedig:

$$\|F(x)\| \leq 10^{-10}, \tag{3}$$

vagy $i_v = 120$, ahol i_v az iterációszám, $\|F(x)\|$ pedig az ún. maximumnorma.

Ezt a feltételt előzetes kísérletezéssel állapítottuk meg. Azt tapasztaltuk, hogy amennyiben 120 iterációs lépésen belül nem kapjuk meg a (3) szerinti korlátot, akkor egy-két eset kivételével az nem is érhető el. Ugyanakkor a 10^{-10} olyan szigorú konstansnak bizonyult, ami kellőképpen széthúzta mind a módszerek, mind az egyenletrendszerek mezőnyét.

4. A számítógépi futtatás eredményeinek kiértékelése

A 3300 futtatás alapján a módszereket 4 szempont szerint értékeltük:

I. iterációszám,

II. futásidő,

III. a megoldott egyenletrendszerek száma,

IV. hogyan alakulnak előbbiek különböző kezdővektorok mellett.

Egy egyenletrendszert akkor tekintettünk megoldottnak, ha a (3) szerinti korlátot 120-nál kevesebb iterációs lépésben elértük.

Az I. és II. szerinti értékeléshez 2-2 mérőszámot vezettünk be. Egyik az átlagos iterációszám, ill. futásidő, a másik pedig az iterációszám, ill. futásidő alapján kialakított, ún. helyezési szám.

Ez utóbbi meghatározásánál az egyes egyenletrendszereket „pontozóbíróknak" tekintettük. Egy adott egyenletrendszer azt a módszert helyezi előbbre, amelyiknél az iterációszám, ill. a futásidő kisebb. Azonos iterációszámú, ill. (századmásodperc pontosságig) azonos futásidejű módszerek azonos helyezési számot kaptak az illető egyenletrendszertől. Egy-egy módszer végső helyezési számát az egyes egyenletrendszerektől kapott helyezési számok összegeként határoztuk meg.

A vizsgálatokat az egyenletrendszerek különböző részhalmazain végeztük. Ennek megfelelően feldolgoztuk az eredményeket úgy, hogy

- valamennyi egyenletrendszert bevontuk a feldolgozásba,
- csak az ARGONNE teszteket vontuk be,
- csak a szinguláris *Jacobi-mátrixú* egyenletrendszereket vontuk be,
- csak a reguláris *Jacobi-mátrixú* egyenletrendszereket vontuk be.

A kiértékelés során kapott táblázatokat hely hiányában nem közöljük, csupán a belőlük készült összefoglaló diagramokat. A csak a reguláris *Jacobi-mátrixú* egyenletrendszerek feldolgozása során kapott diagramok közlésétől is eltekintettünk, mert belőlük ugyanolyan következtetések vonhatók le, mint az 1. diagramból. (Ez nem meglepő, mivel a 97 megoldható egyenletrendszer nagyrészének reguláris a *Jacobi-mátrixa* a megoldás helyén.)

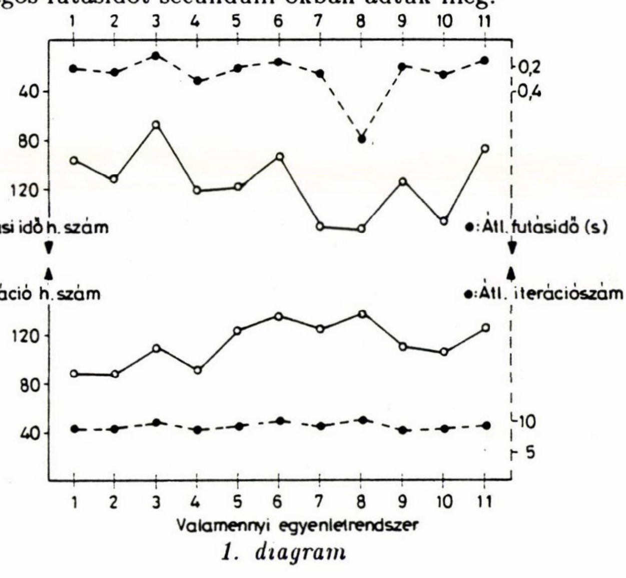
Az iterációszám és a futásidő szerinti összehasonlításnál egy-egy egyenletrendszercsaládból értelemszerűen csak azokat vettük figyelembe, amelyeket valamennyi módszer megoldott. A II. táblázatban foglaltuk össze, hogy egy-egy feladatosztályból hány egyenletrendszert oldott meg valamennyi módszer a [11]-ben megadott kezdő értékek 1-szeresével, 10-szeresével és 100-szorosával (zárójelben feltüntettük a feladatosztályokban szereplő megoldható egyenletrendszerek számát is):

II. táblázat

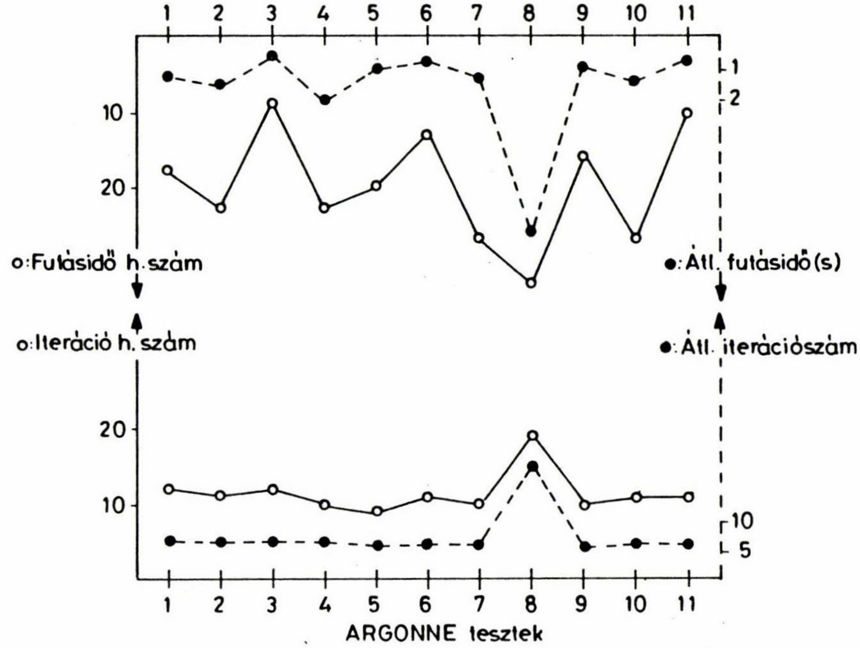
kezdőérték	1-szeres	10-szeres	100-szoros
minden egyenletrendszerből (97 db)	50	32	23
ARGONNE tesztekből (14 db)	5	4	2
szinguláris *Jacobi-m.* esetén (21 db)	11	9	8
reguláris *Jacobi-m.* esetén (76 db)	39	23	15

A módszerek iterációszám és futásidő szerinti összehasonlítására a standard kezdő adatokkal való futtatás alapján készült 1.–3. diagramok szolgálnak. Az iterációszám és a futásidő szerinti mérőszámokat együtt ábrázoltuk, az ordináta tengelyeket egymással szembe fordítva. A vízszintes tengelyen a módszerek sorszámai szerepelnek.

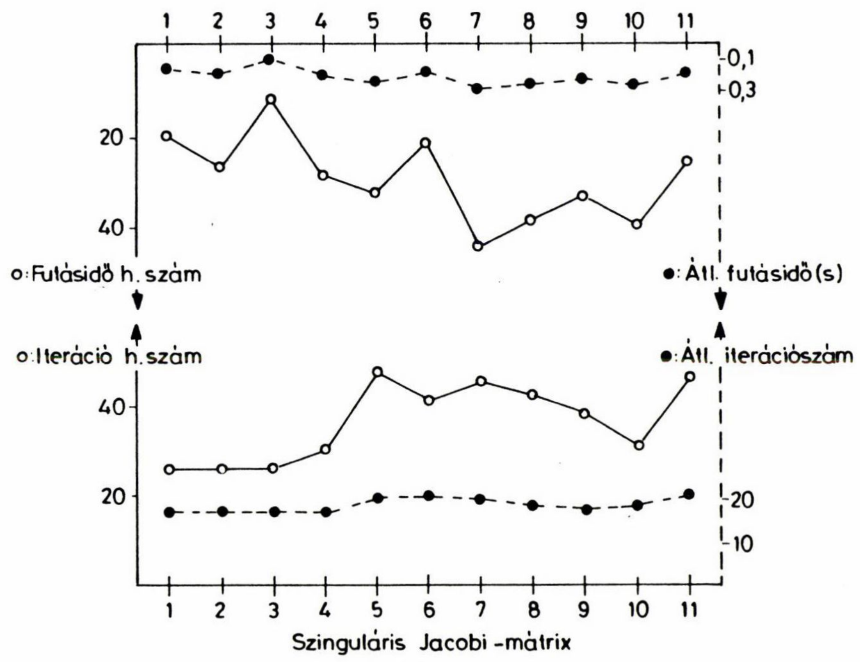
Az iterációszám és a futásidő szerinti helyezési számoknak megfelelő pontokat folytonos vonallal, az átlagos értékeknek megfelelőeket pedig szaggatott vonallal kötöttük össze. Előbbiek skálabeosztása a baloldali, utóbbiaké a jobboldali tengelyen van. Az átlagos futásidőt secundum-okban adtuk meg.



1. diagram



2. diagram



3. diagram

I. Az iterációszámok alakulása

Az átlagos iterációszámok (1.–3. diagramok alsó szaggatott vonalai) igen kis különbséget mutatnak. A különbségeket a helyezési szám felerősíti (a diagramok alsó folytonos vonalai). A tendencia szinte mindenütt ugyanaz, mint az átlagos iterációszám esetén.

Valamennyi egyenletrendszert tekintve az ABS1, ABS4 és az ABSX mutatkozik a legjobbnak, ezeket követi az ABSLU, DABSX, BRENT, majd a többiek.

Az ARGONNE tesztek alapján a helyezési számok sem mutatnak számottevő ingadozást, egyedül a DABSINV1 mutatkozik kiugróan rosszabbnak a többinél.

A szinguláris esetekben az átlagos iterációszám minden módszernél közel duplája, mint ami a valamennyi egyenletrendszer figyelembe vételével adódott. (A regulárisoknál tapasztaltakhoz képest pedig közel háromszorosa.) A folytonos módszerek általában is jobbaknak tűnnek a diszkréteknél.

II. A futásidők alakulása

Az átlagos értékeket az 1.–3. diagramok felső szaggatott vonalai, a futásidő alapján megállapított helyezési számokat a felső folytonos vonalak ábrázolják.

Abszolut számokban kis futásidőket kaptunk, de az egymáshoz viszonyított eltérések nagyobbak, mint az átlagos iterációszámok esetében. Itt is igaz az a megállapítás, hogy a helyezési számokat tekintve a tendenciák ugyanazok, mint az átlagos értékeknél, de a különbségek felerősödnek.

Valamennyi diagram egyértelműen az ABSLU és a BROWN fölényét mutatja, alig valamivel gyengébb a DABSLU.

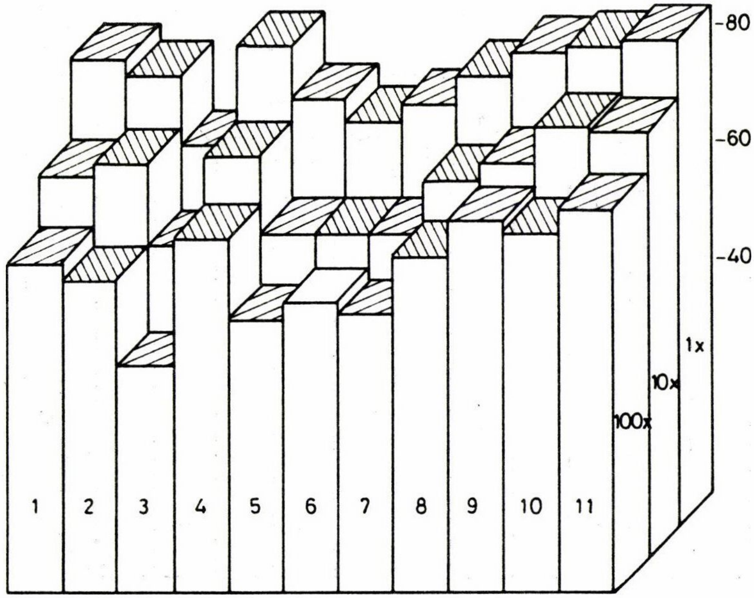
A *Huang-algoritmus* (ABS1) érthetően gyorsabb, mint a módosított (ABS4), hiszen utóbbi egy újabb projekciót hajt végre és ez – mint látható – általában nem eredményezett iterációszám csökkenést.

A következő kategóriát az ABSX, DABS4 és a DABSX alkotja, a DABSINV1 a szinguláris esetet nem nézve kirívóan lassú.

A szinguláris esetben az átlagos iterációszám növekedéssel párhuzamosan nem nő az átlagos futásidő. Ennek az a magyarázata, hogy a relatíve nagy futásidejű ARGONNE tesztek megemelik az átlagot, de külön a szingulárisok között csak egyikük szerepelt.

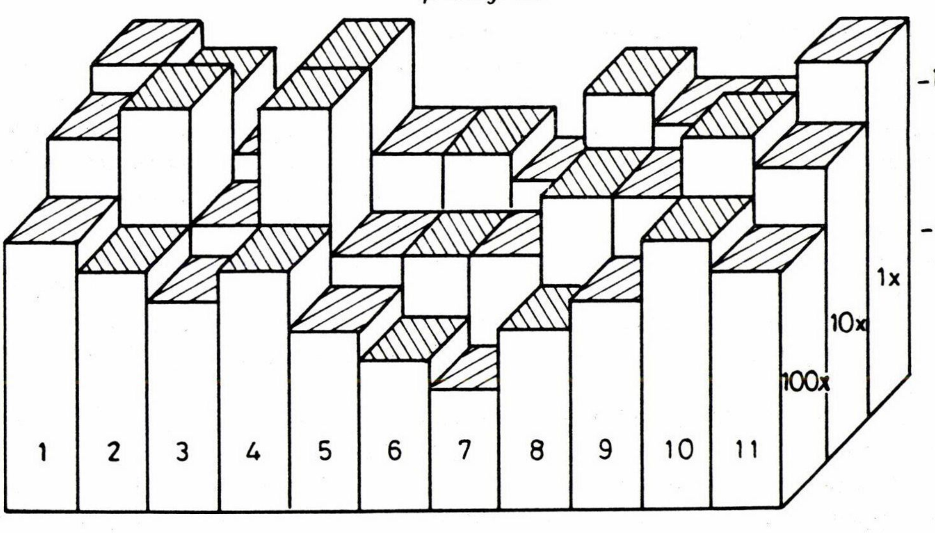
Itt jegyezzük meg, hogy a futásidő nem közvetlenül a módszerre jellemző, hanem a programra és a számítógépre. Ugyanakkor feltételezhetjük, hogy azonos programozó által azonos feltételek között készített programoknak ugyanazon gépen kapott futási ideje a módszer gyorsaságát is hűen tükrözi.

A futásidők közötti eltérések azt mutatják, hogy nem mindig elegendő csak a függvény-kiértékelések számát figyelembe venni. Ez az általunk vizsgált diszkrét, ill. folytonos módszereken belül azonos volt, időben mégis lényeges eltérések adódtak. Az egyéb aritmetikai műveletek száma nagy különbséggel éppen az ABSLU, DABSLU és a BROWN módszereknél a legkisebb, míg a legnagyobb a BRENT, DABSINV és a DABSINV1 módszereknél. Ugyanakkor pedig az iterációszámokat tekintve a DABSLU és a BROWN a legrosszabbak közé tartozik és az ABSLU is a leggyengébb a folytonosak között.



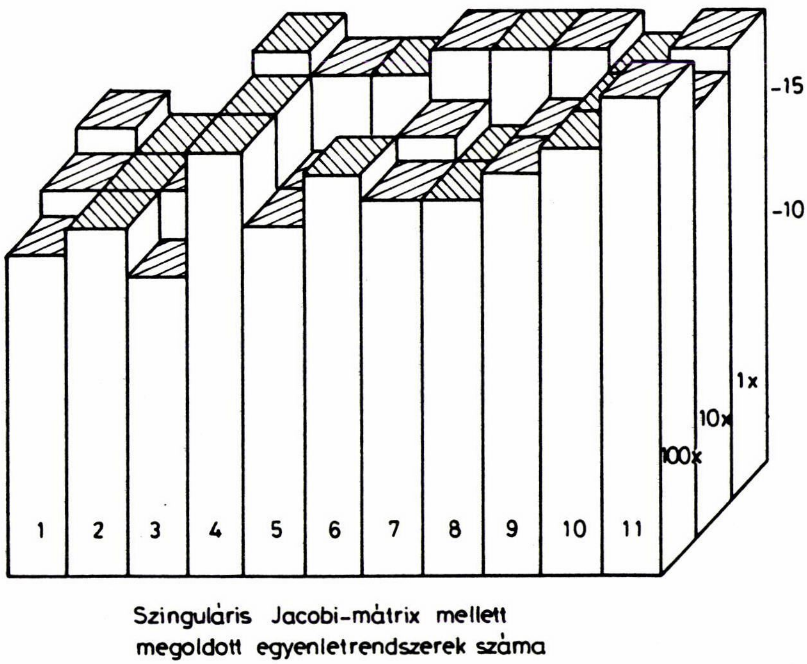
Valamennyi egyenletrendszerből
megoldott egyenletrendszerek száma

4. diagram



ARGONNE tesztekből
megoldott egyenletrendszerek száma

5. diagram



6. diagram

III. A megoldott egyenletrendszerek száma

Az egyes módszerek által megoldott egyenletrendszerek számát a 4.–6. diagramok ábrázolják. A térbeli oszlopdiagramok a standard kezdőértékek 1-szerese, 10-szerese és 100-szorosa mellett megoldott egyenletrendszerek számát mutatják. Vízszintes irányban itt is az egyes módszereknek csupán a sorszámát tüntettük fel.

Tekintettel arra, hogy az általunk vizsgált, kevés számú ismeretlent tartalmazó egyenletrendszerek esetén az iterációszámok és a futásidők jórészt kicsinyek, ezért a megoldott egyenletrendszerek számát tartjuk a legfontosabb jellemzőnek.

Megállapítható, hogy az eddig is ismert ABS1, ABS4, BRENT és BROWN módszerek mellett az általunk bevezetett folytonos ABSX és a diszkrét DABSX egyaránt a legjobbak között van. A következő kategóriába a DABS4 és a DABSINV1 sorolható.

A standard kezdőértékekkel való futás alapján egyértelműen leggyengébb a folytonos ABSLU és ennek diszkretizáltja, a DABSLU.

Lényeges szerepe van a főelem kiválasztásnak, hisz az említett leggyengébb kettő éppen a legjobbak közé tartozó BROWN (ill. a *folytonos Brown*) *módszer* főelem kiválasztás nélküli megfelelői.

Az ARGONNE tesztek alapján a BRENT és a DABSX helyett a DABSINV1 csatlakozik a legjobbakhoz, a DABSINV pedig a leggyengébb.

A szinguláris eseteket figyelembe vevő 6. diagram legfontosabb tanulsága, hogy ezeknél globális különbség jelentkezik a diszkrét módszerek javára a folytonosokkal szemben.

IV. A kezdő adatok befolyása

Az iterációszámok és a futásidő alakulását más kezdő vektorok esetén nem foglaltuk diagramba. Kissé növekszenek, ahogy távolodunk a kezdő adatokkal a megoldástól, de ez a növekedés nem számottevő. Az a tapasztalatunk, hogy valamely módszer egy adott vektorból kiindulva vagy megoldja az egyenletrendszert néhány lépésben, vagy egyáltalán nem talál megoldást.

Valamennyi egyenletrendszert figyelembe véve 10-szeres kezdeti értékek mellett a megoldott egyenletrendszerek száma a standardhoz képest átlagosan kb. 20%-kal, 100-szoros kezdő értékek esetén 30%-kal csökkent. Ebből a szempontból érdemleges különbséget nem tapasztaltunk az egyes módszerek között.

Az ARGONNE teszteket és a szinguláris esetet tekintve a változás nem ilyen szabályszerű. Pl. a *Brent-* és az ABS4-nél (*módosított Huang*) nem változott a megoldott egyenletrendszerek száma a kezdővektorok változtatásával. Szembeötlik, hogy vannak módszerek, amelyek távolabbról indítva több egyenletrendszert oldottak meg, mint közelebbről. Ez azért nem meglepő, mert a módszerek ún. attraktivitási tartománya nem szükségszerűen gömb, sőt esetleg nem is összefüggő.

V. Megjegyzések az egyenletrendszerekről

A 97 megoldható egyenletrendszerből egyik módszer sem oldott meg 8-at.

Megemlítjük, hogy azt a négyismeretlenes egyenletrendszert ([11], 114. sorszám), amelyik *Jacobi-mátrixának* rangja a megoldás helyén csak kettő, kizárólag diszkrét módszer oldotta meg, mégpedig standard kezdőérték mellett három, 10-szeres mellett egy, 100-szoros mellett pedig hat módszer.

Az egyetlen, nagyon bonyolultnak tekinthető, változtatható ismeretlenszámú egyenletrendszerrel ([11], 215. sorszám) hat ismeretlen esetén egyik módszer sem boldogult. (Két ismeretlennel még valamennyi módszer megoldotta.)

Az ARGONNE tesztek, ill. a szinguláris *Jacobi-mátrixúak* körében végzett vizsgálatokból levonható következtetések kissé eltérnek a valamennyi egyenletrendszer figyelembe vételével kaphatóakétól. Ennek az lehet az oka, hogy egyrészt kisebb a minta (pl. a 14, ill. 21 egyenletrendszer közül minden módszer 5-öt, ill. 11-et oldott meg standard kezdőértékekkel), másrészt az ARGONNE tesztek között lényegesen nagyobb arányban szerepelnek a bonyolultabb, tízismeretlenes egyenletrendszerek, szingularitás esetén pedig a konvergencia-tételek sem garantálják az adott módszerrel való megoldhatóságot.

VI. Megjegyzések a dolgozatban bevezetett módszerekről

A vizsgált 11 módszer közül 6 módszer, nevezetesen a 4.-9. sorszámúak, jelen dolgozatban szerepelnek először.

Az 1.–6. diagramok alapján megállapítható, hogy a vizsgált szempontok szerint majdnem kivétel nélkül állják a versenyt a referencia-módszerekkel, sőt esetenként a legjobbak között vannak.

A DABSINV1 meglehetősen gyenge az iterációszám és a futásidő szerint, de az általunk igen fontosnak tartott megoldó képesség tekintetében (azaz, hogy egyáltalán hányat oldott meg a vizsgált egyenletrendszerek közül), a javasolhatók között van ez az eljárás is.

Nem ajánljuk a DABSLU-t (mint ahogy az eddig is ismert, folytonos implicit LU algoritmust, az ABSLU-t sem). Helyette a BROWN-nal ekvivalens, főelem kiválasztást alkalmazó változat kidolgozását javasoljuk.

IRODALOM

[1] ABAFFY, J, "A lineáris egyenletrendszerek általános megoldásának egy direkt módszerosztálya", *Alkalmazott Matematikai Lapok* 5 (1979), 233-240.

[2] ABAFFY, J., BROYDEN, C. G., SPEDICATO, E., "A class of direct methods for linear systems", *Numerische Mathematik* 45 (1984), 361-376.

[3] ABAFFY, J., BROYDEN, C.G., SPEDICATO, E., *Numerical performance of the pseudosymmetric algorithm in the ABS class versus LU factorization with iterative refinement* (Rapporto SOFTMAT 8/83, Universita di Bergamo, 1983).

[4] ABAFFY, J., GALÁNTAI, A., *Conjugate direction methods for linear and nonlinear systems of algebraic equations. Colloquia Mathematica Societatis János Bolyai 50* (Numerical Methods, Miskolc, 1986).

[5] ABAFFY, J., GALÁNTAI, A., SPEDICATO, E., "The local convergence of ABS methods for nonlinear algebraic equations.", *Numerische Mathematik* 51 (1987), 429-439.

[6] ABAFFY, J., C.G., SPEDICATO, E., *ABS projection algorithms: Mathematical Techniques for Linear and Nonlinear Equations* (Ellis Horwood Limited, Chichester, 1989).

[7] JENEY, A., "Nemlineáris egyenletrendszerek megoldására szolgáló ABS-módszerek egy részosztályának diszkretizációja.", (Kézirat).

[8] SCHMIDT, J.W., HOYER, W., "Ein Konvergenzsatz für Verfahren vom Brown-Brent-Typ", *ZAMM* 57 (1977), 397-405.

[9] SCHWETLICK, H., *Numerische Lösung nichtlinearer Gleichungen* (VEB Deutscher Verlag der Wissenschaften, Berlin, 1979).

[10] SPEDICATO, E., BODON, E., *Numerical behaviour of the implicit QR algorithm of the ABS class for nonlinear systems. Consiglio Nazionale delle Ricerche Progretto Finalizzato Informatica SOFTMAT 90/S* (1990).

[11] *Test examples of systems of nonlinear equations, Version 3-90* (Estonian Software and Computer Service Company, Tallin, 1990).

(Beérkezett: 1991. szeptember 18.)

JENEY ANDRÁS
MISKOLCI EGYETEM MATEMATIKAI INTÉZET
3515 MISKOLC, EGYETEMVÁROS

NUMERICAL INVESTIGATION OF ABS-METHODS FOR SOLVING SYSTEMS OF NONLINEAR ALGEBRAIC EQUATIONS

A. JENEY

In paper [7] we introduced an efficient discretization of the continuous ABS methods for solving systems of nonlinear algebraic equations. In the present paper we are testing numerically five discretized as well as four continuous ABS-algorithms. The discrete Brown and Brent algorithms are also tested for comparison. The eleven methods were run on one hundred test problems ([11]) with three different initial points in each case. The numerical results are analyzed from several points of view and are also shown on diagrams. The computational results clearly show that the new discretized ABS methods can be useful for efficient solving systems of nonlinear equations.

NEMLINEÁRIS EGYENLETRENDSZEREK MEGOLDÁSÁRA SZOLGÁLÓ ABS-MÓDSZEREK EGY RÉSZOSZTÁLYÁNAK DISZKRETIZÁCIÓJA

JENEY ANDRÁS

Miskolc

A dolgozatban a nemlineáris egyenletrendszerek megoldására szolgáló ABS-módszerek egy részosztályának diszkretizálását végeztük el. Megmutattuk, hogy az így kapott módszercsalád ekvivalens a SCHMIDT és HOYER szerint általánosított Brown-Brent-módszerek egy eddig nem vizsgált részosztályával. A bevezetett diszkrét ABS-módszerek hatékonysága ezek szerint megegyezik a Brown-Brent módszerekével. Megadtuk azokat a paramétereket, amelyek alkalmazásával a konkrét Brown-, ill. Brent módszerrel ekvivalens algoritmust kapjuk.

1. Bevezetés

A dolgozatban az

$$F(x) = 0; \quad x \in \mathbb{R}^n; \quad F : D \subset \mathbb{R}^n \to \mathbb{R}^n \tag{1}$$

(vagy skalárisan: $f_i(x) = 0; \quad x \in \mathbb{R}^n; \quad f_i : D \subset \mathbb{R}^n \to \mathbb{R}; \quad i = 1, 2, \dots, n$) alakú, n ismeretlenes nemlineáris egyenletrendszerek néhány ismert megoldási módszerével és azokból levezethető új módszerekkel foglalkozunk.

ABAFFY [1], valamint ABAFFY, BROYDEN és SPEDICATO [2] javasolták módszerek egy családját lineáris egyenletrendszerekre, mely eljárások ABS-*módszerek* néven váltak ismertté. Később ABAFFY, GALÁNTAI és SPEDICATO [4], valamint ABAFFY és GALÁNTAI [3] a lineáris egyenletrendszerek megoldására szolgáló ezen direkt módszerosztályt kiterjesztették nemlineáris egyenletrendszerekre is.

SCHMIDT és HOYER [11]-ben a *Brown- és Brent-módszereket* egységesen tárgyalva, azokat tartalmazó, általános diszkrét módszercsaládot ismertet. Ezeket az eljárásokat a dolgozatban *Schmidt-Hoyer-féle általánosított Brown-Brent-típusú módszereknek* fogjuk nevezni, illetve röviden: S-H-*féle* B-B-*típusú módszereknek* .

A deriváltak numerikus közelítésével a folytonos ABS-*módszerekből* diszkrét eljárások származtathatók. Az ilyen irányú kutatások közül SPEDICATO, CHEN és DENG [14] munkáját említjük. Az általuk kidolgozott algoritmust SCD *algoritmusnak* nevezzük.

A dolgozatban a bázis ABS-*módszerek* egy lehetséges diszkretizációját vezetjük be. Az így kapott algoritmusokra DABS névvel fogunk hivatkozni. Ezen módszerek egy részosztályát kis számítási költségű, $1+\alpha$ (ahol α értéke elérheti az 1-et is) R-konvergenciarendű, hatékony eljárások alkotják, melyeket HDABS algoritmusnak nevezünk. Tisztázzuk az említett diszkrét módszerek egymáshoz való viszonyát.

Vektornormaként az euklideszi-, mátrixnormaként a spektrál-normát használjuk; az e_j, ill. I a megfelelő méretű, mindenkori j-edik egységvektort, ill. egységmátrixot jelöli.

2. Az ABS- és az S-H-féle B-B-típusú módszerek

A diszkretizációt a skálázás és blokkosítás nélküli, ún. bázis ABS-*módszernél* hajtottuk végre, ezért csak azt közöljük [4] alapján (az $a_k^T(y_k)$ a *Jacobi-mátrix* k-adik sorvektora):

I. Algoritmus (bázis ABS)

Az x^* megoldáshoz elég közeli $x_0 \in \mathbf{R}^n$-ből kiindulva határozzuk meg az $\mathbf{R}^n$-beli $x_1, x_2, \ldots$ sorozat elemeit az alábbiak szerint $(i = 0, 1, \ldots)$:

1.lépés. Legyen $k := 1$; $y_1 := x_i$; $H_1 := I$

2.lépés. Válasszunk egy z_k paraméter vektort és számítsuk a p_k vektort:

$$p_k := H_k^T z_k$$

A z_k paramétert úgy választjuk, hogy $p_k^T a_k(y_k) \neq 0$ teljesüljön.

3.lépés. Legyen $y_{k+1} := y_k - \dfrac{f_k(y_k) p_k}{p_k^T a_k(y_k)}$

4.lépés. Válasszuk meg a w_k paramétert úgy, hogy $w_k^T H_k a_k(y_k) \neq 0$ teljesüljön, majd ezzel legyen

$$H_{k+1} := H_k - \frac{H_k a_k(y_k) w_k^T H_k}{w_k^T H_k a_k(y_k)}$$

5.lépés. Ha $k < n$, akkor legyen $k := k+1$ és ismétlés a 2.lépéstől, egyébként legyen $x_{i+1} := y_{n+1}$; $i := i+1$ és ismétlés az 1.lépéstől. □

(Megjegyezzük, hogy egész precíz írásmód esetén a k index mellett a főiteráció i indexét is fel kellene mindenütt tüntetni. Ezt a jobb áttekinthetőség érdekében hagytuk el itt és a később tárgyalt algoritmusoknál is.)

Az ABS-*módszerekben* a z_k és w_k paraméterek az algoritmusban közölt megkötésektől eltekintve szabadon választhatók. Ezen paraméterek rögzítésével számos konkrét módszer adódik.

A *Schmidt-Hoyer-féle általánosított Brown-Brent-típusú módszerek* osztálya [11] alapján a következőképpen adható meg:

II. Algoritmus (S-H-*féle* B-B)

Induljunk ki az (1) egyenletrendszer x^* megoldásához elég közel lévő $x_0 \in \mathbf{R}^n$ vektorból. Az $x_1, x_2, \ldots \in \mathbf{R}^n$ sorozat elemeit egymás után állítsuk elő mindaddig, amíg a kilépési feltétel nem teljesül, a következőképpen ($i = 0, 1, \ldots$):

1.lépés. Válasszunk egy $h_i \neq 0$ diszkretizálási lépésközt és egy reguláris n x n méretű Q_i mátrixot. Legyen $h := h_i$; $k := 1$; $R_1 := Q_i$; $y_1 := x_i$

2.lépés. Határozzuk meg az $a_k \in \mathbf{R}^n$ vektort a következőképpen:

$$a_k := \frac{1}{h}\begin{bmatrix} 0 \\ \vdots \\ 0 \\ f_k(y_k + hR_ke_k) - f_k(y_k) \\ f_k(y_k + hR_ke_{k+1}) - f_k(y_k) \\ \vdots \\ f_k(y_k + hR_ke_n) - f_k(y_k) \end{bmatrix}$$

3.lépés. Állítsunk elő egy n x n méretű reguláris P_k mátrixot, amelyre

$$P_k^T a_k = s_k e_k \qquad (s_k \neq 0, \text{ skalár})$$

és amelynek alakja:

$$P_k = \left[\begin{array}{c|c} I & 0 \\ \hline 0 & \tilde{P}_k \end{array}\right] \tag{2}$$

(Az I alkotja a P_k bal felső, $(k-1)$ x $(k-1)$ méretű minorját; a jobb alsó, $n - k + 1$ sorból és oszlopból álló minor pedig a $\tilde{P}_k$; a P_k mátrix többi eleme zérus).

4.lépés. Legyen $R_{k+1} := R_k P_k$

5.lépés. Legyen $y_{k+1} := y_k - \dfrac{f_k(y_k)}{s_k} R_{k+1} e_k$

6.lépés. Ha $k < n$, akkor legyen $k := k + 1$ és ismétlés a 2.lépéstől, egyébként legyen $x_{i+1} := y_{n+1}$; $i := i + 1$ és ismétlés az 1.lépéstől. □

A Q_i és a P_k paraméterek különböző megválasztásával itt is más és más módszerek adódnak. (A [9] szerint Q_i nem játszik lényeges szerepet.)

A h_i számára [11] a következő lehetőségeket ajánlja ($0 < |h_0|$, elegendően kicsi):

$$0 < |h_{i+1}| \leq \|x_{i+1} - x_i\| \tag{3}$$

$$0 < |h_{i+1}| \leq \|x_{i+1} - x_i\|^2 \tag{4}$$

$$0 < |h_{i+1}| \leq \| F(x_{i+1}) \| \tag{5}$$

A (3) szerinti választással a R-konvergenciarend $\frac{(1+\sqrt{5})}{2}$, utóbbiaknál pedig 2. Esetenként kielégítő lehet a $0 < |h_i| \leq 10^{-\tau}(\|x_{i+1}\| + 10^{-\tau})$ választás is (itt a τ értéke legtöbbször 3). Ekkor a konvergenciarend általában csak 1.

A skalár függvényérték meghatározások száma a (3) és (4) szerinti választásokkal $\frac{n(n+3)}{2}$, az (5) szerintinél pedig $\frac{n(n+5)}{2}$.

Egységnek véve az egy vektorértékű függvény kiszámítási költségét, az *Ostrowski-féle aszimptotikus hatékonyság* [10] a következő:

$$E = \begin{cases} \dfrac{2\ln\frac{1+\sqrt{5}}{2}}{n+3} & \text{, a (3) szerint} \\ \dfrac{2\ln 2}{n+3} & \text{, a (4) szerint} \\ \dfrac{2\ln 2}{n+5} & \text{, az (5) szerint} \end{cases} \tag{6}$$

Az aritmetikai műveletek száma már természetesen függ a Q_i és a P_k paraméterek megválasztásától. A *Brown-módszer* [11] szerinti leírásában ez $\frac{1}{2}n^3 + O(n^2)$, a *Brent-módszernél* pedig $\frac{5}{2}n^3 + O(n^2)$.

3. A diszkrét ABS-algoritmus levezetése

A bázis ABS-*módszerekben* (I.algoritmus) a $H_k a_k(y_k)$ vektor egyes komponensei az f_k függvényeknek az y_k helyen vett, a H_k sorainak megfelelő irányú iránymenti deriváltjai:

$$H_k a_k(y_k) = H_k \nabla f_k(y_k) = \lim_{h\to 0} \bar{a}_k \,,$$

ahol

$$\bar{a}_k = \frac{1}{h}\begin{bmatrix} f_k(y_k + hH_k^T e_1) - f_k(y_k) \\ f_k(y_k + hH_k^T e_2) - f_k(y_k) \\ \vdots \\ f_k(y_k + hH_k^T e_n) - f_k(y_k) \end{bmatrix}$$

A $H_k a_k(y_k)$ vektor tehát approximálható $\bar{a}_k$-val. Ha az I.algoritmusban mindenütt az $\bar{a}_k$-t használjuk a $H_k a_k(y_k)$ helyett, diszkrét bázis ABS-*módszercsaládot* kapunk, melyet nevezzünk DABS-*módszercsaládnak.*

A DABS-*módszerekben* megszorítás nélküli paraméter választás esetén a függvényhivatkozások száma $n(n+1)$, azaz közel kétszer annyi, mint az S-H-*féle* B-B-*típusú módszerekben.* Az aritmetikai műveletek száma is sokkal nagyobb, mintegy $7n^3 + O(n^2)$. A később tárgyalt kapcsolat alapján feltehetjük, hogy a konvergenciarend nem magasabb, így ezen eljárások hatékonysága általában jóval kedvezőtlenebb, mint a referencia-módszereknek tekinthető *Brown-Brent-módszereké.* Speciális paraméter választással azonban ugyanolyan hatékonyságú algoritmusok adódnak.

Tekintsük a bázis ABS-*módszerek* közül azokat, amelyeknél a z_k és a w_k paramétereket úgy választjuk, hogy mindkettő első $k-1$ komponense zérus:

$$(7)\qquad \begin{aligned} z_k^T &= [\,0^T \mid \tilde{z}_k^T\,] = [0,\dots,0,z_k^{(k)},\dots,z_k^{(n)}];\\ w_k^T &= [\,0^T \mid \tilde{w}_k^T\,] = [0,\dots,0,w_k^{(k)},\dots,w_k^{(n)}] \end{aligned}$$

$$(k = 2,3,\dots,n)$$

Particionáljuk a H_k mátrixot az első $k-1$, ill. az utolsó $n-k+1$ sorára:

$$(8)\qquad H_k = \begin{bmatrix} S_k \\ - \\ \tilde{H}_k \end{bmatrix}$$

A költségcsökkentés alapjául a következő tétel szolgál:

3.1. TÉTEL. *Ha az I.algoritmusban a paraméterek (7)-nek megfelelően vannak megválasztva, akkor az y_{k+1} vektort és a $\tilde{H}_{k+1}$ mátrixot az y_k, z_k, w_k mellett a $\tilde{H}_k$ mátrix egyértelműen meghatározza. Így az algoritmus végrehajtása során a (8) szerinti particióban szereplő S_k elemei közömbösek.* $(k = 2,\dots,n;\ i = 0,1,\dots)$

Bizonyítás. Az I.algoritmus 3.lépésében számolt y_{k+1} a p_k-n keresztül függ a H_k-tól. A (7)-et és a (8)-at figyelembe véve, a 2.lépésben

$$p_k = H_k^T z_k = [\,S_k^T \mid \tilde{H}_k^T\,] \begin{bmatrix} 0 \\ - \\ \tilde{z}_k \end{bmatrix} = \tilde{H}_k^T \tilde{z}_k$$

adódik. Az S_k-nak tehát nincs szerepe az y_{k+1} meghatározásában.

Jelölje c^T a H_{k+1} mátrix k-adik sorvektorát. Az H_{k+1} utolsó $n-k+1$ sorának számítása az algoritmus 4.lépése alapján (tekintettel (7)-re és (8)-ra):

$$\begin{bmatrix} c^T \\ - \\ \tilde{H}_{k+1} \end{bmatrix} = \tilde{H}_k - \frac{\tilde{H}_k a_k(y_k)[\,0^T \mid \tilde{w}_k^T\,] \begin{bmatrix} S_k \\ - \\ \tilde{H}_k \end{bmatrix}}{[\,0^T \mid \tilde{w}_k^T\,] \begin{bmatrix} S_k \\ - \\ \tilde{H}_k \end{bmatrix} a_k(y_k)} = \tilde{H}_k - \frac{\tilde{H}_k a_k(y_k) \tilde{w}_k^T \tilde{H}_k}{\tilde{w}_k^T \tilde{H}_k a_k(y_k)}$$

Tehát a $\tilde{H}_{k+1}$ meghatározásakor is közömbös, hogy milyen az S_k.

Minthogy y_k és H_k rekurziv definíciójában csak a megelőző indexű elemek szerepelnek, továbbá a belső iteráció kezdő adatait sem befolyásolja egyetlen S_k sem, állításunkat igazoltuk. □

A tételből kiderül, hogy az algoritmus végrehajtása során a H_k mátrix első $k-1$ sorát, illetve a $H_k a_k(y_k)$ vektor első $k-1$ komponensét nem kell kiszámítani. Így már folytonos esetben is csökkenthető az aritmetikai műveletek száma, mégpedig a legáltalánosabb esetben $9n^3 + O(n^2)$-ről $\frac{9}{2}n^3 + O(n^2)$-re.

A diszkretizált változatban magának az $\bar{a}_k$-nak nem kell az első $k-1$ komponensével törődni, ill. vehetjük azokat bárminek, célszerűen zérusnak.

A hatékony diszkrét algoritmus tehát a következő:

III. Algoritmus (HDABS)

Az x^* megoldáshoz elég közeli $x_0 \in \mathrm{R}^n$-ből kiindulva határozzuk meg az R^n-beli $x_1, x_2, \ldots$ sorozat elemeit az alábbiak szerint $(i = 0, 1, \ldots)$:

1.lépés. Legyen $k := 1$; $y_1 := x_i$; $H_1 := I$; $h := h_i \neq 0$

2.lépés. Határozzuk meg az $\bar{a}_k \in \mathrm{R}^n$ vektort a következőképpen:

$$\bar{a}_k := \frac{1}{h} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ f_k(y_k + hH_k^T e_k) - f_k(y_k) \\ f_k(y_k + hH_k^T e_{k+1}) - f_k(y_k) \\ \vdots \\ f_k(y_k + hH_k^T e_n) - f_k(y_k) \end{bmatrix}$$

3.lépés. Válasszunk egy z_k paraméter vektort, amelynek első $k-1$ komponense zérus és amelyre $z_k^T \bar{a}_k \neq 0$ teljesül. Legyen $p_k := H_k^T z_k$

4.lépés. Legyen $y_{k+1} := y_k - \dfrac{f_k(y_k) p_k}{z_k^T \bar{a}_k}$

5.lépés. Válasszuk meg a w_k paramétert úgy, hogy első $k-1$ komponense zérus legyen és $w_k^T \bar{a}_k \neq 0$ teljesüljön, majd ezzel legyen $H_{k+1} := H_k - \dfrac{\bar{a}_k w_k^T H_k}{w_k^T \bar{a}_k}$

6.lépés. Ha $k < n$, akkor legyen $k := k + 1$ és ismétlés a 2.lépéstől, egyébként legyen $x_{i+1} := y_{n+1}$; $i := i + 1$ és ismétlés az 1.lépéstől. □

Megjegyezzük, hogy az 5.lépésben a H_{k+1} a következőképpen is felírható:

$$H_{k+1} = \left[I - \frac{\bar{a}_k w_k^T}{w_k^T \bar{a}_k} \right] H_k \tag{9}$$

4. Kapcsolat az S-H-féle B-B-típusú módszerekkel

Annak érdekében, hogy a kapcsolatot lehetőleg egyszerűen megmutathassuk, fogalmazzuk át a III.algoritmust úgy, hogy formailag is minél jobban hasonlítson a II.algoritmushoz:

– alkalmazzuk az $\bar{R}_k = H_k^T$ jelölést,

– vezessük be a $\bar{P}_k$ mátrixot a (9)-beli kifejezés első tényezőjének transzponáltjára:

$$\bar{P}_k := I - \frac{w_k \bar{a}_k^T}{w_k^T \bar{a}_k} \tag{10}$$

– cseréljük fel az y_{k+1} és a $H_{k+1} = \bar{R}_{k+1}^T$ kiszámításának sorrendjét,

– végül ne az $\bar{R}_{k+1}^T = H_{k+1}$-et számoljuk, hanem az $\bar{R}_{k+1}$-et (mintha a III.algoritmusban a H_k^T-t módosítanánk H_{k+1}^T-re).

Ezekkel a változtatásokkal a III.algoritmusnak csak a 4.–5.lépései módosulnak (ill. az előző lépésekben annyi a változás, hogy az említett más jelölést alkalmazzuk):

4/a.lépés. Válasszuk meg a w_k paramétert úgy, hogy első $k-1$ komponense zérus legyen és $w_k^T \bar{a}_k \neq 0$ teljesüljön. A (10) szerinti $\bar{P}_k$-val legyen $\bar{R}_{k+1} := \bar{R}_k \bar{P}_k$

5/a.lépés. Legyen $\quad y_{k+1} := y_k - \dfrac{f_k(y_k)p_k}{z_k^T \bar{a}_k}$

Nevezzük az így kapott módszercsaládot III/a.algoritmusnak.

A III. és a III/a.algoritmusok ekvivalenciája nyilvánvaló. Amennyiben a III.algoritmus 5.lépésében a H_k és a III/a.algoritmus 4.lépésében az $\bar{R}_k$ módosítását egymásnak megfelelő műveletsorozattal végezzük, akkor numerikusan is teljes az azonosság.

Térjünk vissza a II.algoritmushoz és tekintsük annál a következő paraméter választásokat: $Q_i = I$; $P_k = \hat{P}_k$, ahol

$$\hat{P}_k e_j = \begin{cases} \left(I - \dfrac{w_k \bar{a}_k^T}{w_k^T \bar{a}_k}\right) e_j = \bar{P}_k e_j & \text{, ha } j \neq k \\ z_k & \text{, ha } j = k \end{cases} \tag{11}$$

$$(j,\, k = 1, 2, \ldots, n;\ i = 0, 1, \ldots)$$

(Azaz a $\hat{P}_k$ mátrix k-adik oszlopvektora legyen a z_k, a többi elem pedig egyezzen meg a III/a.algoritmusnál bevezetett, (10)-zel definiált $\bar{P}_k$ megfelelő elemeivel.)

A w_k és a z_k paraméterek első $k-1$ komponense zérus, továbbá $w_k^T \bar{a}_k$, $z_k^T \bar{a}_k \neq 0$. Legyen az így kapott részosztály a II/1.algoritmus.

4.1. TÉTEL. *A II/1.algoritmus* $w_k^{(k)} \neq 0$ $(k = 1, 2, \ldots, n;\ i = 0, 1, \ldots)$ *esetén a II.algoritmus által meghatározott család részosztálya, továbbá ekvivalens a III/a. és így a III.algoritmussal.*

Ez egyben azt is jelenti, hogy a III/a. (ill. III.) algoritmussal leírt módszercsalád a *Schmidt-Hoyer-féle általánosított Brown-Brent-típusú módszerek* egy részosztálya.

Bizonyítás. Az állítás első részéhez azt kell megmutatnunk, hogy $\hat{P}_k$ a II.algoritmus 3.lépésében leírt (2) alakú, reguláris és

$$\hat{P}_k^T \bar{a}_k = s_k e_k \quad (s_k \neq 0) \tag{12}$$

Mivel $\bar{a}_k$, w_k, z_k mindegyikének első $k-1$ komponense zérus, ezért $\hat{P}_k$ szerkezete (11) alapján a (2)-nek megfelelő.

$\hat{P}_k^T$ csak a k-adik sorában különbözik $\bar{P}_k^T$-tól és ez a sor éppen a z_k^T, amelyre $z_k^T \bar{a}_k = s_k \neq 0$. Ebből és a $\bar{P}_k^T \bar{a}_k = I\bar{a}_k - \frac{\bar{a}_k w_k^T}{w_k^T \bar{a}_k} \bar{a}_k = 0$-ból következik (12).

A regularitás igazolásához hivatkozhatnánk a következő fejezetre, ahol explicite meg is adjuk $\hat{P}_k$ inverzét. Annak létezése azonban bizonyítható a konkrét ismerete nélkül is.

Indirekt úton bizonyítunk. Tegyük fel, hogy $\hat{P}_k$ nem reguláris. Mivel a szerkezete (2)-nek megfelelő, ez csak úgy lehet, ha utolsó $n-k+1$ oszlopvektora lineárisan nem független. Ekkor létezik nem csupa zérus α_j, hogy

$$\sum_{j=k}^{n} \alpha_j \hat{P}_k e_j = \alpha_k z_k + \sum_{j=k+1}^{n} \alpha_j \hat{P}_k e_j = 0 \tag{13}$$

Szorozzuk meg (13)-at balról $\bar{a}_k^T$-vel. Mivel $\bar{a}_k^T z_k \neq 0$, továbbá (12)-ből $\bar{a}_k^T \hat{P}_k e_j = 0$ $(j = k+1, \dots n)$, ezért az $\bar{a}_k^T$-vel szorzott (13)-ból $\alpha_k = 0$ adódik. Ezt felhasználva, magából a (13)-ból a $\sum_{j=k+1}^{n} \alpha_j \hat{P}_k e_j = 0$ egyenletet kapjuk.

Ez utóbbit írjuk fel komponensenként, figyelemmel a $\hat{P}_k$ mátrix (11) szerinti definíciójára:

$$-\frac{w_k^{(k)}}{w_k^T \bar{a}_k} \sum_{j=k+1}^{n} \alpha_j \bar{a}_k^{(j)} = 0 \tag{14}$$

$$\alpha_l - \frac{w_k^{(l)}}{w_k^T \bar{a}_k} \sum_{j=k+1}^{n} \alpha_j \bar{a}_k^{(j)} = 0 \quad (l = k+1, \dots, n) \tag{15}$$

Ha $w_k^{(k)} \neq 0$, akkor (14) miatt $\sum_{j=k+1}^{n} \alpha_j \bar{a}_k^{(j)} = 0$. Ezt (15)-be helyettesítve, $\alpha_l = 0$ $(l = k+1, \dots, n)$ következik. Az ellentmondás miatt $\hat{P}_k$ tehát reguláris.

Az állítás második részéhez (azaz a II/1. és a III/a. algoritmusok ekvivalenciájának megmutatásához) azt kell belátni, hogy a főiteráció tetszőleges i-edik lépésében a két algoritmus által előállított $\{y_k\}$ sorozat megegyezik (azonos $y_1 = x_i$ mellett). Ezt párhuzamosan bizonyítjuk azzal, hogy R_k és $\bar{R}_k$ utolsó $n-k+1$ oszlopa azonos. (Első $k-1$ oszlopuk az a_k, ill. $\bar{a}_k$ meghatározásánál is közömbös.)

A bizonyításhoz k szerinti indukciót alkalmazunk.

$k = 1$-re a két algoritmus definíciójából következik az egyezés. Feltéve, hogy k-ra igaz az állítás $(1 \leq k < n+1)$, a 4.lépés alapján nyilvánvaló az R_{k+1}és $\bar{R}_{k+1}$ említett oszlopainak azonossága. ($\hat{P}_k$ és $\bar{P}_k$ csak a k-adik oszlopukban különböznek, így a 4.lépésben az R_{k+1} és a $\bar{R}_{k+1}$ mátrixok számításakor csak ott következhet be eltérés.)

A II/1.algoritmus 5.lépésében R_{k+1} helyett írjuk a vele egyenlő $R_k\hat{P}_k$ kifejezést, valamint helyettesítsük be az $s_k = e_k^T\hat{P}_k^T a_k$-t:

$$y_{k+1} = y_k - \frac{f_k(y_k)R_k\hat{P}_k e_k}{e_k^T\hat{P}_k^T a_k} = y_k - \frac{f_k(y_k)R_k z_k}{z_k^T a_k} = y_k - \frac{f_k(y_k)\bar{R}_k z_k}{z_k^T \bar{a}_k} = y_k - \frac{f_k(y_k)p_k}{z_k^T \bar{a}_k}$$

Itt kihasználtuk az R_k és az $\bar{R}_k$ utolsó $n-k+1$ oszlopának előbb igazolt egyenlőségét, továbbá, hogy z_k első $k-1$ komponense zérus, valamint, hogy indukciós feltevésünkből $\bar{a}_k = a_k$ is következik. □

Vizsgáljuk meg a [14]-ben, részben heurisztikusan bevezetett algoritmus és az S-H-*féle* B-B-*típusú módszerek* kapcsolatát is. Az algoritmus:

IV. Algoritmus (SCD)

Induljunk ki az (1) egyenletrendszer x^* megoldásához elég közel lévő $x_0 \in \mathbf{R}^n$ vektorból. Az $x_1, x_2, \ldots \in \mathbf{R}^n$ sorozat elemeit egymás után állítsuk elő mindaddig, amíg a kilépési feltétel nem teljesül, a következőképpen $(i = 0, 1, \ldots)$:

1.lépés. Legyen $k := 1$; $y_1 := x_i$.
Válasszunk egy n x n méretű reguláris H_1 mátrixot.

2.lépés. Válasszunk egy pozitív skalár K-t, valamint ehhez az u_k, $v_k \in \mathbf{R}^n$ vektorokat és a μ, μ_j skalár számokat $(j = k, \ldots, n)$ úgy, hogy

$$\|u_k - x_i\|, \|v_k - x_i\| \leq K\|F(x_i)\|; \quad 0 < \mu, \mu_j \leq K\|F(x_i)\| \tag{16}$$

teljesüljenek.

Határozzuk meg az $n-k+1$ komponensű a_k vektort a következőképpen:

$$a_k := \begin{bmatrix} \dfrac{f_k(v_k + \mu_k H_k e_k) - f_k(v_k)}{\mu_k} \\ \dfrac{f_k(v_k + \mu_{k+1} H_k e_{k+1}) - f_k(v_k)}{\mu_{k+1}} \\ \vdots \\ \dfrac{f_k(v_k + \mu_n H_k e_n) - f_k(v_k)}{\mu_n} \end{bmatrix}$$

3.lépés. Válasszunk egy $n-k+1$ komponensű z_k paraméter vektort úgy, hogy $z_k^T a_k \neq 0$ teljesüljön és számítsuk a p_k vektort: $p_k = H_k^T z_k$

4.lépés. Legyen $s_k := y_k - u_k$. Ha $s_k \neq 0$, akkor legyen

$$\delta_k := \frac{f_k(v_k + \mu \frac{s_k}{\|s_k\|}) - f_k(v_k)}{\mu}$$

és helyettesítsük a δ_k-t a $\delta_k\|s_k\|$-val; egyébként legyen $\delta_k := 0$.

Legyen $y_{k+1} := y_k - \dfrac{[f_k(u_k) + \delta_k]p_k}{z_k^T a_k}$

5.lépés. Válasszunk egy $(n-k) \times (n-k+1)$ méretű, teljes sorrangú V_k paraméter mátrixot úgy, hogy $V_k a_k = 0$ teljesüljön, majd ezzel legyen $H_{k+1} := V_k H_k$

6.lépés. Ha $k < n$, akkor legyen $k := k+1$ és ismétlés a 2.lépéstől, egyébként legyen $x_{i+1} := y_{n+1}$; $i := i+1$ és ismétlés az 1.lépéstől. □

Megjegyezzük, hogy a más algoritmusokkal való analógiák felismerése érdekében az eredeti, [14]-ben közölthöz képest minimális formai változtatást hajtottunk végre a leírásban: egy-két helyen más jelölés, illetve más sorrend.

Az 5.lépésben használt V_k olyan általános mátrix, amivel a módszer kilép az ABS keretei közül. A szerzők a V_k mátrix számára egy szűkebb, az ABS szerkezethez illeszkedő explicit formulát is adtak. Eszerint V_k a következő V_k' mátrix első sorának elhagyásával jön létre:

$$V_k' = I - \frac{(a_k - \sigma_k e_1)w_k^T}{w_k^T a_k}, \tag{17}$$

ahol w_k olyan paraméter vektor, amelyre $w_k^T a_k \neq 0$ teljesül, σ_k skalár.

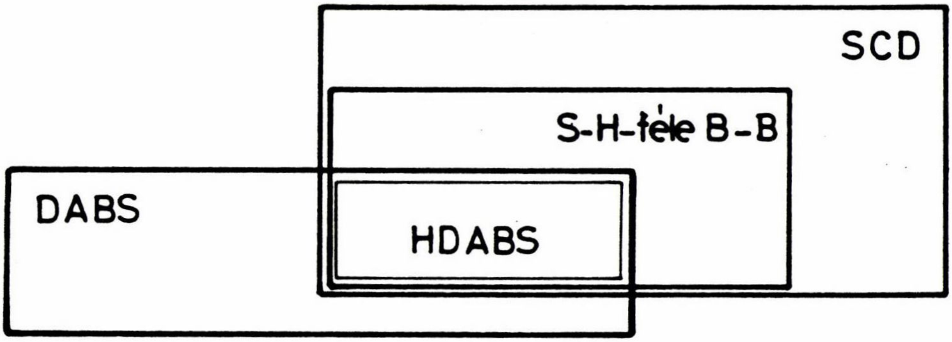
Észrevehetjük, hogy a σ_k paraméternek nincs szerepe. Ti. a (17) szerinti V_k' mátrixnak csak az első sora függ a σ_k-tól, a V_k pedig éppen ezen sor törlésével keletkezik. Ez a sor a más algoritmussal való ekvivalencia kimutatásánál érdekes, akkor, ha a z_k^T-vel azonosra választjuk.

Tekintsük a IV.algoritmusban az $u_k = v_k = y_k$ választást és legyen μ_j egy főiteráción belül állandó. Engedjünk meg μ_j számára negatív értékeket is. (Ezt valószínűleg az eredeti, IV.algoritmusban is megtehetjük, csupán az abszolut értékére kell a (16) szerinti korlátozást adni.)

A 4.1.tétel bizonyításához hasonló módon látható, hogy így az algoritmus a fent említett u_k és v_k-val nem más, mint maga a *Schmidt-Hoyer-féle általánosított Brown-Brent-típusú módszerek családja.* (A z_k-val – mint első oszloppal – bővített V_k^T-nek a $\tilde{P}_k$ felel meg. A két algoritmus leírásában a dimenziók eltérésének nincs jelentősége, hisz a konkrét végrehajtás során az S-H-*féle* B-B-*típusú módszereknél* is csak a mindenkori jobb alsó, $n-k+1$ sorból és oszlopból álló minorokkal számolunk.)

A IV.algoritmus 2.lépésében az u_k, v_k paramétereket nem vehetjük fel távol az x_i-től, így az y_k-tól sem, és a μ_j sem ingadozhat nagyon, ezért az általános, IV.algoritmus az S-H-*féle* B-B-*típusú módszerek* perturbált változatának tekinthető.

A tárgyalt diszkrét algoritmusok kapcsolatát a következő ábrán foglaltuk össze (az SCD-ben μ_j-re negatív értékeket is megengedve):



1. ábra

Mind a *Brown-*, mind a *Brent-módszer* eleme a HDABS *algoritmus-családnak* (és így a felsorolt valamennyi diszkrét módszerosztálynak). A nekik megfelelő paraméter választásokat a 6. fejezetben adjuk meg.

5. A diszkrét algoritmus konvergenciája és hatékonysága

Vezessük be a reguláris zérushely fogalmát [13] szerint.

5.1. Definíció. Az $F : D \subset \mathbf{R}^n \to \mathbf{R}^n$ függvénynek $x^* \in D$ reguláris zérushelye, ha:

(a) $F(x^*) = 0$,

(b) D-ben van az x^*-nak olyan ϱ sugarú zárt $\bar{B}(x^*, \varrho)$ környezete ($\varrho > 0$), hogy ott $F(x)$ differenciálható, továbbá $[F'(x^*)]^{-1}$ létezik,

(c) létezik $L > 0$ konstans, hogy $x, y \in \bar{B}$ esetén

$$\|F'(x) - F'(y)\| \leq L\,\|x - y\|$$

Fenti definícióból következik, hogy x^* izolált és egyszeres gyök. □

Ezek után kimondjuk a III.algoritmus konvergencia tételét.

5.2.TÉTEL. *Ha teljesülnek az alábbi feltételek:*

(f1) az x^* az $F : D \in \mathbf{R}^n \to \mathbf{R}^n$ függvény reguláris zérushelye,

(f2) $\| [F'(x^*)]^{-1} \| \leq \beta$,

(f3) léteznek az $m_1, m_2, m_3, M_1, M_2, M_3$ valós konstansok, úgy, hogy a

$$0 < m_1 \leq \left|w_k^T \bar{a}_k\right| \leq M_1 \,; \quad 0 < m_2 \leq \left|z_k^T \bar{a}_k\right| \,; \quad 0 < m_3 \leq \left|w_k^{(k)}\right| \,;$$
$$\|w_k\| \leq M_2 \,; \quad \|z_k\| \leq M_3 \,; \; (k = 1, 2, \ldots, n \,;\; i = 0, 1, \ldots)$$

egyenlőtlenségek fennállnak, akkor léteznek a γ, δ valós konstansok, amelyekkel teljesülnek a következők:

$$\left\| \hat{P}_k \right\|, \left\| [\hat{P}_k]^{-1} \right\| \leq \gamma \; ; \quad (\gamma \geq 1; \quad k = 1, 2, \ldots, n; \quad i = 0, 1, \ldots) \tag{18}$$

$$\| Q_i \|, \left\| Q_i^{-1} \right\| \leq \delta \; ; \quad (\delta \geq 1; \quad i = 0, 1, \ldots) \tag{19}$$

Továbbá, ha az $\varepsilon > 0$ valós számot úgy választjuk, hogy az

$$2c_1\varepsilon < \varrho \; ; \quad 2c_4\varepsilon < 1$$

egyenlőtlenségek teljesülnek az alábbi konstansokkal:

$$c_0 = \frac{1}{2}L\varrho + \| F'(x^*) \| ; \quad c_1 = (1 + 2c_0\beta\gamma^{2n}\delta^2)^n ; \quad c_2 = Lc_1^2(1 + \gamma^{n-1}\sqrt{n}) ;$$
$$c_3 = (n+1)c_2\sqrt{n} ; \quad c_4 = \beta(Lc_1^2 + c_3) ,$$

akkor a III.algoritmus minden olyan $x_0 \in \mathrm{R}^n$ és h_i esetén konvergens, amelyekre $x_0 \in \bar{B}(x^*, \varepsilon)$ és $0 < |h_i| \leq \varepsilon$ fennállnak és

$$\|x_{i+1} - x^*\| \leq c_4(\|x_i - x^*\| + |h_i|) \|x_i - x^*\| ; \quad (i = 0, 1, \ldots)$$

Bizonyítás. A [11]-ben a II.algoritmusra adott konvergencia tétel jelen tételünktől annyiban különbözik, hogy ott az (f3) helyett eleve a (18) és (19) teljesülését követeljük meg. (Eredetileg a (18)-ban a még nem specializált P_k szerepel, itt viszont éppen a specialitás kihasználásával, közvetlenül a paraméter vektorokra adunk feltételeket.)

A 4.1. tétel értelmében a III.algoritmus beletartozik a II.algoritmus által meghatározott osztályba, ezért itt csak azt kell megmutatnunk, hogy a feltételekből egyaránt következik (18) és (19) is.

$Q_i = Q_i^{-1} = I$ miatt (19) mindig teljesül $\delta = 1$ értékkel.

A (11)-el definiált $\hat{P}_k$ a következő alakban is írható:

$$\hat{P}_k = I - \frac{w_k \bar{a}_k^T}{w_k^T \bar{a}_k} + q_k e_k^T , \tag{20}$$

ahol

$$q_k = z_k - (e_k - \frac{w_k \bar{a}_k^T}{w_k^T \bar{a}_k} e_k) = z_k - (I - \frac{w_k \bar{a}_k^T}{w_k^T \bar{a}_k}) e_k \tag{21}$$

Becsüljük meg $\hat{P}_k$ normáját (felhasználva (21)-et):

$$\left\| \hat{P}_k \right\| = \max_{\|x\|=1} \left\| \hat{P}_k x \right\| = \max_{\|x\|=1} \left\| x - w_k \frac{\bar{a}_k^T x}{w_k^T \bar{a}_k} + q_k e_k^T x \right\| \leq$$

$$\leq 1 + \frac{\sum_{j=1}^{n} \left|\bar{a}_k^{(j)}\right|}{\left|w_k^T \bar{a}_k\right|} \|w_k\| + \|z_k\| + 1 + \frac{\left|\bar{a}_k^{(k)}\right|}{\left|w_k^T \bar{a}_k\right|} \|w_k\| \leq 2 + \frac{2M}{m_1} M_2 + M_3 ,$$

ahol M az $\bar{a}_k$ vektor komponensei abszolutérték összegének felső korlátja ($k = 1, 2, \ldots, n$; $i = 0, 1, \ldots$). Az M létezése következik az (f1) feltételből.

A (20) szerint a $\hat{P}_k = I - u\bar{a}_k^T + q_k e_k^T$ mátrix inverzét keressük, ahol $u = \dfrac{w_k}{w_k^T \bar{a}_k}$. Ilyen típusú mátrix inverze kereshető

$$[\hat{P}_k]^{-1} = I + \alpha_1 u\bar{a}_k^T + \alpha_2 u e_k^T + \alpha_3 q_k \bar{a}_k^T + \alpha_4 q_k e_k^T \tag{22}$$

alakban.

A $\hat{P}_k[\hat{P}_k]^{-1} = I$ egyenletből $\alpha_1, \alpha_2, \alpha_3, \alpha_4$-re olyan lineáris algebrai egyenletrendszert kapunk, amely $\bar{a}_k^T q_k$, $e_k^T u \neq 0$ esetén megoldható és

$$\alpha_1 = \frac{1 + e_k^T q_k}{(\bar{a}_k^T q_k)(e_k^T u)} ; \quad \alpha_2 = -\frac{1}{e_k^T u} ; \quad \alpha_3 = -\frac{1}{\bar{a}_k^T q_k} ; \quad \alpha_4 = 0$$

következik. Helyettesítsünk (22)-be és írjuk vissza az u-t. Kihasználva az egyszerűen belátható $\bar{a}_k^T q_k = z_k^T \bar{a}_k$ egyenlőséget, kapjuk:

$$[\hat{P}_k]^{-1} = I + \frac{1 + q_k^{(k)}}{(z_k^T \bar{a}_k) w_k^{(k)}} w_k \bar{a}_k^T - \frac{1}{w_k^{(k)}} w_k e_k^T - \frac{1}{z_k^T \bar{a}_k} q_k \bar{a}_k^T$$

Hasonlóan becsléssel, mint a $\hat{P}_k$ esetén, adódik, hogy $[\hat{P}_k]^{-1}$ normája is k-tól és i-től független korlát alatt marad. □

A h_i lépésközt a (3), (4), vagy (5) szerint állapíthatjuk meg. Mivel a konvergenciarend, és az egy iterációs lépésben igényelt függvényhivatkozások száma is megegyezik az S-H-*féle* B-B-*típusú módszerekével*, a HDABS *módszerek* hatékonysága is (6)-tal azonos.

Az aritmetikai műveletek száma és az igényelt tárkapacitás különböző paraméterek esetén más és más lehet.

6. A paraméter vektorok megválasztása

A számos konkrét algoritmust szolgáltató paraméterek közül itt most csak a *Brown*-, ill. a *Brent-módszert*, valamint a szimmetrikus $\bar{P}_k$ mátrixot eredményezőeket soroljuk fel.

I. $w_k = z_k = e_k$. Az implicit LU felbontás. A w_k és z_k legtermészetesebbnek látszó olyan rögzítése, amelynél első $k-1$ komponensük eltűnik. Nem nehéz megmutatni, hogy ez éppen a főelem kiválasztás nélküli *Brown-eljárás.*

II. $w_k = z_k = \bar{a}_k$. Módosított *Huang-szerű algoritmus*. Nem tekinthetjük a folytonos módosított *Huang-eljárás* (I.algoritmus, $w_k = z_k = H_k a_k(y_k)$ választással) direkt diszkretizációjának, mert az $\bar{a}_k$-nak csak az utolsó $n-k+1$ komponense approximálja a $H_k a_k(y_k)$ vektor megfelelő komponenseit.

Egyik előnye, hogy a $\bar{P}_k$ szimmetrikus, ezért az aritmetikai műveletek száma redukálható (és ha az a cél, akkor az igénybevett tárolóterület is.) További előnye, hogy a paraméterek előállításával ugyanúgy nem kell törődni, mint az I. választás során.

III. $w_k = \bar{a}_k \pm \|\bar{a}_k\| e_k$;

$$z_k = -\frac{w_k^{(k)}}{w_k^T \bar{a}_k} \bar{a}_k = -\frac{\bar{a}_k^{(k)} \pm \|\bar{a}_k\|}{w_k^T \bar{a}_k} \bar{a}_k \tag{23}$$

Egyszerű számolással kapjuk, hogy ez esetben a $\hat{P}_k$ mátrix a *Householder-transzformáció* mátrixa:

$$\hat{P}_k = I - \frac{(\bar{a}_k \pm \|\bar{a}_k\| e_k)(\bar{a}_k \pm \|\bar{a}_k\| e_k)^T}{(\bar{a}_k \pm \|\bar{a}_k\| e_k)^T \bar{a}_k} \tag{24}$$

Ezzel a paraméter választással tehát pontosan az eredeti *Brent-módszerrel* ekvivalens eljárás adódik, mind az $\{y_k\}$, mind az irányvektorok tekintetében.

A (23)-ban szereplő $\pm$ előjelek közül az $\bar{a}_k^{(k)}$ előjelével azonosat választjuk, hogy a w_k, $z_k \neq 0$ biztosan teljesüljön, ha $\bar{a}_k \neq 0$.

Megjegyezzük, hogy a *Brent-módszert* szisztematikusan is megtalálhatjuk, ha abból indulunk ki, hogy ez esetben a $\hat{P}_k$ ortogonális (azaz $\hat{P}_k \hat{P}_k^T = I$) és szimmetrikus. Szimmetrikus mátrix pontosan akkor ortogonális, ha involutórius is (azaz $\hat{P}_k \hat{P}_k = I$). Levezetve az egyidejűleg szimmetrikus és involutórius $\hat{P}_k$-ot, a (23) adódik.

Végezetül megemlítjük, hogy a 3.fejezetben bevezetett algoritmus-családot több paraméter választás mellett numerikusan is vizsgáltuk. Ennek eredményeit egy másik dolgozatban adjuk közre.

IRODALOM

[1] ABAFFY, J., "A lineáris egyenletrendszerek általános megoldásának egy direkt módszerosztálya", *Alkalmazott Matematikai Lapok* 5 (1979), 233-240.

[2] ABAFFY, J., BROYDEN, C.G., SPEDICATO, E., "A class of direct methods for linear systems", *Numerische Mathematik* 45 (1984), 361-376.

[3] ABAFFY, J., GALÁNTAI, A., "Conjugate direction methods for linear and nonlinear systems of algebraic equations.", *Colloquia Mathematica Societatis János Bolyai 50.*, Numerical Methods, Miskolc (1986).

[4] ABAFFY, J., GALÁNTAI, A., SPEDICATO, E., "The local convergence of ABS methods for nonlinear algebraic equations", *Numerische Mathematik* 51 (1987), 429-439.

[5] ABAFFY, J., C.G., SPEDICATO, E., *ABS projection algorithms: Mathematical Techniques for Linear and Nonlinear Equations* (Ellis Horwood Limited, Chichester, 1989).

[6] BRENT, R.P., "Some efficient algorithms for solving systems of nonlinear equations", *SIAM J. Numer. Anal.* 10 (1973), 327-344.

[7] BROWN, K.M., "A quadratically convergent Newton-like method based upon Gaussian-elimination", *SIAM J. Numer. Anal.* 6 (1969), 560-569.
[8] GAY, D.M., "Brown's method and some generalizations, with applications to minimization problems", *Cornell University, Comput. Sci. Techn. Rep.* (1975), 75-225.
[9] MORÉ, J.J., COSNARD, M.Y., "Numerical solution of nonlinear equations", *ACM. Trans. Math. Soft.* 5 (1979), 64-85.
[10] OSTROWSKI, A.M., *Solution of equations and systems of equations* (Academic Press, New York, 1960).
[11] SCHMIDT, J.W., HOYER, W., "Ein Konvergenzsatz für Verfahren vom Brown-Brent-Typ.", *ZAMM* 57 (1977), 397-405.
[12] SCHMIDT, J.W., HOYER, W., "Die Verfahren vom Brown-Brent-Typ bei gemischt linearen-nichtlinearen Gleichungssystemen", *ZAMM* 58 (1978), 425-428.
[13] SCHWETLICK, H., *Numerische Lösung nichtlinearer Gleichungen* (VEB Deutscher Verlag der Wissenschaften, Berlin, 1979).
[14] SPEDICATO, E., CHEN, Z., DENG, N., "A class of difference ABS-type algorithms for a nonlinear system of equations", *Quaderni del Dipartimento di Matematica, Statistica, Informatica e Applicazioni* 6 (1991).
[15] STEWART, G.W., "Conjugate direction methods for solving systems of linear equations", *Numerische Mathematik* 21 (1973), 285-297.

(Beérkezett: 1991. szeptember 18.)

JENEY ANDRÁS
MISKOLCI EGYETEM MATEMATIKAI INTÉZET
3515 MISKOLC, EGYETEMVÁROS

DISCRETIZATION OF A SUBCLASS OF THE ABS-METHODS FOR THE SOLUTION OF SYSTEMS OF NONLINEAR ALGEBRAIC EQUATIONS

A. JENEY

We discretize a subclass of the continuous ABS methods for solving systems of nonlinear algebraic equations. The discretized ABS subclass is equivalent with a subset of the Schmidt-Hoyer generalization of the discretized Brown-Brent methods. The computational efficiency of the members of the discretized ABS subclass is identical to the efficiency of the Brown-Brent methods. We also give the parameters in the ABS subclass which result the Brown and Brent methods respectively.

Alkalmazott Matematikai Lapok 15 (1990–91) 381–403

A LINEÁRIS ÉS A LOGLINEÁRIS ELOSZLÁSCSALÁD DUALITÁSI KAPCSOLATÁNAK VIZSGÁLATA GEOMETRIAI PROGRAMOZÁSSAL

LOURDES GONZALEZ

Miskolc

A dolgozat első részében a CSISZÁR [2] által vizsgált lineáris, valamint külső loglineáris feltételekkel meghatározott eloszláscsaládok dualitási összefüggését geometriai programozással tárgyaljuk. A második részben a feladat kiterjesztését adjuk. A feladat az input-output táblák kalibrálásának általánosításaként is tekinthető [1], [3], [4], [6].

1. Bevezetés

A valószínűségeloszlások egymástól való eltérésére a KULLBACK-LEIBLER [7] által bevezetett diszkrimináló információt fogjuk használni. Minthogy ezen eltérés nem szimmetrikus, célszerű a két eloszlást megkülönböztetni. RÉNYI [8] dolgozatát követve az „a priori" és „a posteriori" elnevezéseket használjuk.

1.1. *Definíció.* Legyen a $\mathrm{q} = (q_1, \ldots, q_n)$ egy „a priori" a $\mathrm{p} = (p_1, \ldots, p_n)$ egy „a posteriori" eloszlás. A két eloszlás eltérésén a következő mennyiséget értjük:

$$D(\mathrm{p}\|\mathrm{q}) := \sum_{j=1}^{n} p_j \log \frac{p_j}{q_j},$$

és ezt diszkrimináló információnak nevezzük.

A logaritmus függvény konkávitásából, azonnal adódik, hogy $D(\mathrm{p}\|\mathrm{q}) \geq 0$ és egyenlőség akkor és csak akkor ha $\mathrm{p} = \mathrm{q}$.
A $D(\mathrm{p}\|\mathrm{q})$ diszkrimináló információt az alábbi módon felbonthatjuk:

$$\sum_{j=1}^{n} p_j \log \frac{p_j}{q_j} = -\left(\sum_{j=1}^{n} p_j \log q_j + \sum_{j=1}^{n} p_j \log \frac{1}{p_j}\right).$$

Vezessük be az alábbi jelöléseket és fogalmakat:

$$H(\mathrm{p}) := \sum_{j=1}^{n} p_j \log \frac{1}{p_j},$$

a p eloszlás „entropiája", és

$$L(\mathbf{p}\|\mathbf{q}) := \sum_{j=1}^{n} p_j \log q_j,$$

a p „a posteriori" és a q „a priori" eloszlások „loglikelihood függvénye".
Így

$$D(\mathbf{p}\|\mathbf{q}) = -L(\mathbf{p}\|\mathbf{q}) - H(\mathbf{p}) \tag{1.1}$$

A diszkrimináló információ felhasználásával két alapvető optimalizálási feladatot fogalmazhatunk meg.

Legyen

$$\mathcal{S}_n := \left\{ \mathbf{a} \,\middle|\, \sum_{j=1}^{n} a_j = 1,\ a_j \geq 0 \right\},$$

az ún. n dimenziós szimplex.

1. Feladat. Minimális diszkrimináló információ feladat (MDI):

Legyen adott $\mathbf{q}^{(0)} \in \mathcal{S}_n$ eloszlás és $\mathcal{L} \in \mathcal{S}_n$ eloszláshalmaz. A feladat: keresendő azon $\mathbf{p} \in \mathcal{L}$, melyre

$$D(\mathbf{p}\|\mathbf{q}^{(0)}) \quad \text{minimális.}$$

2. Feladat. Maximum likelihood feladat (ML):

Legyen adott $\mathbf{p}^{(0)} \in \mathcal{S}_n$ eloszlás és $\mathcal{E} \in \mathcal{S}_n$ eloszláshalmaz. A feladat: keresendő azon $\mathbf{q} \in \mathcal{E}$, melyre

$$D(\mathbf{p}^{(0)}\|\mathbf{q}) \quad \text{minimális.}$$

Minthogy (1.1) alapján

$$D(\mathbf{p}^{(0)}\|\mathbf{q}) = -L(\mathbf{p}^{(0)}\|\mathbf{q}) - H(\mathbf{p}^{(0)}),$$

ezért ezen feladat ekvivalens a

$$\max_{\mathbf{q}\in\mathcal{E}} L(\mathbf{p}^{(0)}\|\mathbf{q})$$

feladattal; azaz maximum likelihood feladat.
A következőkben a két feladat dualitási kapcsolatát mutatjuk meg abban az esetben amikor az $\mathcal{L}$ eloszláshalmaz belső lineáris feltételekkel (implicit módon) van megadva és az $\mathcal{E}$ eloszláshalmaz külső loglineáris formában (explicit módon) adott.

2. Lineáris és loglineáris eloszláscsalád optimalizációs feladatának dualitási kapcsolata

Legyen **A** egy tetszőleges adott $m \times n$-es mátrix; oszlopvektorait jelöljük $\mathbf{a}_j$-vel. Az **A** mátrixon keresztül definiáljuk az $\mathcal{L}$ lineáris és az $\mathcal{E}$ loglineáris eloszláscsaládot.

Belső feltételes, lineáris eloszláscsalád:

Legyen $\mathbf{p}^{(0)}$ egy adott eloszlás. Az alábbi implicit módon, lineáris feltételekkel adott $\mathbf{p}$ eloszlások összességét $\mathcal{L}$ lináris eloszláscsaládnak nevezzük.

$$\mathbf{A}\mathbf{p} = \mathbf{A}\mathbf{p}^{(0)}. \tag{2.1}$$

Azonnal látható, hogy $\mathbf{p}^{(0)} \in \mathcal{L}$ és bármelyik $\mathbf{p} \in \mathcal{L}$ átveheti $\mathbf{p}^{(0)}$ szerepét.

Külső feltételes, loglineáris eloszláscsalád:

Legyen $\mathbf{q}^{(0)} > 0$ egy adott eloszlás. Az alábbi explicit módon, loglineáris feltételekkel adott $\mathbf{q}$ eloszlások összességét $\mathcal{E}$ loglineáris eloszláscsaládnak nevezzük:

$$\left.\begin{array}{rl} & q_j = q_j^{(0)} \exp(\mathbf{t}\mathbf{a}_j + u), \quad (j = 1, \dots, n) \\ \text{ahol} & \mathbf{t} = (t_1, \dots, t_m) \text{ tetszőleges és} \\ & u = -\log \sum_{j=1}^{n} q_j^{(0)} \exp(\mathbf{t}\mathbf{a}_j) \end{array}\right\} \tag{2.2}$$

(az u fenti megválasztása azért szükséges, hogy q_j eloszlás legyen). Azonnal látható, hogy $\mathbf{q}^{(0)} \in \mathcal{E}$ és bármelyik $\mathbf{q} \in \mathcal{E}$ átveheti $\mathbf{q}^{(0)}$ szerepét. Megjegyezzük, hogy $\mathbf{q}^{(0)} > 0$ miatt minden $\mathbf{q} \in \mathcal{E}$ eloszlásra $\mathbf{q} > 0$; és ha valamely j-re $q_j^{(0)} = 0$ értéket megengednénk, akkor minden $\mathbf{q} \in \mathcal{E}$ eloszlásra $q_j = 0$ lenne, vagyis vizsgálatunk eggyel kisebb számú eseményre történhetne.

Optimalizálási feladatok

A következőkben egy adott $\mathbf{A}$ mátrixhoz tartozó $\mathcal{L}$ és $\mathcal{E}$ eloszláscsaládokon adjuk meg az MDI és ML alapfeladatokat:

Minimális diszkrimináló információ feladat (MDI):	*Maximum likelihood feladat (ML):*
Legyen $\mathbf{q}^{(0)} \in \mathcal{E}$	Legyen $\mathbf{p}^{(0)} \in \mathcal{L}$
$\min D(\mathbf{p}\|\|\mathbf{q}^{(0)})$	$\min D(\mathbf{p}^{(0)}\|\|\mathbf{q})$
$\mathbf{p} \in \mathcal{L}$	$\mathbf{q} \in \mathcal{E}$

Az első feladatot D vetítésnek a másodikat L vetítésnek is szokás nevezni. Az alábbi lemma amely elemi eszközökkel belátható, már előrevetíti a dualitást.

2.1. LEMMA. *Legyen* $\mathbf{p}^{(0)} \in \mathcal{L}$, $\mathbf{q}^{(0)} \in \mathcal{E}$ *rögzített és* $\mathbf{p} \in \mathcal{L}$, $\mathbf{q} \in \mathcal{E}$ *pedig tetszőlegesek. Ekkor az alábbi egyenlőtlenség fenn áll.*

$$D(\mathbf{p}^{(0)}\|\mathbf{q}^{(0)}) \leq D(\mathbf{p}\|\mathbf{q}^{(0)}) + D(\mathbf{p}^{(0)}\|\mathbf{q}) \tag{2.3}$$

és egyenlőség akkor és csak akkor, ha

$$\mathbf{p} = \mathbf{q} \tag{2.4}$$

Bizonyítás. A (2.3) egyenlőséget kirészletezve és egyszerűsítve kapjuk, hogy

$$\sum_{j=1}^{n} p_j^{(0)} \log \frac{q_j}{q_j^{(0)}} \le \sum_{j=1}^{n} p_j \log \frac{p_j}{q_j^{(0)}}.$$

Ennek belátásához induljunk ki a

$$\sum_{j=1}^{n} p_j \log \frac{p_j}{q_j} \ge 0, \tag{2.5}$$

egyenlőtlenségből, ahol egyenlőség akkor és csak akkor teljesül, ha $\mathbf{p} = \mathbf{q}$.

Felhasználva (2.2)-t, (2.5)-ből kapjuk, hogy

$$\sum_{j=1}^{n} p_j \log \frac{p_j}{q_j^{(0)}} \ge \sum_{j=1}^{n} p_j(\mathbf{t}\mathbf{a}_j + u). \tag{2.6}$$

Felhasználva (2.1)-et, (2.6)-ból kapjuk, hogy

$$\sum_{j=1}^{n} p_j \log \frac{p_j}{q_j^{(0)}} \ge \sum_{j=1}^{n} p_j^{(0)}(\mathbf{t}\mathbf{a}_j + u) \tag{2.7}$$

De (2.7) jobb oldalán (2.2)-t helyettesítve

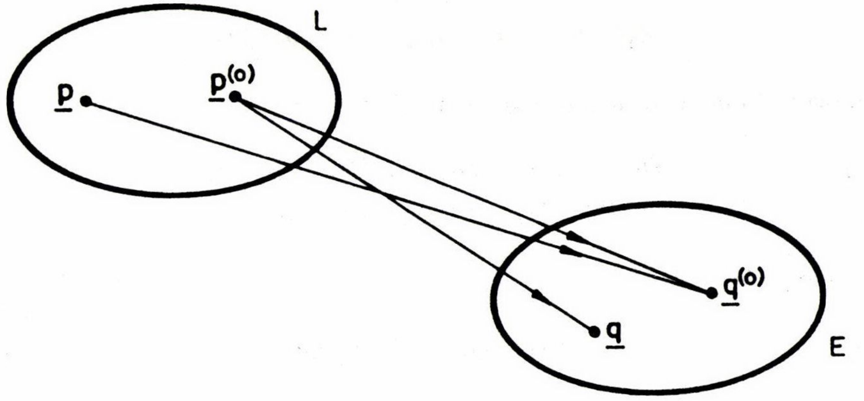
$$\sum_{j=1}^{n} p_j \log \frac{p_j}{q_j^{(0)}} \ge \sum_{j=1}^{n} p_j^{(0)} \log \frac{q_j}{q_j^{(0)}}.$$

Ezzel állításunkat bebizonyítottuk. □

Az alábbi következmény könnyen belátható.

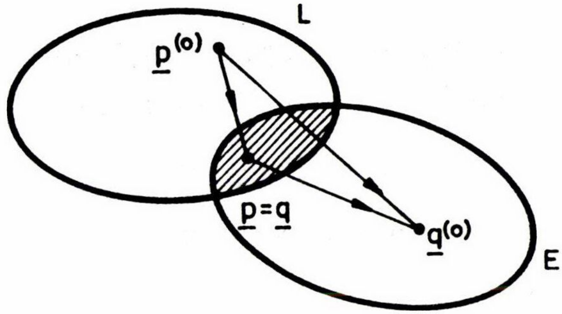
2.2. KÖVETKEZMÉNY. (Gyenge equilibrium). *Ha* (2.3)-*ban egyenlőség teljesül valamely* $\mathbf{p}^*$ *és* $\mathbf{q}^*$ *eloszlásra, akkor* $\mathbf{p}^*$ *az MDI a* $\mathbf{q}^*$ *pedig az ML feladat optimális megoldása.*

Megjegyzés. A lemma állítását sematikus ábrán szemléltetjük; a diszkrimináló információt, az eltérést, nyíllal ábrázoljuk. A nyíl talpa az „a posteriori" a nyíl hegye az „a priori".



1. ábra

Ha lenne az $\mathcal{L}$ és $\mathcal{E}$ halmaznak közös része, akkor a közös részben lévő minden $\mathbf{p} = \mathbf{q}$ pontra a lemma állítása egyenlőséggel teljesülne; szemléltetve



2. ábra

Megjegyezzük, hogy az $\mathcal{L}$ zárt, az $\mathcal{E}$ azonban nem zárt részhalmaza a szimplexnek. Megjegyezzük, hogy az $\mathcal{L}$ és $\mathcal{E}$ halmazoknak legfeljebb egy közös pontjuk lehet. Ugyanis indirekte, tegyük fel, hogy van kettő:

$$\mathbf{p}^* = \mathbf{q}^* \text{ és } \bar{\mathbf{p}} = \bar{\mathbf{q}} \tag{2.8}$$

akkor a lemma (2.4) és (2.3) állítása miatt $p^* = q^*$ majd $\bar{p} = \bar{q}$ pontokra:

$$D(p^{(0)}\|q^{(0)}) = D(p^*\|q^{(0)}) + D(p^{(0)}\|q^*), \tag{2.9}$$

$$D(p^{(0)}\|q^{(0)}) = D(\bar{p}\|q^{(0)}) + D(p^{(0)}\|\bar{q}). \tag{2.10}$$

A lemma (2.3) miatt a $(p^*, \bar{q})$ majd a $(\bar{p}, q^*)$ pontokra:

$$D(p^{(0)}\|q^{(0)}) \leq D(p^*\|q^{(0)}) + D(p^{(0)}\|\bar{q}), \tag{2.11}$$

$$D(p^{(0)}\|q^{(0)}) \leq D(\bar{p}\|q^{(0)}) + D(p^{(0)}\|q^*). \tag{2.12}$$

A (2.9) és (2.11)-ből:

$$D(p^{(0)}\|q^*) \leq D(p^{(0)}\|\bar{q}) \tag{2.13}$$

A (2.10) és (2.12)-ből:

$$D(p^{(0)}\|\bar{q}) \leq (p^{(0)}\|q^*) \tag{2.14}$$

Így (2.13) és (2.14) egybevetéséből adódik, hogy

$$D(p^{(0)}\|\bar{q}) = D(p^{(0)}\|q^*) \tag{2.15}$$

Ezt a (2.15)-öt a (2.9) és (2.10)-be visszatéve:

$$D(p^{(0)}\|q^{(0)}) = D(p^*\|q^{(0)}) + D(p^{(0)}\|\bar{q}), \tag{2.16}$$

$$D(p^{(0)}\|q^{(0)}) = D(\bar{p}\|q^{(0)}) + D(p^{(0)}\|q^*). \tag{2.17}$$

A (2.16) és (2.17)-ből a lemma szerint adódik, hogy

$$p^* = \bar{q} \quad \text{és} \quad \bar{p} = q^*.$$

De (2.8) miatt akkor

$$p^* = \bar{p} \quad \text{és} \quad q^* = \bar{q},$$

tehát valóban legfeljebb csak egy közös pont lehet. Könnyen végig gondolható, hogy az $\mathcal{L}$ és $\mathrm{cl}\mathcal{E}$ halmazoknak is csak egy közös pontjuk lehet, ahol $\mathrm{cl}\mathcal{E}$ az $\mathcal{E}$ halmaz lezártja, ugyanis a (2.3) egyenlőtlenség $q \in \mathrm{cl}\mathcal{E}$ esetén is igaz (megengedve a jobb oldalon a $+\infty$ értéket is).

A következő dualitási tétel szemléletes tartalma:

— (Gyenge forma). Ha van $\mathcal{L}$-nek pozitív pontja, akkor $\mathcal{L} \bigcap \mathcal{E}$ egy és csak egy pontból áll.

— (Erős forma). $\mathcal{L} \bigcap \mathrm{cl}\mathcal{E}$ egy és csak egy pontból áll, ahol $\mathrm{cl}\mathcal{E}$ az $\mathcal{E}$ halmaz lezártja.

2.3. TÉTEL.

1° (Gyenge forma). *Tegyük fel, hogy van olyan* $\mathbf{p} \in \mathcal{L}$, *hogy* $\mathbf{p} > \mathbf{0}$. *Ekkor létezik egy és csak egy* $\mathbf{p}^* \in \mathcal{L}$ *és* $\mathbf{q}^* \in \mathcal{E}$, $(\mathbf{p}^* = \mathbf{q}^*)$, *melyekre a*

$$D(\mathbf{p}^{(0)}\|\mathbf{q}^{(0)}) = D(\mathbf{p}^*\|\mathbf{q}^{(0)}) + D(\mathbf{p}^{(0)}\|\mathbf{q}^*) \tag{2.18}$$

ún. Pitagorasz-összefüggés teljesül.

2° (Erős forma). *Létezik egy és csak egy* $\mathbf{p}^* \in \mathcal{L}$ *és* $\mathbf{q}^* \in \mathrm{cl}\mathcal{E}$ $(\mathbf{p}^* = \mathbf{q}^*)$ *melyekre (2.18) fennáll.*

Bizonyítás. Az, hogy legfeljebb egy $\mathbf{p}^* \in \mathcal{L}$ van amelyre az állítások teljesülnek a lemma után tett megjegyzésünkből nyilvánvaló. Azt kell csak megmutatni, hogy legalább egy van.
A tétel bizonyításához a geometriai programozás gyenge és erős dualtitási tételeit fogjuk használni ([5]).

Először a feladatot geometriai programozási formára átírjuk. Adottak $\mathbf{q}^{(0)}$, $\mathbf{p}^{(0)}$ eloszlásvektorok és $\mathbf{A}$ mátrix. Legyen $\mathbf{b} := \mathbf{A}\mathbf{p}^{(0)}$.

Ekkor belső feltételes feladat

$$\left.\begin{aligned} \mathbf{A}\mathbf{p} &= \mathbf{b} \\ \sum_{j=1}^{n} p_j &= 1 \\ \mathbf{p} &\geq \mathbf{0} \end{aligned}\right\}$$

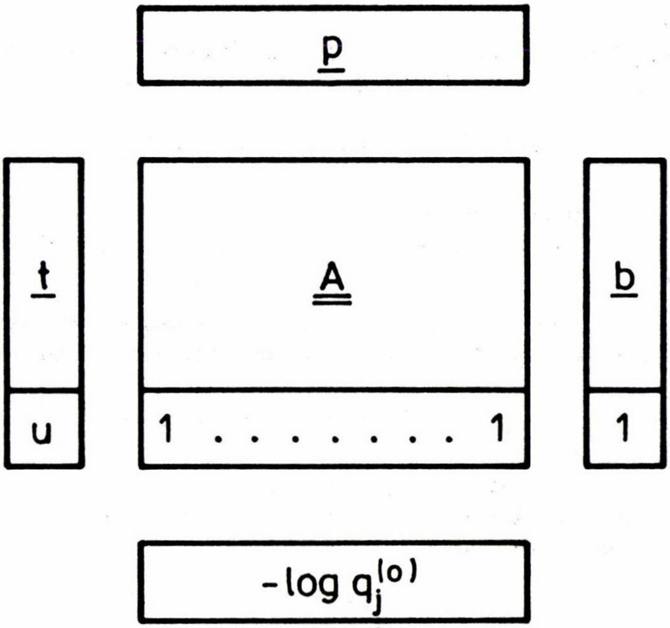
$$\sum_{j=1}^{n} p_j \log \frac{p_j}{q_j^{(0)}} \quad \text{minimalizálandó!}$$

A célfüggvény még az alábbi formában is írható

$$\sum_{j=1}^{n} p_j(-\log q_j^{(0)}) + \log \frac{\prod_{j=1}^{n} p_j^{p_j}}{\left(\sum_{j=1}^{n} p_j\right)^{\sum_{j=1}^{n} p_j}}$$

minimalizálandó!

Ez egy geometriai programozási duálfeladat. A könnyebb követhetőség érdekében sematikusan szemléltejük a paramétereket:



3. ábra

Legyenek a primálváltozók $\mathbf{t}$ m dimenziós vektor és u skalár.
A primál feladat feltételi halmaza:

$$\sum_{j=1}^{n} \exp(\mathbf{t}\mathbf{a}_j + u + \log q_j^{(0)}) \leq 1 \tag{2.19}$$

a céfüggvénye:

$$\mathbf{t}\mathbf{b} + u \quad \text{maximalizálandó.}$$

Azonnal látható, hogy a primál feladat *Slater reguláris*.
Az egyensúlyi feltételek

$$q_j^{(0)} \exp(\mathbf{t}\mathbf{a}_j + u) = p_j, \quad (j = 1, \ldots, n). \tag{2.20}$$

1° A geometriai programozás gyenge dualitási tételét használva az alábbiakat kapjuk:

Ha van $\mathbf{p} \in \mathcal{L}$, hogy $\mathbf{p} > \mathbf{0}$ akkor léteznek olyan $\mathbf{p}^* \in \mathcal{L}$ és $\mathbf{t}^*$, u^* primál megoldások, amelyek a (2.20) egyensúlyi feltételt teljesítik, azaz

$$q_j^{(0)} \exp(\mathbf{t}^*\mathbf{a}_j + u^*) = p_j^*, \quad (j = 1, \ldots, n) \tag{2.21}$$

és a célfüggvényértékekre

$$\sum_{j=1}^{n} p_j^* \log \frac{p_j^*}{q_j^{(0)}} = \mathbf{t}^*\mathbf{b} + u^* \tag{2.22}$$

A (2.21)-ből adódik, hogy

$$u^* = -\log \sum_{j=1}^{n} q_j^{(0)} \exp(\mathbf{t}^*\mathbf{a}_j).$$

Jelöljük

$$q_j^* := q_j^{(0)} \exp(\mathbf{t}^*\mathbf{a}_j + u^*) \quad (j = 1, \ldots, n)$$

Ebből

$$\log \frac{q_j^*}{q_j^{(0)}} = \mathbf{t}^*\mathbf{a}_j + u^* \quad (j = 1, \ldots, n) \tag{2.23}$$

A (2.22)-ben b értéket visszaírva kapjuk, hogy

$$\sum_{j=1}^{n} p_j^* \log \frac{p_j^*}{q_j^{(0)}} = \sum_{j=1}^{n} (\mathbf{t}^*\mathbf{a}_j + u^*) p_j^{(0)}$$

A (2.23)-at visszaírva a fenti egyenlőségbe írhatjuk, hogy

$$\sum_{j=1}^{n} p_j^* \log \frac{p_j^*}{q_j^{(0)}} = \sum_{j=1}^{n} p_j^{(0)} \log \frac{p_j^{(0)}}{q_j^{(0)}} - \sum_{j=1}^{n} p_j^{(0)} \log \frac{p_j^{(0)}}{q_j^*}.$$

Azaz

$$D(\mathbf{p}^{(0)} \| \mathbf{q}^{(0)}) = D(\mathbf{p}^{(0)} \| \mathbf{q}^*) + D(\mathbf{p}^* \| \mathbf{q}^{(0)}).$$

2° A geometriai programozás erős dualitási tétele szerint létezik $\mathbf{p}^* \in \mathcal{L}$ úgy, hogy

$$\sum_{j=1}^{n} p_j^* \log \frac{p_j^*}{q_j^{(0)}} = \sup_{(\mathbf{t},u)} (\mathbf{t}\mathbf{b} + u) \tag{2.24}$$

és a szupremumot nem csökkenti, ha (2.19)-ben egyenlőséget tételezünk fel, azaz, hogy

$$u = -\log \sum_{j=1}^{n} q_j^{(0)} \exp(\mathbf{t}\mathbf{a}_j). \tag{2.25}$$

Jelöljük

$$q_j := q_j^{(0)} \exp(\mathrm{t}a_j + u) \quad (j = 1, \dots, n) \tag{2.26}$$

Így (2.24) az alábbi formát ölti:

$$\sum_{j=1}^{n} p_j^* \log \frac{p_j^*}{q_j^{(0)}} = \sup_{(q)} \left(\sum_{j=1}^{n} p_j^{(0)} \log \frac{q_j}{q_j^{(0)}} \right)$$

és q a (2.25), (2.26) feltételeket teljesíti. Minthogy a (2.25) és (2.26) feltételek az $\mathcal{E}$ halmazt definiálják és $\mathcal{E}$ része a szimplexnek a log függvény folytonossága miatt:

$$\sup_{\mathrm{q}\in\mathcal{E}} \left(\sum_{j=1}^{n} p_j^{(0)} \log \frac{q_j}{q_j^{(0)}} \right) = \max_{\mathrm{q}\in \mathrm{cl}\mathcal{E}} \left(\sum_{j=1}^{n} p_j^{(0)} \log \frac{q_j}{q_j^{(0)}} \right)$$

Azonban

$$\max_{\mathrm{q}\in \mathrm{cl}\mathcal{E}} \left(\sum_{j=1}^{n} p_j^{(0)} \log \frac{q_j}{q_j^{(0)}} \right) = \sum_{j=1}^{n} p_j^{(0)} \log \frac{p_j^{(0)}}{q_j^{(0)}} - \min_{\mathrm{q}\in \mathrm{cl}\mathcal{E}} \left(\sum_{j=1}^{n} p_j^{(0)} \log \frac{p_j^{(0)}}{q_j}. \right)$$

Legyen $\mathrm{q}^* \in \mathrm{cl}\mathcal{E}$ olyan, hogy

$$\min_{\mathrm{q}\in \mathrm{cl}\mathcal{E}} \left(\sum_{j=1}^{n} p_j^{(0)} \log \frac{p_j^{(0)}}{q_j} \right) = \sum_{j=1}^{n} p_j^{(0)} \log \frac{p_j^{(0)}}{q_j^*}.$$

Így kapjuk, hogy

$$D(\mathrm{p}^{(0)}\|\mathrm{q}^{(0)}) = D(\mathrm{p}^{(0)}\|\mathrm{q}^*) + D(\mathrm{p}^*\|\mathrm{q}^{(0)}) \qquad \square$$

2.4. KÖVETKEZMÉNY. (Erős equilibrium): *Ha a* $\mathrm{q}^* \in \mathcal{L}$ *az MDI feladat optimális megoldása és ha a* $\mathrm{q}^* \in \mathrm{cl}\mathcal{E}$ *az ML feladat optimális megoldása (azaz* $\{\inf D(\mathrm{p}^{(0)}\|\mathrm{q}) \mid \mathrm{q} \in \mathcal{E}\} = D(\mathrm{p}^{(0)}\|\mathrm{q}^*)$, $\mathrm{q}^* \in \mathrm{cl}\mathcal{E}$) *akkor a (2.18) Pitagorasz-összefüggés fennáll.*

Bizonyítás. Legyen $\bar{\mathrm{p}}$, $\bar{\mathrm{q}}$ a 2.3 tétel által biztosított optimális megoldások. Ezekre tudjuk, hogy

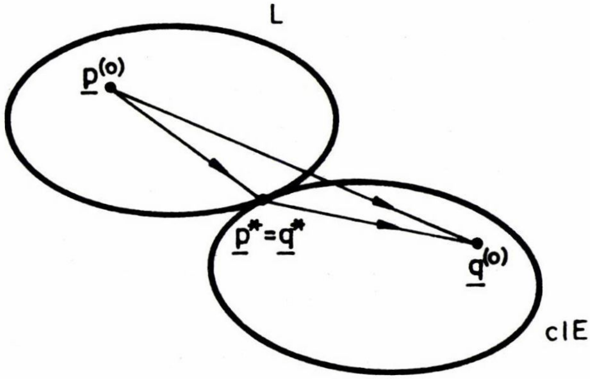
$$D(\bar{\mathrm{p}}\|\mathrm{q}^{(0)}) + D(\mathrm{p}^{(0)}\|\bar{\mathrm{q}}) = D(\mathrm{p}^{(0)}\|\mathrm{q}^{(0)}). \tag{2.27}$$

Minthogy p^* is és q^* is optimálisak így

$$\begin{aligned} D(\mathrm{p}^*\|\mathrm{q}^{(0)}) &= D(\bar{\mathrm{p}}\|\mathrm{q}^{(0)}) \\ D(\mathrm{p}^{(0)}\|\mathrm{q}^*) &= D(\mathrm{p}^{(0)}\|\bar{\mathrm{q}}) \end{aligned} \tag{2.28}$$

A (2.28)-at (2.27)-be helyettesítve:

$$D(\mathrm{p}^*\|\mathrm{q}^{(0)}) + D(\mathrm{p}^{(0)}\|\mathrm{q}^*) = D(\mathrm{p}^{(0)}\|\mathrm{q}^{(0)}) \qquad \square$$



4. *ábra*

Megjegyzés. A tétel állítását sematikus ábrán szemléltethetjük (jelölés ugyanaz mint a lemma utáni megjegyzésünkben)

$$D(\mathbf{p}^*\|\mathbf{q}^{(0)}) + D(\mathbf{p}^{(0)}\|\mathbf{q}^*) = D(\mathbf{p}^{(0)}\|\mathbf{q}^{(0)}).$$

Ezt az alábbi formában is írhatjuk:

$$D(\mathbf{p}^{(0)}\|\mathbf{q}^{(0)}) = \min_{(\mathbf{p})} D(\mathbf{p}\|\mathbf{q}^{(0)}) + \inf_{(\mathbf{q})} D(\mathbf{p}^{(0)}\|\mathbf{q}),$$

ahol $\min_{(\mathbf{p})} D(\mathbf{p}\|\mathbf{q}^{(0)})$ az úgynevezett diszkrimináló vetítés (röviden **D** vetítés) és $\inf_{(\mathbf{q})} D(\mathbf{p}^{(0)}\|\mathbf{q})$ az úgynevezett likelihood vetítés (röviden **L** vetítés). Ha van $\mathbf{p} > \mathbf{0}$ pont $\mathcal{L}$-ben, akkor az infimum minimumra cserélhető.

2.5. KÖVETKEZMÉNY. *Legyenek* $\mathbf{p}^{(0)}$, $\mathbf{q}^{(0)}$ *adott eloszlások.*

Legyen $\tilde{\mathbf{A}}$ *olyan mátrix, amit az* $\mathbf{A}$ *mátrixból újabb sorok hozzávételével kapunk. Az* $\tilde{\mathcal{L}}$ *eloszláscsaládot a következőképpen definiáljuk:*

$$\tilde{\mathcal{L}} := \{\mathbf{p} \mid \tilde{\mathbf{A}}\mathbf{p} = \tilde{\mathbf{A}}\mathbf{p}^{(0)}\}$$

Az $\mathcal{L}$ *eloszláscsalád továbbra is:*

$$\mathcal{L} := \{\mathbf{p} \mid \mathbf{A}\mathbf{p} = \mathbf{A}\mathbf{p}^{(0)}\}.$$

így nyilvánvaló, hogy $\mathbf{p}^{(0)} \in \tilde{\mathcal{L}} \subset \mathcal{L}$.

Az $\tilde{\mathcal{E}}$ *eloszláscsaládot az alábbiak szerint definiáljuk:*

$$\tilde{\mathcal{E}} := \{\mathbf{q} \mid q_j = q_j^{(0)} \exp(\mathbf{t}\tilde{\mathbf{a}}_j + u), \quad (j = 1, \ldots, n)\}$$

$\mathbf{t} = (t_1, \ldots, t_m)$ *tetszőleges,* $u = -\log \sum_{j=1}^{n} q_j^{(0)} \exp(\mathbf{t}\tilde{a}_j)$, $(j = 1, \ldots, n)$.

Az $\mathcal{E}$ *eloszláscsalád pedig most is:*

$$\mathcal{E} := \{\mathbf{q} \mid q_j = q_j^{(0)} \exp(\mathbf{t}\mathbf{a}_j + u), \quad (j = 1, \ldots, n)\}$$

$\mathbf{t} = (t_1, \ldots, t_m)$ *tetszőleges,* $u = -\log \sum_{j=1}^{n} q_j^{(0)} \exp(\mathbf{t}\mathbf{a}_j)$, $(j = 1, \ldots, n)$. *Nyilvánvaló, hogy* $\mathbf{q}^{(0)} \in \mathcal{E} \subset \tilde{\mathcal{E}}$.

Az alábbiakban a tétel következményeként egy állítást bizonyítunk be: Legyenek $\mathbf{p}^{(0)} \in \tilde{\mathcal{L}}$ és $\mathbf{q}^{(0)} \in \mathrm{cl}\mathcal{E}$ tetszőlgesek. Legyen

$$\min_{\mathbf{p} \in \mathcal{L}} D(\mathbf{p} \| \mathbf{q}^{(0)}) = D(\mathbf{p}^* \| \mathbf{q}^{(0)}), \tag{2.29}$$

$$\min_{\mathbf{p} \in \tilde{\mathcal{L}}} D(\mathbf{p} \| \mathbf{q}^{(0)}) = D(\tilde{\mathbf{p}}^* \| \mathbf{q}^{(0)}). \tag{2.30}$$

Legyen

$$\min_{\mathbf{q} \in \mathrm{cl}\mathcal{E}} D(\mathbf{p}^{(0)} \| \mathbf{q}) = D(\mathbf{p}^{(0)} \| \mathbf{q}^*), \tag{2.31}$$

$$\min_{\mathbf{q} \in \mathrm{cl}\tilde{\mathcal{E}}} D(\mathbf{p}^{(0)} \| \mathbf{q}) = D(\mathbf{p}^{(0)} \| \tilde{\mathbf{q}}^*). \tag{2.32}$$

ÁLLÍTÁS.

1) $\min_{\mathbf{p} \in \tilde{\mathcal{L}}} D(\mathbf{p} \| \mathbf{p}^*) = D(\tilde{\mathbf{p}}^* \| \mathbf{p}^*)$.

A $\tilde{\mathbf{p}}^*$*-ot a* $\mathbf{p}^*$*-nak* $\tilde{\mathcal{L}}$ *halmazra vonatkozó* D *vetületének nevezzük.*

2) $\min_{\mathbf{q} \in \mathrm{cl}\mathcal{E}} D(\tilde{\mathbf{q}}^* \| \mathbf{q}) = D(\tilde{\mathbf{q}}^* \| \mathbf{q}^*)$.

A $\mathbf{q}^*$*-ot a* $\tilde{\mathbf{q}}^*$*-nak* $\mathrm{cl}\mathcal{E}$ *halmazra vonatkozó* L *vetületének nevezzük.*

Bizonyítás. (2.29) és (2.31)-ből, valamint az erős equilibriumból adódik, hogy

$$D(\mathbf{p}^{(0)} \| \mathbf{q}^{(0)}) = D(\mathbf{p}^* \| \mathbf{q}^{(0)}) + D(\mathbf{p}^{(0)} \| \mathbf{q}^*) \quad \text{és} \quad \mathbf{p}^* = \mathbf{q}^* \tag{2.33}$$

(2.30) és (2.32)-ből és ugyancsak az erős equilibriumból adódik, hogy

$$D(\mathbf{p}^{(0)} \| \mathbf{q}^{(0)}) = D(\tilde{\mathbf{p}}^* \| \mathbf{q}^{(0)}) + D(\mathbf{p}^{(0)} \| \tilde{\mathbf{q}}^*) \quad \text{és} \quad \tilde{\mathbf{p}}^* = \tilde{\mathbf{q}}^*. \tag{2.34}$$

Mivel $\mathbf{q}^{(0)} \in \mathrm{cl}\mathcal{E}$ tetszőleges, most vegye át a szerepét $\mathbf{q}^*$ (2.34)-ből és (2.33) szerint figyelembe véve, hogy $\mathbf{q}^* = \mathbf{p}^*$; az alábbit kapjuk:

$$D(\mathbf{p}^{(0} \| \mathbf{p}^*) = D(\tilde{\mathbf{p}}^* \| \mathbf{p}^*) + D(\mathbf{p}^{(0)} \| \tilde{\mathbf{p}}^*).$$

Mivel $D(\mathbf{p}^{(0)} \| \tilde{\mathbf{p}}^*) \geq 0$, azért

$$D(\mathbf{p}^{(0)} \| \mathbf{p}^*) \geq D(\tilde{\mathbf{p}}^* \| \mathbf{p}^*).$$

A fenti egyenlőtlenségből és minthogy $p^{(0)}$ szerepét bármilyen $\mathbf{p} \in \tilde{\mathcal{L}}$ veheti át, ezért

$$\min_{\mathbf{p}\in\tilde{\mathcal{L}}} D(\mathbf{p}\|\mathbf{p}^*) = D(\tilde{\mathbf{p}}^*\|\mathbf{p}^*).$$

Így az állítás első részét bebizonyítottuk.
$\mathbf{p}^{(0)} \in \tilde{\mathcal{L}}$ tetszőleges, így most a szerepét $\tilde{\mathbf{p}}^*$ vegye át. (2.33)-ból és (2.34) szerint figyelembe véve, hogy $\tilde{\mathbf{p}}^* = \tilde{\mathbf{q}}^*$, azt kapjuk hogy:

$$D(\tilde{\mathbf{q}}^*\|\mathbf{q}^{(0)}) = D(\mathbf{q}^*\|\mathbf{q}^{(0)}) + D(\tilde{\mathbf{q}}^*\|\mathbf{q}^*).$$

Mivel $D(\mathbf{q}^*\|\mathbf{q}^{(0)} \geq 0$, azért

$$D(\tilde{\mathbf{q}}^*\|\mathbf{q}^{(0)}) \geq D(\tilde{\mathbf{q}}^*\|\mathbf{q}^*).$$

A fenti egyenlőtlenségből és minthogy $\mathbf{q}^{(0)}$ szerepét bármilyen $\mathbf{q}$ veheti át, azt kapjuk, hogy:

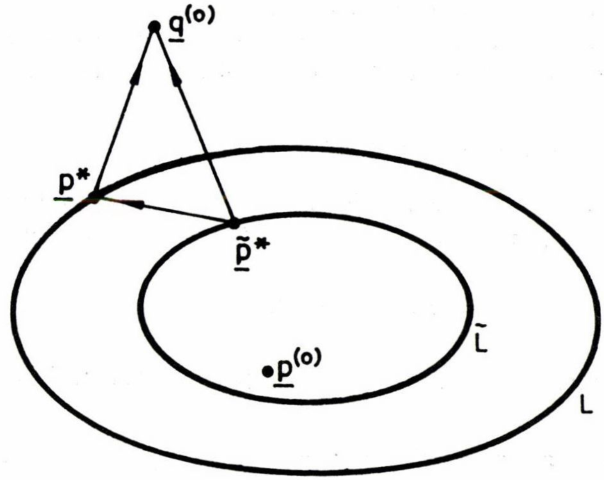
$$\min_{\mathbf{q}\in\mathrm{cl}\mathcal{E}} D(\tilde{\mathbf{q}}^*\|\mathbf{q}) = D(\tilde{\mathbf{q}}^*\|\mathbf{q}^*).$$

Tehát az állítás második részét is bebizonyítottuk. □

Ha feltételezzük még, hogy van olyan $\mathbf{p} \in \tilde{\mathcal{L}}$, $\mathbf{p} > \mathbf{0}$, akkor az előbbiekben a cl$\mathcal{E}$-re illetve cl$\tilde{\mathcal{E}}$-re vonatkozóak $\mathcal{E}$-re és $\tilde{\mathcal{E}}$-ra is érvényesek.

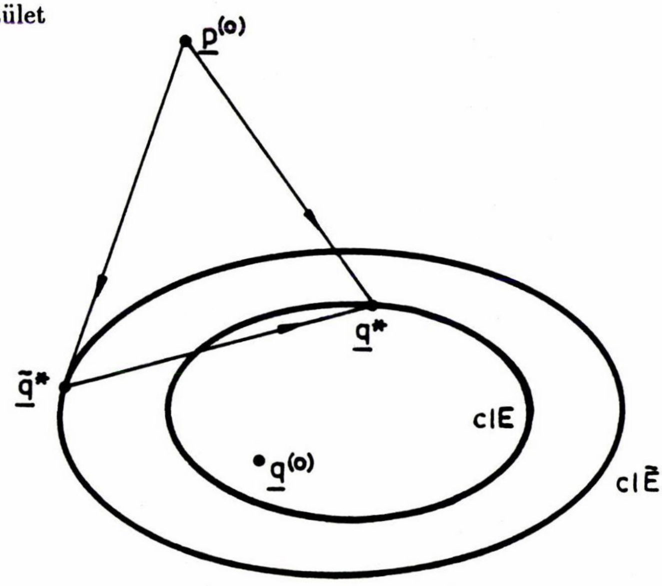
A következmény állításait ábrán szemléltetve

1) **D** vetület



5. ábra

2) **L** vetület



6. ábra

3. Lineáris és exponenciális eloszláscsalád optimalizációs feladatának dualitási kapcsolata: az összetett eset

Legyen **A** tetszőlegesen adott $m \times n$-es mátrix, oszlopvektorait jelöljük $\mathbf{a}_j$-vel, és legyen **B** tetszőleges adott $m \times k$-s mátrix, oszlopvektorait jelöljük $\mathbf{b}_i$-vel. Az **A** és **B** mátrix segítségével definiáljuk az $\mathcal{L}$ és $\mathcal{E}$ eloszláscsaládokat.

Belső feltételes, lineáris eloszláscsalád

Legyenek $\mathbf{p}^{(0)}$ és $\mathbf{r}^{(0)}$ adott eloszlások. Az alábbi implicit módon lineáris feltételekkel adott $(\mathbf{p},\mathbf{r})$ eloszlások összességét $\mathcal{L}$ lineáris családnak nevezzük.

(3.1) $$\mathbf{Ap} + \mathbf{Br} = \mathbf{Ap}^{(0)} + \mathbf{Br}^{(0)}.$$

Nyilvánvalóan $(\mathbf{p}^{(0)},\mathbf{r}^{(0)}) \in \mathcal{L}$ és bármelyik $(\mathbf{p},\mathbf{r})$ átveheti $(\mathbf{p}^{(0)},\mathbf{r}^{(0)})$ szerepét.

Külső feltételes, exponenciális eloszlás

Legyenek $\mathbf{q}^{(0)} > \mathbf{0}$ és $\mathbf{s}^{(0)} > \mathbf{0}$ adott eloszlások. Az alábbi explicit módon exponenciális (loglineáris) feltételekkel adott $(\mathbf{q},\mathbf{s})$ eloszlások összességét $\mathcal{E}$ exponenciális eloszláscsaládnak nevezzük.

$$
(3.2)\qquad
\begin{aligned}
q_j &= q_j^{(0)} \exp(\mathbf{t}\mathbf{a}_j + u), \quad (j = 1, \dots, n)\\
s_j &= s_j^{(0)} \exp(\mathbf{t}\mathbf{b}_j + v), \quad (j = 1, \dots, k)\\
\text{ahol}\quad \mathbf{t} &= (t_1, \dots, t_m) \quad \text{és}\\
u &= -\log \sum_{j=1}^{n} q_j^{(0)} \exp(\mathbf{t}\mathbf{a}_j)\\
v &= -\log \sum_{j=1}^{k} s_j^{(0)} \exp(\mathbf{t}\mathbf{b}_j)
\end{aligned}
$$

(az u, v előbbi megválasztása azért szükségs, hogy q_j és s_j eloszlások legyenek)

Azonnal látható, hogy $(\mathbf{q}^{(0)}, \mathbf{s}^{(0)}) \in \mathcal{E}$ és bármelyik $(\mathbf{q}, \mathbf{s})$ pár átveheti $(\mathbf{q}^{(0)}, \mathbf{s}^{(0)})$ szerepét.

Ahogy az előbbiekben is, megjegyezzük, hogy $(\mathbf{q}^{(0)}, \mathbf{s}^{(0)}) > 0$ miatt minden $(\mathbf{q}, \mathbf{s}) \in \mathcal{E}$ eloszlásra $(\mathbf{q}, \mathbf{s}) > 0$.

Optimalizálási feladatok az összetett esetre

A következőkben adott $\mathbf{A}$ és $\mathbf{B}$ mátrixhoz tartozó $\mathcal{L}$ és $\mathcal{E}$ eloszláscsaládokon adjuk meg az MDI és ML alapfeladatokat.

MDI-feladat (D vetítés)
Legyenek $(\mathbf{q}^{(0)}, \mathbf{s}^{(0)}) \in \mathcal{E}$
$$\min_{(\mathbf{p},\mathbf{r})\in\mathcal{E}} D(\mathbf{p}\|\mathbf{q}^{(0)}) + D(\mathbf{r}\|\mathbf{s}^{(0)})$$

ML-feladat (L vetítés)
Legyenek $(\mathbf{p}^{(0)}, \mathbf{r}^{(0)}) \in \mathcal{L}$
$$\min_{(\mathbf{p},\mathbf{s})\in\mathcal{E}} D(\mathbf{p}^{(0)}\|\mathbf{q}) + D(\mathbf{r}^{(0)}\|\mathbf{s})$$

Az előző esethez hasonlóan a következő lemmát látjuk be.

3.1. Lemma. *Legyenek* $(\mathbf{p}^{(0)}, \mathbf{r}^{(0)}) \in \mathcal{L}$, $(\mathbf{q}^{(0)}, \mathbf{s}^{(0)}) \in \mathcal{E}$ *rögzítettek,* $(\mathbf{p}, \mathbf{r}) \in \mathcal{L}$ *és* $(\mathbf{q}, \mathbf{s}) \in \mathcal{E}$ *pedig tetszőlegesek. Az alábbi egyenlőtlenség fenn áll.*

$$
(3.3)\qquad
\begin{aligned}
D(\mathbf{p}^{(0)}\|\mathbf{q}^{(0)}) + D(\mathbf{r}^{(0)}\|\mathbf{s}^{(0)}) \le & D(\mathbf{p}\|\mathbf{q}^{(0)}) + D(\mathbf{p}^{(0)}\|\mathbf{q}) +\\
& + D(\mathbf{r}\|\mathbf{s}^{(0)}) + D(\mathbf{r}^{(0)}\|\mathbf{s})
\end{aligned}
$$

és egyenlőség akkor és csak akkor, ha

$$(\mathbf{p}, \mathbf{r}) = (\mathbf{q}, \mathbf{s}).$$

Ebben az esetben fennállnak az alábbi egyenlőségek is:

$$(3.4)\qquad D(\mathbf{p}^{(0)}\|\mathbf{q}^{(0)}) = D(\mathbf{p}\|\mathbf{q}^{(0)}) + D(\mathbf{p}^{(0)}\|\mathbf{q})$$

és

$$(3.5)\qquad D(\mathbf{r}^{(0)}\|\mathbf{s}^{(0)}) = D(\mathbf{r}\|\mathbf{s}^{(0)}) + D(\mathbf{r}^{(0)}\|\mathbf{s})$$

Megjegyzés. Mielőtt a bizonyításra rátérnénk megmutatjuk, hogy ha a (3.3)-ban egyenlőség van akkor az két (3.4) és (3.5) egyenlőségre esik szét.

A megjegyzés indoklása.

Az egyenlőség minden $(\mathbf{p},\mathbf{r}) \in \mathcal{L}$ és minden $(\mathbf{q},\mathbf{s}) \in \mathcal{E}$ esetén teljesül, ugyanis $(\mathbf{p}^{(0)},\mathbf{r}^{(0)})$ és $(\mathbf{q}^{(0)},\mathbf{s}^{(0)})$ szerepét bármilyen $(\mathbf{p},\mathbf{r}) \in \mathcal{L}$ és $(\mathbf{q},\mathbf{s}) \in \mathcal{E}$ átveheti. Így kapjuk, hogy

$$D(\mathbf{p}\|\mathbf{q}) + D(\mathbf{p}\|\mathbf{q}) + D(\mathbf{r}\|\mathbf{s}) + D(\mathbf{r}\|\mathbf{s}) = D(\mathbf{p}\|\mathbf{q}) + D(\mathbf{r}\|\mathbf{s}).$$

Tehát

$$D(\mathbf{p}\|\mathbf{q}) + D(\mathbf{r}\|\mathbf{s}) = 0.$$

De a divergencia mindig nem negatív. Tehát a fenti egyenlőség akkor és csak akkor érvényes, ha

$$D(\mathbf{p}\|\mathbf{q}) = 0 \quad \text{és} \quad D(\mathbf{r}\|\mathbf{s}) = 0.$$

Amiből az következik, hogy $\mathbf{p} = \mathbf{q}$ és $\mathbf{r} = \mathbf{s}$. Külön-külön (3.3)-ban (egyenlőség esetén) visszahelyettesítve kapjuk, hogy

$$D(\mathbf{p}\|\mathbf{q}^{(0)}) + D(\mathbf{p}^{(0)}\|\mathbf{q}) = D(\mathbf{p}^{(0)}\|\mathbf{q}^{(0)}),$$

és

$$D(\mathbf{r}\|\mathbf{s}^{(0)}) + D(\mathbf{r}^{(0)}\|\mathbf{s}) = D(\mathbf{r}^{(0)}\|\mathbf{s}^{(0)}).$$

Azaz blokkonként érvényes a *Pitagorasz-összefüggés.*

Bizonyítás. Ha a (3.3) egyenlőtlenséget kirészletezzük és a lehetséges egyszerűsítéseket elvégezzük, az alábbi egyenlőtlenséget kapjuk

$$\sum_{j=1}^{n} p_j^{(0)} \log \frac{q_j}{q_j^{(0)}} + \sum_{j=1}^{n} r_j^{(0)} \log \frac{s_j}{s_j^{(0)}} \le \sum_{j=1}^{n} p_j \log \frac{p_j}{q_j^{(0)}} + \sum_{j=1}^{k} r_j \log \frac{r_j}{s_j^{(0)}}.$$

A bizonyításhoz induljunk ki abból, hogy

$$\sum_{j=1}^{n} p_j \log \frac{p_j}{q_j} \ge 0 \tag{3.6}$$

és

$$\sum_{j=1}^{k} r_j \log \frac{r_j}{s_j} \ge 0 \tag{3.7}$$

és egyenlőség akkor és csak akkor teljesül, ha

$$\mathbf{p} = \mathbf{q} \quad \text{illetve} \quad \mathbf{r} = \mathbf{s}$$

Felhasználva (3.2)-t, (3.6) és (3.7)-ből kapjuk, hogy

$$\sum_{j=1}^{n} p_j \log p_j \geq \sum_{j=1}^{n} p_j(\log q_j^{(0)} + \mathbf{t}\mathbf{a}_j + u)$$

és

$$\sum_{j=1}^{k} r_j \log r_j \geq \sum_{j=1}^{k} r_j(\log s_j^{(0)} + \mathbf{t}\mathbf{b}_j + v).$$

Majd a két egyenlőséget átrendezve és összeadva az alábbi egyenlőtlenséget kapjuk:

$$\sum_{j=1}^{n} p_j \log \frac{p_j}{q_j^{(0)}} + \sum_{j=1}^{k} r_j \log \frac{r_j}{s_j^{(0)}} \geq \sum_{j=1}^{n} p_j \mathbf{t}\mathbf{a}_j + u + \sum_{j=1}^{k} r_j \mathbf{t}\mathbf{b}_j + v. \tag{3.8}$$

Felhasználva (3.1)-et (3.8) az alábbi alakot ölti:

$$\sum_{j=1}^{n} p_j \log \frac{p_j}{q_j^{(0)}} + \sum_{j=1}^{k} r_j \log \frac{r_j}{s_j^{(0)}} \geq \sum_{j=1}^{n} p_j^{(0)}(\mathbf{t}\mathbf{a}_j + u) + \sum_{j=1}^{k} r_j^{(0)}(\mathbf{t}\mathbf{b}_j + v).$$

(3.2)-ből pedig

$$\sum_{j=1}^{n} p_j \log \frac{p_j}{q_j^{(0)}} + \sum_{j=1}^{k} r_j \log \frac{r_j}{s_j^{(0)}} \geq \sum_{j=1}^{n} p_j^{(0)} \log \frac{q_j}{q_j^{(0)}} + \sum_{j=1}^{k} r_j^{(0)} \log \frac{s_j}{s_j^{(0)}}$$

és egyenlőség akkor és csak akkor, ha

$$(\mathbf{p}, \mathbf{r}) = (\mathbf{q}, \mathbf{s}) \qquad \square$$

Az alábbi következmény könnyen belátható.

3.2. KÖVETKEZMÉNY. (Gyenge equilibrium)
Ha (3.3)-ban teljesül az egyenlőség valamely $(\mathbf{p}^*, \mathbf{r}^*)$ *és* $(\mathbf{q}^*, \mathbf{s}^*)$ *eloszlásra, akkor a* $(\mathbf{p}^*, \mathbf{r}^*)$ *az MDI, a* $(\mathbf{q}^*, \mathbf{s}^*)$ *pedig az ML feladat optimális megoldása.*

Megjegyzés. Egyenlőség esetén két (3.4), (3.5) egyenlőségi állítást kapunk. Ebből, minthogy a szimpla esetben beláttuk, hogy $\mathcal{L} \bigcap \mathrm{cl}\mathcal{E}$ csak egy pontból állhat, következik, hogy most is csak legfeljebb egy közös pont lehet.

3.3. TÉTEL.
1° (Gyenge forma). *Tegyük fel, hogy van olyan* $(\mathbf{p}, \mathbf{r}) \in \mathcal{L}$, *hogy* $(\mathbf{p}, \mathbf{r}) > 0$. *Ekkor létezik egy és csak egy* $(\mathbf{p}^*, \mathbf{r}^*) \in \mathcal{L}$ *és* $(\mathbf{q}^*, \mathbf{s}^*) \in \mathcal{E}$ *úgyhogy* $(\mathbf{p}^*, \mathbf{r}^*) = (\mathbf{q}^*, \mathbf{s}^*)$, *amelyekre a*

$$D(\mathbf{p}^{(0)} \| \mathbf{q}^{(0)}) = D(\mathbf{p}^* \| \mathbf{q}^{(0)}) + D(\mathbf{p}^{(0)} \| \mathbf{q}^*) \tag{3.9}$$

és

$$D(\mathbf{r}^{(0)}\|\mathbf{s}^{(0)}) = D(\mathbf{r}^*\|\mathbf{s}^{(0)}) + D(\mathbf{r}^{(0)}\|\mathbf{s}^*) \tag{3.10}$$

úgynevezett Pitagorasz-összefüggések telejsülnek.

2° (Erős forma). *Létezik egy és csak egy* $(\mathbf{p}^*, \mathbf{r}^*) \in \mathcal{L}$ *és* $(\mathbf{q}^*, \mathbf{s}^*) \in \mathrm{cl}\mathcal{E}$, *úgyhogy* $(\mathbf{p}^*, \mathbf{r}^*) = (\mathbf{q}^*, \mathbf{s}^*)$, *melyekre (3.9) és (3.10) fennáll.*

Bizonyítás. Az, hogy egy és csak egy $(\mathbf{p}^*, \mathbf{r}^*)$ van amelyre az állítások teljesülnek a fent tett megjegyzésünkből nyilvánvalóan adódik. Az alábbiakban azt bizonyítjuk be, hogy legalább egy van. A szimpla esethez hasonlóan, a tétel bizonyításához itt is a geometriai programozás gyenge és erős tételeit fogjuk használni [5]. Mindenekelőtt a feladatot geometriai programozási formára írjuk át.

Adottak $\mathbf{p}^{(0)}$, $\mathbf{r}^{(0}$, $\mathbf{q}^{(0)}$, $\mathbf{s}^{(0)}$ eloszlásvektorok és $\mathbf{A}$, $\mathbf{B}$ mátrixok. Legyen $\mathbf{d} := \mathbf{A}\mathbf{p}^{(0)} + \mathbf{B}\mathbf{r}^{(0)}$. Ekkor a belső feltételes feladat

$$\begin{aligned} &\mathbf{A}\mathbf{p} + \mathbf{B}\mathbf{r} = \mathbf{d} \\ &\sum_{j=1}^{n} p_j = 1 \qquad p_j \geq 0, \quad (j = 1, \ldots, n) \\ &\sum_{j=1}^{k} r_j = 1 \qquad r_j \geq 0, \quad (j = 1, \ldots, k) \end{aligned}$$

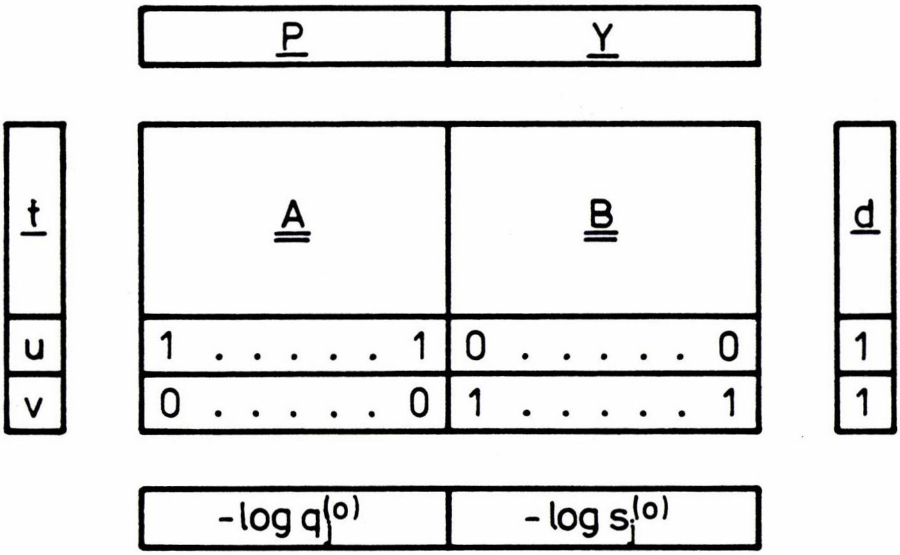
$$\sum_{j=1}^{n} p_j \log \frac{p_j}{q_j^{(0)}} + \sum_{j=1}^{k} r_j \log \frac{r_j}{s_j^{(0)}} \quad \text{minimalizálandó!}$$

A célfüggvény még az alábbi formában is írható

$$\sum_{j=1}^{n} p_j(-\log q_j^{(0)}) + \sum_{j=1}^{k} r_j(-\log s_j^{(0)}) + \log \frac{\prod_{j=1}^{n} (p_j)^{p_j}}{\left(\sum_{j=1}^{n} p_j\right)^{\sum_{j=1}^{n} p_j}} + $$
$$+ \log \frac{\prod_{j=1}^{k} (r_j)^{r_j}}{\left(\sum_{j=1}^{k} r_j\right)^{\sum_{j=1}^{k} r_j}}$$

Ez nem más mint egy geometriai programozási duálfeladat. A paramétereket sematikusan szemléltetve:

Legyenek a primálváltozók $\mathbf{t}$ m dimenziós vektor és u, v skalárok.



7. ábra

A primál feladat feltételi halmaza:

$$
(3.11) \qquad
\begin{aligned}
&\sum_{j=1}^{n} \exp(\mathbf{t}\mathbf{a}_j + u + \log q_j^{(0)}) \le 1 \\
&\sum_{j=1}^{k} \exp(\mathbf{t}\mathbf{b}_j + v + \log s_j^{(0)}) \le 1
\end{aligned}
$$

célfüggvénye pedig:

$$\mathbf{t}\mathbf{d} + u + v \quad \text{maximalizálandó!}$$

Azonnal látható, hogy a primál feladat *Slater reguláris.*

Az egyensúlyi feltételek:

$$(3.12) \qquad q_j^{(0)} \exp(\mathbf{t}\mathbf{a}_j + u) = p_j, \quad (j = 1, \ldots, n)$$

$$(3.13) \qquad s_j^{(0)} \exp(\mathbf{t}\mathbf{b}_j + v) = r_j, \quad (j = 1, \ldots, k)$$

1° A geometriai programozás gyenge dualitási tételét felhasználva az alábbiakat kapjuk:

Ha van $(\mathbf{p}, \mathbf{r}) \in \mathcal{L}$, hogy $(\mathbf{p}, \mathbf{r}) > 0$, akkor léteznek olyan $(\mathbf{p}^*, \mathbf{r}^*) \in \mathcal{L}$ és $\mathbf{t}^*, u^*, v^*$ primál megoldások, amelyek a (3.12) és (3.13) egyensúlyi feltételeket teljesítik, azaz

$$(3.14) \qquad q_j^{(0)} \exp(\mathbf{t}^*\mathbf{a}_j + u^*) = p_j^*, \quad (j = 1, \ldots, n)$$

és

$$s_j^{(0)} \exp(\mathbf{t}^*\mathbf{b}_j + v^*) = r_j^*, \quad (j = 1, \ldots, k) \tag{3.15}$$

és a céfüggvényértékekre

$$\sum_{j=1}^{n} p_j^* \log \frac{p_j^*}{q_j^{(0)}} + \sum_{j=1}^{k} r_j^* \log \frac{r_j^*}{s_j^{(0)}} = \mathbf{t}\mathbf{d} + u^* + v^*. \tag{3.16}$$

A (3.14) és (3.15)-ből adódik, hogy

$$u^* = -\log \sum_{j=1}^{n} q_j^{(0)} \exp(\mathbf{t}^*\mathbf{a}_j)$$

és

$$v^* = -\log \sum_{j=1}^{k} s_j^{(0)} \exp(\mathbf{t}^*\mathbf{b}_j).$$

Legyen

$$q_j^* := q_j^{(0)} \exp(\mathbf{t}\mathbf{a}_j + u^*) \quad (j = 1, \ldots, n)$$

és

$$s_j^* := s_j^{(0)} \exp(\mathbf{t}\mathbf{b}_j + v^*) \quad (j = 1, \ldots, k)$$

Ebből

$$\log \frac{q_j^*}{q_j^{(0)}} = \mathbf{t}^*\mathbf{a}_j + u^* \quad (j = 1, \ldots, n) \tag{3.17}$$

és

$$\log \frac{s_j^*}{s_j^{(0)}} = \mathbf{t}^*\mathbf{b}_j + v^* \quad (j = 1, \ldots, k) \tag{3.18}$$

(3.16)-ban $\mathbf{d}$ értékét visszaírva

$$\sum_{j=1}^{n} p_j^* \log \frac{p_j^*}{q_j^{(0)}} + \sum_{j=1}^{k} r_j^* \log \frac{r_j^*}{s_j^{(0)}} = \sum_{j=1}^{n} (\mathbf{t}^*\mathbf{a}_j + u^*) p_j^{(0)} + \\ + \sum_{j=1}^{k} (\mathbf{t}^*\mathbf{b}_j + v^*) r_j^{(0)}$$

Ebbe (3.17)-et és (3.18)-at visszaírva

$$\sum_{j=1}^{n} p_j^* \log \frac{p_j^*}{q_j^{(0)}} + \sum_{j=1}^{k} r_j^* \log \frac{r_j^*}{s_j^{(0)}} = \sum_{j=1}^{n} p_j^{(0)} \log \frac{p_j^{(0)}}{q_j^{(0)}} - \sum_{j=1}^{n} p_j^{(0)} \log \frac{p_j^{(0)}}{q_j^*} + $$
$$+ \sum_{j=1}^{k} r_j^{(0)} \log \frac{r_j^{(0)}}{s_j^{(0)}} - \sum_{j=1}^{k} r_j^{(0)} \log \frac{r_j^{(0)}}{s_j^*},$$

azaz

$$D(\mathbf{p}^{(0)}\|\mathbf{q}^{(0)}) + D(\mathbf{r}^{(0)}\|\mathbf{s}^{(0)}) = D(\mathbf{p}^{(0)}\|\mathbf{q}^*) + D(\mathbf{p}^*\|\mathbf{q}^{(0)}) + D(\mathbf{r}^{(0)}\|\mathbf{s}^*) + D(\mathbf{r}^*\|\mathbf{s}^{(0)}).$$

A lemma szerint, az egyenlőség miatt, ebben az esetben az alábbi két egyenlőség áll fenn:

$$D(\mathbf{p}^{(0)}\|\mathbf{q}^{(0)}) = D(\mathbf{p}^{(0)}\|\mathbf{q}^*) + D(\mathbf{p}^*\|\mathbf{q}^{(0)})$$

és

$$D(\mathbf{r}^{(0)}\|\mathbf{s}^{(0)}) = D(\mathbf{r}^{(0)}\|\mathbf{s}^*) + D(\mathbf{r}^*\|\mathbf{s}^{(0)})$$

2° A geometriai programozás erős dualitási tétele szerint létezik $(\mathbf{p}^*, \mathbf{r}^*) \in \mathcal{L}$ úgy, hogy

$$\sum_{j=1}^{n} p_j^* \log \frac{p_j^*}{q_j^{(0)}} + \sum_{j=1}^{k} r_j^* \log \frac{r_j^*}{s_j^{(0)}} = \sup_{(\mathbf{t},u,v)} (\mathbf{td} + u + v), \tag{3.19}$$

és a szupremumot nem csökkenti, ha (3.11)-ben egyenlőséget tételezünk fel, azaz, hogy

$$\begin{aligned} u &= -\log \sum_{j=1}^{n} q_j^{(0)} \exp(\mathbf{ta}_j) \\ v &= -\log \sum_{j=1}^{k} s_j^{(0)} \exp(\mathbf{tb}_j). \end{aligned} \tag{3.20}$$

Legyen

$$\begin{aligned} q_j &:= q_j^{(0)} \exp(\mathbf{ta}_j + u) \quad (j = 1, \ldots, n) \\ s_j &:= s_j^{(0)} \exp(\mathbf{tb}_j + v) \quad (j = 1, \ldots, k) \end{aligned} \tag{3.21}$$

Így (3.19) az alábbi formát ölti:

$$\sum_{j=1}^{n} p_j^* \log \frac{p_j^*}{q_j^{(0)}} + \sum_{j=1}^{k} r_j^* \log \frac{r_j^*}{s_j^{(0)}} = \sup_{(\mathbf{q},\mathbf{s})} \left(\sum_{j=1}^{n} p_j^{(0)} \log \frac{q_j}{q_j^{(0)}} + \sum_{j=1}^{k} r_j^{(0)} \log \frac{s_j}{s_j^{(0)}} \right)$$

és $(\mathbf{q},\mathbf{s})$ a (3.20), (3.21) feltételeket teljesítik. Mivel a (3.20), (3.21) feltételek az $\mathcal{E}$ halmazt definiálják és $\mathcal{E}$ része a szimplexek *Descartes-szorzatának*, továbbá a log függvény folytonossága miatt:

$$\sup_{(\mathbf{q},\mathbf{s})\in\mathcal{E}}\left(\sum_{j=1}^{n} p_j^{(0)}\log\frac{q_j}{q_j^{(0)}}+\sum_{j=1}^{k} r_j^{(0)}\log\frac{s_j}{s_j^{(0)}}\right)=$$
$$=\max_{(\mathbf{q},\mathbf{s})\in \mathrm{cl}\mathcal{E}}\left(\sum_{j=1}^{n} p_j^{(0)}\log\frac{q_j}{q_j^{(0)}}+\sum_{j=1}^{k} r_j^{(0)}\log\frac{s_j}{s_j^{(0)}}\right).$$

Azonban

$$\max_{(\mathbf{q},\mathbf{s})\in \mathrm{cl}\mathcal{E}}\left(\sum_{j=1}^{n} p_j^{(0)}\log\frac{q_j}{q_j^{(0)}}+\sum_{j=1}^{k} r_j^{(0)}\log\frac{s_j}{s_j^{(0)}}\right)=$$
$$=\sum_{j=1}^{n} p_j^{(0)}\log\frac{p_j^{(0)}}{q_j^{(0)}}+\sum_{j=1}^{n} r_j^{(0)}\log\frac{r_j^{(0)}}{s_j^{(0)}}-$$
$$-\min_{(\mathbf{q},\mathbf{s})\in \mathrm{cl}\mathcal{E}}\left(\sum_{j=1}^{n} p_j^{(0)}\log\frac{p_j^{(0)}}{q_j}+\sum_{j=1}^{k} r_j^{(0)}\log\frac{r_j^{(0)}}{s_j}\right).$$

Legyen $(\mathbf{q}^*,\mathbf{s}^*)\in \mathrm{cl}\mathcal{E}$ olyan, hogy

$$\min_{(\mathbf{q},\mathbf{s})\in \mathrm{cl}\mathcal{E}}\left(\sum_{j=1}^{n} p_j^{(0)}\log\frac{p_j^{(0)}}{q_j},\sum_{j=1}^{k} s_j^{(0)}\log\frac{s_j^{(0)}}{r_j}\right)=\sum_{j=1}^{n} p_j^{(0)}\log\frac{p_j^{(0)}}{q_j^*}+\sum_{j=1}^{k} r_j^{(0)}\log\frac{r_j^{(0)}}{s_j^*}.$$

Így kapjuk, hogy

$$D(\mathbf{p}^{(0)}\|\mathbf{q}^{(0)})+D(\mathbf{r}^{(0)}\|\mathbf{s}^{(0)})=D(\mathbf{p}^{(0)}\|\mathbf{q}^*)+D(\mathbf{p}^*\|\mathbf{q}^{(0)})+D(\mathbf{r}^{(0)}\|\mathbf{s}^*)+D(\mathbf{r}^*\|\mathbf{s}^{(0)}).$$

De a lemma miatt (3.19) és (3.20) is fenn áll. □

3.4. KÖVETKEZMÉNY. (Erős equilibrium).
Ha a $(\mathbf{p}^*,\mathbf{r}^*)\in\mathcal{L}$ *az MDI feladat optimális megoldása és a* $(\mathbf{q}^*,\mathbf{s}^*)\in \mathrm{cl}\mathcal{E}$ *az ML feladat optimális megoldása, (azaz* $\{\inf D(\mathbf{p}^{(0)}\|\mathbf{q})+D(\mathbf{r}^{(0)}\ \|\ \mathbf{s})|(\mathbf{q},\mathbf{s})\in\mathcal{E}\}=D(\mathbf{p}^{(0)}\|\mathbf{q}^*)+D(r^{(0)}\|\mathbf{s}^*)$, $(\mathbf{q}^*,\mathbf{s}^*)\in \mathrm{cl}\mathcal{E}$*), akkor a (3.19) és (3.20) Pitagorasz-összefüggések fennállnak.*

Bizonyítás. Legyen $(\bar{\mathbf{p}},\bar{\mathbf{r}})$ és $(\bar{\mathbf{q}},\bar{\mathbf{s}})$ a tétel által biztosított optimális megoldások. Ezekre teljesül, hogy:

$$D(\bar{\mathbf{p}}\|\mathbf{q}^{(0)})+D(\mathbf{p}^{(0)}\|\bar{\mathbf{q}})=D(\mathbf{p}^{(0)}\|\mathbf{q}^{(0)}),\tag{3.22}$$
$$D(\bar{\mathbf{r}}\|\mathbf{s}^{(0)})+D(\mathbf{r}^{(0)}\|\bar{\mathbf{s}})=D(\mathbf{r}^{(0)}\|\mathbf{s}^{(0)}),$$

Minthogy $(\mathbf{p}^*, \mathbf{r}^*)$ is és $(\mathbf{q}^*, \mathbf{s}^*)$ is optimálisak így

$$
\begin{aligned}
D(\mathbf{p}^*\|\mathbf{q}^{(0)}) &= D(\bar{\mathbf{p}}\|\mathbf{q}^{(0)}),\\
D(\mathbf{p}^{(0)}\|\mathbf{q}^*) &= D(\mathbf{p}^{(0)}\|\bar{\mathbf{q}}),\\
D(\mathbf{r}^*\|\mathbf{s}^{(0)}) &= D(\bar{\mathbf{s}}\|\mathbf{s}^{(0)}),\\
D(\mathbf{r}^{(0)}\|\mathbf{s}^*) &= D(\mathbf{r}^{(0)}\|\bar{\mathbf{s}}).
\end{aligned}
\tag{3.23}
$$

A (3.23)-at a (3.22)-be visszahelyettesítve azt kapjuk, hogy

$$D(\mathbf{p}^*\|\mathbf{q}^{(0)}) + D(\mathbf{p}^{(0)}\|\mathbf{q}^*) = D(\mathbf{p}^{(0)}\|\mathbf{q}^{(0)})$$

és

$$D(\mathbf{r}^*\|\mathbf{s}^{(0)}) + D(\mathbf{r}^{(0)}\|\mathbf{s}^*) = D(\mathbf{r}^{(0)}\|\mathbf{s}^{(0)}) \qquad \square$$

Végezetül megjegyezzük, hogy a módszer tetszőleges számú blokkra általánosítható.

IRODALOM

[1] CSENCOV, N. N., *Statisztikai Döntési Szabályozók és Optimális Következetetések*, (orosz nyelven) (Nauka, Moszkva, 1972).

[2] CSISZÁR, I., "A minimális diszkrimináló információ módszere; kontigenciatáblázatok elemzése", *Móri, F. T.-Székely, J. G., Többváltozós statisztikai analízis* (Műszaki Könyvkiadó, Budapest, 1986, XII. fejezet).

[3] GOKHALE-KULLBACK, *The Information in Contingency Tables* (Marcel Dekker, New York, 1978).

[4] KLAFSZKY, E., "Az input-output tábla előrebecsléséről", *MTA SZTAKI Közlemények* **10** (1973), 3–13.

[5] KLAFSZKY, E., "Geometriai Programozás és néhány alkalmazása", *MTA SZTAKI Tanulmányok* **8** (1973).

[6] KLAFSZKY, E., "A theoretical prediction of the input-output tables", in Lecture Notes in Computer Science No. 4. (Springer-Verlag, 1974), pp. 484–492.

[7] KULLBACK-LEIBLER, "On information and sufficiency", *Ann. Math. Statist.* **22** (1951), 79–86.

[8] RÉNYI, A., "Az információelmélet néhány alapvető kérdése", *MTA III. Osztály Közleményei* (1960), 251-282.

(Beérkezett: 1989. május 30.)

LORDES GONZALEZ BENITEZ, ASPIRÁNS
NME MATEMATIKAI INTÉZET, SZÁMÍTECHNIKAI TANSZÉK
3515 MISKOLC
INSTITUTO SUPERIOR TÉCHNICS
GAVETA PORTAL 57, HOLQUIN
CUBA

ANALYSIS OF THE DUALITY BETWEEN LINEAR AND LOGLINEAR FAMILIES OF DISTRIBUTIONS BY GEOMETRIC PROGRAMMING

L. GONZÁLEZ

Given a family of probability distributions defined by linear internal constraints and another family defined by loglinear external constraints (See CSISZÁR [2], we analyse the duality between the two families by applying geometric programming. This allows us to generalize CSISZÁR's results.

The problem may be considered as the generalization of the method of calibration of input-output tables (see [1], [3], [4], [6]).

Alkalmazott Matematikai Lapok 15 (1990–91) 405–418

PÁRHUZAMOS SZÁMÍTÓGÉPEK FELÉPÍTÉSE

DEÁK ISTVÁN

Budapest

A több processzort használó, párhuzamos működést lehetővé tevő számítógépek, valamint az ezekkel szoros kapcsolatban álló, legújabb felépítési elvek szerint működő számítógépek témakörét tekintjük át a felépítési és működési elvek szerint. A történelmi előzmények és az alapfogalmak összefoglalása után az ilyen számítógépek felépítésének problémáit, valamint a hatékonyság mérésének kérdését vizsgáljuk meg. Rövid leírásukkal együtt felsoroljuk a jelenleg kapható legfontosabb gépeket.

1. Bevezetés

A félvezetőipar és a számítógépgyártás folyamatosan csökkenti a számítógépek alkatrészeinek méretét és ezzel növeli az ezekből előállítható számítógépek sebességét. A felhasználók azonban mindig többet kívánnak. Az egy processzorral működő számítógépek fejlődési lehetőségének határai immár világosan láthatók. A kiutat részben új felépítési és működési elvek alkalmazása, részben pedig a felhasznált processzorok számának növelése jelentette.

Az 1940–es évek végén kifejlesztett számítógépek működésének alapelve — egy processzor sorban végrehajtja a tárban elhelyezett programutasításokat — az 1970–es évek végéig egyeduralkodó volt. Ezt a fajta felépítési elvet *Neumann rendszernek* vagy *működési elvnek* is nevezik. Az 1960–as években kezdtek el foglalkozni olyan számítógépek kifejlesztésével, amelyek több processzorral vannak felszerelve, egyszerre több utasítást is végre tudnak hajtani, párhuzamos feldolgozásra alkalmasak, illetőleg olyan módosításokat hajtottak végre a számítógép felépítésén, hogy bizonyos fajta műveleteket egyszerre lehessen elvégezni (például Bemenet/Kimenet műveletek és aritmetikai műveletek).

A párhuzamos számítógépek (*parallel computers*) mellett néha a szuperszámítógépek (*supercomputers*) elnevezést is használják ezekre a gépekre, bár a leggyorsabb gépek nem mind párhuzamos működési elvűek (viszont a párhuzamos számítógépek mind szuperszámítógépnek tekinthetők). Szoros kapcsolatuk miatt mindkét típust tárgyaljuk, de a hangsúlyt a párhuzamos számítógépekre, azok felépítésének és működésének elveire helyezzük.

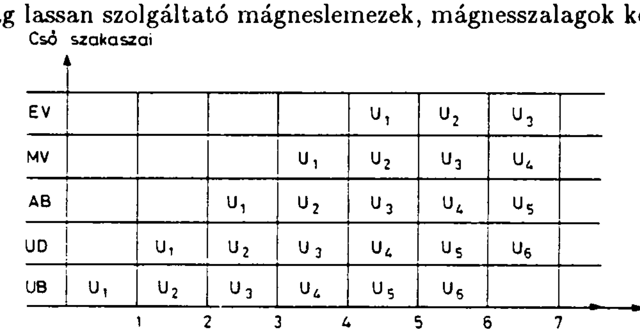
A párhuzamos számítógépek felépítésével foglalkozó általános művek közül a [7], [11], [12], [13] és [14] műveket ajánljuk az érdeklődőknek, a programozással kapcsolatos kérdésekkel foglalkozó általános témájúak közül a [3], [7] és [14] könyveket. Egy speciális területtel, a szimulációval foglalkozó művek közül a [2], [4], [5] és [6] ajánlható, valamint a [4] cikkben található harmincnál több hivatkozás.

2. Alapfogalmak

Rendkívül sokféle felépítési elv jellemzi a párhuzamos számítógépeket, ezért nagyon nehéz egy minden gépre kiterjedő, egyértelmű és jellemző osztályozást megadni. A leggyakrabban használt kategorizálás FLYNN [8] nevéhez fűződik. A SISD (*Single Instruction, Single Data stream* kezdőbetűiből) típusú gépek az egy utasításfolyam — egy adatfolyam elve szerint működnek, vagyis ezek a hagyományos számítógépek. A SIMD (*Single Instruction, Multiple Data*) gépeken egy utasításfolyam van és többszörös adatfolyam, vagyis egy utasítást egyszerre több adaton hajtanak végre. Az MISD géptípus (*Multiple Instruction, Single Data*) gyakorlatilag nem létezik; elvileg ez egy olyan gép lenne, amely több utasítást hajt végre egyszerre egy adaton. Az MIMD (*Multiple Instruction, Multiple Data*) számítógép több utasításfolyamot hajt végre párhuzamosan, egyidejűleg több adatfolyamon dolgozva.

Egy másik felosztás szerint a gépeket a beépített processzorok száma szerint három csoportba soroljuk. Eszerint vannak egyprocesszoros gépek, egészen kevés (2–32) processzorral rendelkező gépek és nagymértékben párhuzamos gépek, amelyek száznál több, esetleg több ezer processzorral dolgoznak.

Egyprocesszoros rendszerben is meg lehet valósítani bizonyos mértékű párhuzamos működést. Így az a számítógép, amely az aritmetikai processzortól függetlenül olvashat egy mágneslemezről vagy szalagról, már párhuzamos működésűnek tekinthető. Hierarchikus memória felépítés esetén is van egy bizonyos párhuzamos működés; ilyenkor a processzor speciális regisztereket, vagy ú.n. gyors tárat (*cache memory*) ér el, amelyekből nagyon gyorsan lehet adatot olvasni vagy írni, míg ezeket a regisztereket vagy gyors tárat a fő tárból látják el adatokkal (a fő tárat pedig mágneslemezekről, esetleg szalagokról). Az adatátvitel a memória különböző szintjei között egymástól függetlenül egyszerre is megvalósulhat. Ezek a felsorolt sorrendben egyre nagyobb kapacitású, de egyre lassabban elérhető tárak segítenek kitölteni azt az űrt, amely a gyors processzor adatigénye és a nagymennyiségű adatot viszonylag lassan szolgáltató mágneslemezek, mágnesszalagok között van.



1. ábra

Öt alegységgel rendelkező „csövesített" processzor működésének vázlata

Egy alapvető fontosságú felépítési elv a cső vagy másnéven szerelőszalag (*pipe-*

line), amelyet egyre szélesebb körben alkalmaznak még egyprocesszoros gépeknél is. Egy processzor esetén ennek az elvnek az alkalmazása abban áll, hogy az elvégzendő utasítást több részre bontják fel, és a különböző részeket a processzor különböző alegységei hajtják végre, egyidejűleg és egymástól függetlenül. Bontsunk fel minden utasítást például öt részre, amelyekről tegyük fel, hogy egyenlő hosszú idő, például egy ciklusidő kell a végrehajtásukhoz: 1. utasítás beolvasása UB, 2. utasítás dekódolása UD, 3. adat beolvasása AB, 4. művelet végrehajtása MV, 5. eredmény visszaírása EV (lásd 1. ábrát). Egy csövesített processzorban öt különböző alegység van, amelyek csak egy-egy részt hajtanak végre a teljes utasításból, ahhoz hasonlóan, ahogy a munkások dolgoznak a futószalag mellett. Tehát a processzor első alegysége egy utasítás végrehajtásából csak annyit vállal, hogy azt kiolvassa a tárból, a többi alegység ezalatt tétlenül várakozik. Mialatt az első alegység a második utasítást olvassa ki a tárból, a második alegység az első utasítás dekódolását végzi. A harmadik időegység alatt a processzor első alegysége a harmadik utasítás beolvasását végzi, a második alegység a második utasítást dekódolja, a harmadik alegység pedig az első utasítás végrehajtásához szükséges adatokat olvassa be a tárból és így tovább.

Egy nem csövesített processzor csak sorban tudja elvégezni az első utasítás öt részét, majd a második utasítás öt részét stb., így öt ciklusidőnként fejez be egy utasítást. A csövesített változat egy ciklusidőnként fejez be egy utasítást (feltéve, hogy az utasításcső már fel van töltve utasításokkal). Így elméletileg a K szakaszos cső K-szor gyorsabb a közönséges processzornál. Ezt az elméletileg elérhető gyorsítást a gyakorlatban csak ritkán lehet megközelíteni és egyébként is függ a konkrét programban található utasításoktól. Ugyanis ha egy IF utasítás végrehajtása után egy másik utasítássorozatra tevődik át a végrehajtás, nem a már a csőbe betöltött utasítássorozatra, akkor meg kell várni, amíg a teljes cső kiürül (kitörlik a feleslegesen betöltött utasítássorozatot, és a másik utasítássorozatot kezdik el betölteni), ami K ciklusideig tart. Ilyen okok miatt a gyakorlatban használatos gépek 4–6 szakaszos csővel dolgoznak, bár létezik 26 szakaszos cső is (*Cyber*-205). A csövesítés elvét itt processzorra írtuk le, de hasonlóan alkalmazható például Bemenet/Kimenet műveleteknél is.

A vektor szuperszámítógépeket egy (vagy több) olyan csővel szerelik fel, amelyik vektorokon tud nagyon gyorsan műveleteket végrehajtani. A cső előbb tárgyalt feltöltési tulajdonsága miatt a hosszú vektorok esetén nagyobb a műveleti sebessége a csőnek, hiszen a feltöltés és a kiürítés adminisztrálása minden vektor esetén ugyanannnyi, de a nagy elemszámú vektoroknál ez kisebb hányadot képvisel a teljes műveleti időből.

A cső felépítési elvét, műveletek egymással való átfedését nemcsak utasítások, aritmetikai műveletek végrehajtásánál alkalmazzák, hasonló elvek alapján lehet szervezni például a Bemenet/Kimenet műveleteket is. A felhasznált adatok természete szerint beszélhetünk vektor és skalár csövekről, ezeken belül megkülönböztetünk aritmetikai, logikai, processzor csöveket. A műveleti sebesség növelése érdekében egy számítógép gyakran többféle csövet is tartalmaz.

3. Felépítési elvek

A jelenleg ismert párhuzamos számítógépek felépítése igen sokféle, szinte gépről-gépre változik. Ebben a szakaszban a leggyakrabban alkalmazott vagy valamilyen szempontból fontosnak tartott felépítési elveket és lehetőségeket foglaljuk össze.

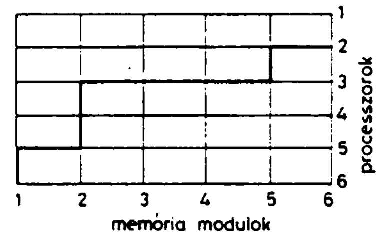
Több processzor alkalmazását feltételezve a memória elérése szerint beszélhetünk kizárólagos memóriáról, vagy közös memóriáról. Az első esetben minden egyes processzorhoz hozzárendelik az általa használható tárterületet, vagy tárakat; az adott processzor a hozzárendelt tárakat kizárólagos hozzáférési joggal használja, más processzorok nem férhetnek ezekhez hozzá. Ebben az esetben a számítógép egy processzorból és a hozzátartozó memóriából álló részekre bontható fel, amelyek egymástól lényegében függetlenül működnek (a kapcsolattartás az összekapcsoló hálózaton továbbított üzenetek formájában valósítható meg). A második esetben, a minden processzor által elérhető közös memória esetében, az ide beírt információk alapján nyílik lehetőség a processzorok közötti együttműködésre. Ekkor viszont szinkronizálási problémák léphetnek fel, továbbá a processzorok és a memória közé kapcsoló egységet (switch) kell beépíteni. (Ilyen kapcsoló egységet kell néha a különböző elérési idejű tárak közé is beépíteni, hiszen a nagyon gyors processzoroknak nagyságrenddel kisebb a ciklusideje, mint a főtárnak, így regiszterek és gyorstárak főtárból való feltöltésére is szükségünk van, ami a kapcsoló hálózaton keresztül történik.)

Szinkronizálási ellentét lép fel, ha ugyanazt az adatot két vagy több processzor akarja egyszerre kiolvasni, felülírni, vagy ugyanazt a memóriamodult elérni.

Jónéhány számítógépen vegyes memóriaelérési stratégiát követnek: minden processzornak van egy kizárólagos tárterülete és ezenkívül van egy nagy közös tár is, amelyet minden processzor elérhet. Memóriakonfliktusok elkerülésére az is egyfajta megoldás lehet, hogy ugyanabból az adatból több (vagy minden) memóriamodulban tárolnak egy példányt — ilyenkor persze az esetleges felülírások miatt az adatok időbeliségét, illetőleg egymásutániságát kell valamilyen módon szinkronozálni és ellenőrizni.

A *Cyber*–205 gépnél a vektorműveletek elvégzésénél várható memóriakonfliktusokat csökkentik azzal, hogy egy vektor elemeit nem folyamatosan tárolják a memóriában, hanem egy konstans számú szóval eltolva, így az egymásutáni elemek külön memóriamodulokba kerülnek (*interleaving*) és ezzel egyszerre elérhetőkké válnak.

A memória és a processzorok (illetőleg a főtár és a gyorstár) között többféle módon lehet kapcsolatot létesíteni. Rákapcsolható egy közös sínre egyik oldalról az összes processzor, másik oldalról a teljes memória. Ilyen felépítésnél nagyon gyakran felléphet adatigények közötti konfliktus, amelynek megoldására többféle algoritmus használható. Legegyszerűbb a legrégebben várakozó igény kielégítése (*First In, First Out* — FIFO — statikus használat), de teljesíthetjük annak az egységnek a kérését a leggyorsabban, amelynek legrégebben volt igénye (*Least Recently Used* — LRU — dinamikus kijelölés), vagy választhatunk véletlenszerűen az igények közül.



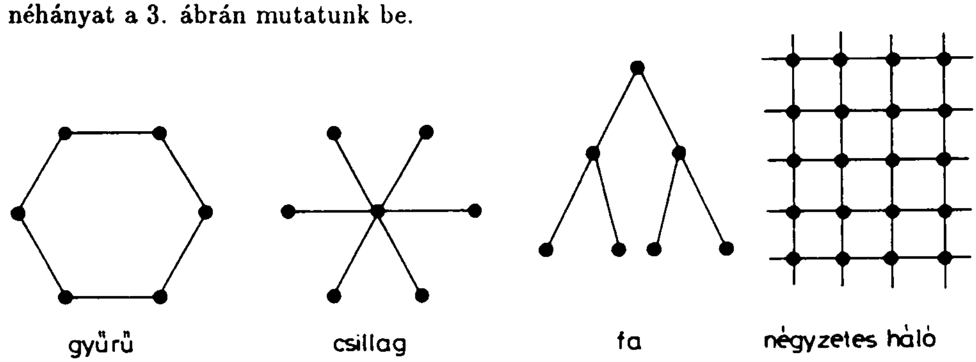
2. ábra

Egy üzenet útja keresztrudas kapcsolásban (vastag vonal)

Egy gyakran használt változat a keresztrudas kapcsolás (*crossbar switch*, lásd a 2. ábrát), ahol akármelyik processzor elérheti akármelyik memóriamodult. A keresztrudas kapcsolás egy csomópontjának képesnek kell lennie arra, hogy egy akármelyik processzortól származó üzenetet akármelyik modulnak továbbítson, továbbá egy olyan részének is kell lennie, amely az egyidejűleg beérkező kéréseket rangsorolja. Persze ilyenkor is kialakulhatnak „forró pontok" az összekötő hálózatban, vagyis olyan csomópontok, amelyeket gyakran igényel több üzenet is az útvonalához. Sok processzor és sok modul esetén egy ilyen rendszer kiépítése elég költségesssé válik.

Ha a keresztrudas összekötő hálózat csomópontjaiban lévő ellenőrző, kapcsoló és elsőbbséget eldöntő funkciókat az egyes memóriamodulokhoz telepítjük, akkor az úgynevezett *multiport rendszert* kapjuk.

A processzorokat gyakran egymással is összekötik, hogy üzenetet lehessen küldeni a tár kikerülésével. Ezek az összekötő hálózatok lehetnek statikusak és dinamikusak (egyes kapcsolók állításával rekonfigurálhatók). A statikus hálózatok közül néhányat a 3. ábrán mutatunk be.



3. ábra

Processzorok összekapcsolásának változatai

Az újabban egyre népszerűbb hálózat az n–dimenziós kocka, amelyen 2^n számú csúcs van. Egy csúcsot egy $(b_1, b_2, \ldots, b_n)$, $b_i = 0$ vagy 1, szám n–essel jelölünk. Egy adott csúcs azokkal a csúcsokkal szomszédos közvetlenül, amelyek szám n–ese pontosan egy bitben különbözik az adott csúcs szám n–esétől. Egy csúcsból egy másik csúcsba küldendő üzenetnek ekkor legfeljebb $n - 1$ másik csúcson kell áthaladnia. Újabban a kukaclyukas (*wormhole routing*) üzenettovábbítást is alkalmazzák, ekkor akármelyik csúcsból bármely másikba azonos idő alatt lehet üzenetet küldeni (ilyen üzenettovábbítás van az iPSC/2 gépen is).

Számítógépek tervezésénél az összekötő hálózat kiválasztásának megítélésében több szempont kap szerepet; a gyakorlatban egy, a teljesítmény (sávszélesség, vagyis a másodpercenként átvihető adat mennyisége) és az ár (a hálózat felépítésének bonyolultsága) közötti kompromisszumot keresnek. Az összekötő hálózat árát elsősorban a hálózat bonyolultsága határozza meg, de ez az ár a technológia fejlődésével csökkenő irányzatot mutat.

A sávszélesség maghatározása azért nehéz, mert a konfliktusok arányát és az átvitel gyorsaságát meghatározó memóriaigények mennyisége és fajtája a programok előre nem jelezhető viselkedésétől függ. Szokás a sávszélességet egyes műveletek esetén elérhető átvihető adatmennyiségnek definiálni, pl. egy mátrixnak vektorral való szorzása esetén a sávszélesség 80 Mbyte.

Itt említjük meg, hogy létezik egy különleges típusú memória, az asszociatív tár, amely a tartalomtól függően címezhető, vagyis a teljes memória keresése egyszerre történik. Így egy utasításidő alatt kikereshető az összes olyan cím, amely egy adott számot tartalmaz. Ilyen típusú tárakra gyorsan változó adatbázisok esetén lehet szükség, például radarjel követése, légtérfigyelés, képfeldolgozás, mesterséges intelligencia modellezés. Asszociatív tárral szerelték fel a *Goodyear Aerospace* STARAN gépét és a *Parallel Processing Ensemble* (PEPE) gépet. Az ilyen típusú memóriák rendkívül drágák.

A több processzort használó rendszereket tekinthetjük *multiprocesszoros gépnek* is (vagyis olyan gépnek, amelyikben több processzor dolgozik együtt), *multiszámítógépnek* is (vagyis több számítógép együttesének). Az első csoportba tartoznak a processzor tömbök, például a *Connection Machine*, míg a másodikba sorolhatók az iPSC *Hypercube* géptípusok. A megkülönböztetés alapja a processzorok közötti együttműködés szorossága, a szinkronizáltság, az üzenetek gyakorisága, az erőforrások közös felhasználása, stb. Ezért a gépek besorolása nem mindig egyértelmű. A multiprocesszoros rendszert — amely esetében a processzorok által végrehajtott műveletek azonosak — szorosan kapcsolt rendszernek, míg a multiszámítógépes rendszert, ahol az egyes processzorok egymástól függetlenül dolgozhatnak, lazán kapcsolt vagy elosztott rendszernek nevezzük (egyesek ide sorolják a számítógép hálózatokat is).

4. Teljesítmény, sebesség, hatékonyság

A számítógépek egyik legfontosabb jellemzője a teljesítmény, amelyet az időegység alatt elvégzett műveletek számával szoktak mérni. A különböző műveletek elvégzéséhez szükséges idő viszont esetenként igen eltérő is lehet. Ezért a teljesítményt többféleképpen szokták mérni. Ilyen lehetőségek a következők: (i) használnak olyan mintaprogramokat, amelyek egyes jól meghatározott feladatokat hajtanak végre (*benchmark*), (ii) egy bizonyos műveletre adják meg a teljesítményt, lebegőpontos összeadásra vagy szorzásra például, (iii) különböző műveletek végrehajtási idejének súlyozott összegét számítják (ún. keverékeket — *mix* — képeznek). Ezért a számítógépek teljesítményére vonatkozó számokat csak bizonyos óvatossággal lehet összehasonlítani.

A teljesítmény mértéke lehet az egy másodperc alatt végrehajtott utasítások száma (egysége a Mip/s — *million instructions per second*, vagyis egymillió utasítás másodpercenként, esetleg a Gip/s = 1000 Mip/s) vagy pedig az egy másodperc alatt végrehajtott lebegőpontos műveletek száma (egysége a Mflop/s — *million floating point operations per second*, vagyis egymillió lebegőpontos művelet másodpercenként, nagyobb egysége a Gflop/s = 1000 Mflop/s).

A párhuzamos számítógépek teljesítményének van egy elméleti maximuma, amit r_∞-nel jelölünk [11], és amit akkor érhet el a számítógép, ha minden processzora állandóan működik. A gyakorlatban ezt sohasem érik el, hiszen a számítógépben lévő csövek feltöltésének és kiürülésének ideje, vagy másmilyen adminisztrációs, esetleg üzenettovábbítási idők a számításokat inproduktív időként terhelik. Használják még az $n_{1/2}$ jelű paramétert: az $n_{1/2}$ hosszúságú vektorok esetén a számítógép az elméletileg lehetséges r_∞ teljesítőképességnek a felét éri el. Így az n hosszúságú vektor esetén a vektorművelet végrehajtási ideje

$$t = (n + n_{1/2})/r_\infty \ .$$

A gyorsítás, amelyet egy p processzort használó párhuzamos algoritmus alkalmazásával érünk el, egy olyan hányados, amelynek számlálója ezen párhuzamos számítógép által végrehajtott leggyorsabb soros program lefuttatásához szükséges idő, míg nevezője az ugyanezen párhuzamos számítógép p processzora által végrehajtott párhuzamos program ideje [14]. A p processzort használó párhuzamos algoritmus hatékonysága a gyorsítás osztva p-vel. Így, ha a soros programot 10 másodperc alatt hajtja végre egy 4 processzoros gép, a párhuzamos algoritmust pedig 4 másodperc alatt, akkor a gyorsítás 2,5-szeres, a hatékonyság 0,625.

Az elérhető gyorsítás lényegesen függ az algoritmus nem párhuzamosítható részének arányától. Ezt az összefüggést *Amdahl törvénye* írja le a következőképpen [1]:

$$S \leq \frac{1}{f + (1-f)/p}$$

ahol S a gyorsítás, f az algoritmus nem párhuzamosítható részének aránya és p a párhuzamos gépen használt processzorok száma. Egy p processzorral működő gép hatékonyságát 0,5-nek feltételezve $S = p/2$ adódik, amelyből

$$p/2 \leq \frac{1}{f + (1-f)/p} \quad \Longrightarrow \quad f \leq \frac{1}{p-1} \ .$$

Tehát egy száz processzoros gépnél ötven százalékos hatékonyságot feltételezve az algoritmus soros része mindössze 1 százalék lehet. Egy kevésbé borulátó álláspont szerint [10] AMDAHL kiindulási feltételei a valóságban nem teljesülnek, mert $1-f$, vagyis az algoritmus párhuzamosítható részének aránya nem független p-től. Másszóval a végzendő munka mérete — és a soros rész aránya — változik a processzorok számával, hiszen ha több processzort lehet használni, akkor a felhasználó általában kiterjeszti a probléma méreteit a lehetőségek határáig (a rácsfelbontás finomsága, iterációs lépések száma, a végeredmény pontossága, stb. állítható be megfelelően).

5. A legfontosabb párhuzamos számítógépek

Jelenleg hatvan felett van a kereskedelemben megvehető, párhuzamos működésre képes számítógép típusok száma [12], [13], de legalább még egyszer ennyi az olyan számítógépek száma, amelyek egyetemeken és kutatóintézetekben működnek, és csak néhány darab készült el belőlük. Ezért a teljesség igénye nélkül mindössze arra vállalkozunk, hogy a valamilyen szempontból fontos (leggyorsabb, elterjedt, újszerű felépítésű, stb.) gépekről adunk rövid áttekintést.

5.1 A CRAY számítógépek

A CRAY-1 nagyon elterjedt típus, az első példányt 1976-ban adták át. A gép órajele 12,5 nsec–os, az 1 M 64 bites szóból álló memóriája 16 csoportra van osztva. A különböző műveleteket (logikai, egész, lebegőpontos, címmanipuláció, vektorműveletek, skalárműveletek) egymástól függetlenül 12 funkcionális egység végzi. Ezek mindegyike regiszterekkel van ellátva és csövesítve van. A regiszterekbe olvassák be a viszonylag lassú memóriából az adatokat, hogy a funkcionális egységeket megfelelő sebességgel tudják adatokkal ellátni. A funkcionális egységeket egymáshoz kapcsolva is lehet csövesíteni (összeláncolás — *chaining*), vagyis ha egy művelet eredménye a következő művelet végrehajtásához kell, akkor az eredményt nem a memóriába, hanem egy regiszterbe írjuk vissza, ahonnan gyorsabban lehet kiolvasni. Mivel a regiszterek (összesen 8 darab van) 64 elemet képesek tárolni, ennél hosszabb vektorokat a műveletek végrahajtásához 64 eleműekre kell széttördelni.

A gyakorlati mérések szerint $n_{1/2}$ értéke 7 és 20 között változik a szokásos próbaprogramok esetén [11]. Ez azt jelenti, hogy a gép nem reagál túlzott teljesítménycsökkenéssel a rövid vektorokkal számoló programokra sem, ezért a CRAY-1 nagyon jól használható nem kifejezetten vektorokat használó programok esetén is. Az r_∞ elméletileg elérhető maximális sebesség (ha a lebegőpontos összeadás és a szorzás csövek egyidejűleg, összeláncolva működnek) a 160 Mflop/sec -hoz van közel. A gyakorlatban a regiszterek és a memória között sávszélesség szűkössége miatt — megfelelő program esetén — a 20–30 Mflop/sec már jó sebességnek számít.

A gép lényegében véve SIMD gép, hiszen egyetlen program fut, bár annak különböző részei párhuzamosan is működhetnek, ha a részek között nincs összefüggés és memóriakonfliktus.

A gép egyik továbbfejlesztése a CRAY–XMP, ami lényegében két, közös memóriából dolgozó CRAY-1 gépet jelent, amely 9,5 ns órajellel dolgozik (1983). Minden egyes processzorhoz négy adatutat építettek ki a memóriától, így a vektor csöveket nehézség nélkül lehet utántölteni (egy $\mathbf{x} = \mathbf{y} + \mathbf{z}$ vektor művelet elvégzéséhez három adatutra van szükség — kettőre az $\mathbf{y}$ és $\mathbf{z}$ kiolvasásához és egyre az $\mathbf{x}$ visszaírásához). Memóriakonfliktusok miatt (ugyanazt a memóriarészt két kérés éri egyszerre) még mindig lassulhat a program végrehajtása. Ezen a gépen standard próbaprogramok felhasználása esetén — meghatározott arányú aritmetikai művelet, ki és beolvasás, stb. — $n_{1/2} \sim 50 - 60$ volt, $r_\infty = 420$ Mflop/s.

Egy újabb típus a CRAY-2, amely négy processzorral 256 M szót használ, 4 ns-os órajellel (1985). Ennek a gépnek az elméleti sebessége 2 Gflop/s. A CRAY-2 mérete igen kicsiny a folyékony hélium hűtésnek köszönhetően, egy méter átmérőjű és 70 cm magas henger, amelyben 40 cm-nél hosszabb vezeték nincsen, ez a kis fizikai méret nyilván egyik oka a nagy sebességének.

A CRAY–XMP gép utódjának tekinthető a CRAY–YMP (1988. február), amelynek elméletileg elérhető legnagyobb sebessége 3 Gflop/s, ezt nyolc processzoros kiépítettség esetén érhetik el.

Mind a CRAY–XMP, mind a CRAY-2 MIMD gépnek tekintendő, mivel a processzorok különböző utasításfolyamatot hajtanak végre. Természetesen a működő példányok sokszor a fentebb leírt alapgépekkel nem teljesen azonos felépítésűek, hiszen az egyes kiszállított gépeket a vevő kívánsága szerint kissé módosíthatják.

5.2. Az FPS-164

A hozzákapcsolt számítógépek típusának jellegzetes képviselője az FPS-164 (*Floating Point System*) amellyel általában alacsony költséggel nagy mennyiségű vektor vagy mátrix számítást lehet elvégezni. Technikailag az ilyen hozzákapcsolt gépeket egy önállóan működő géppel (VAX-11 sorozat vagy IBM 308X sorozat) kötik össze, s míg az önállóan működő főgép gyakorolja a teljes rendszer feletti ellenőrzést, a bemenet/kimenet csatornák és a perifériák ellenőrzését, addig az FPS-164 vagy más hozzákapcsolt gép végzi a nagymennyiségű lebegőpontos számítást. Egy ilyen munkamegosztás akár 100-szoros felgyorsítást is jelenthet egy mini sebességéhez képest.

Az FPS-164 egy 185 ns órajellel és 64 bit szóhosszúsággal dolgozó gép, melynek 16 M szó memóriája van. Nyolc, egymástól független csővezetéke van a gépnek, ezek között van egy lebegőpontos szorzó és egy összeadó, valamint egy egész értékűekkel számoló aritmetikai-logikai egység, továbbá többféle memória. A gép elméletileg elérhető legnagyobb sebessége 11 Mflop/s, a gyakorlatban 1–8 Mflop/s sebességet lehet elérni a problémától és a programozásra fordított figyelemtől függően.

A MAX tábla (ahol a MAX a *Matrix Algebra Accelerator* kifejezést takarja) két darab FPS-164-es gépből van összeállítva, mindegyik géphez kiegészítésként 4 darab, egyenként 2048 elemet tartalmazó vektor regisztert csatoltak. A MAX táblákat használva lehet összeállítani az FPS 164/MAX gépet, amelynek a fő vezérlője egy FPS 164 és ehhez még legfeljebb 15 darab MAX táblát lehet hozzácsatolni. Ezzel az FPS 164/MAX teljes elméleti sebessége 341 Mflop/s. A rendszer rendkívül alkalmas lineáris algebrai műveletek (mátrixszorzás, sajátérték számítás, stb.) végzésére.

5.3. Az Intel Hypercube

Az n-dimenziós kocka formájára kialakított összeköttetések jellemzők az *Intel* cég által kifejlesztett *Hypercube* számítógépre ($n = 3, 4, 5, 6$ a kapható változatok). Az egész rendszert a *Cube Manager* nevű személyi számítógép felügyeli. Az n-dimenziós kocka egy csúcspontjában egy processzor és egy lebegőpontos társprocesszor van, továbbá helyi memória (128 K-tól 512 K-ig). A kocka egy csúcsa n szomszédos csúcshoz van kötve, egy csúcstól egy másikig átlagosan $n/2$ élen kell az üzenetnek végighaladnia. Az első típusok esetén a kommunikációs idők viszonylag nagyok, mintegy ötven lebegőpontos szorzást tud a gép végrehajtani azalatt, amíg egyik csúcsból a másikba átér az üzenet. A csúcsban elhelyezett memóriát mintegy 1/3 részben foglalja le az operációs rendszer és az üzenetek továbbítására szolgáló program. A rendszert lényegében úgy lehet jellemezni, hogy 2^n személyi számítógép van összekapcsolva, egy gép felügyelete alatt.

A következő változatban (1985) 80286-os mikroprocesszor van a csúcsokban 80282-es társprocesszorral és ezt egy *Intel* 310-es mikrogép ellenőrzi. Az újabb, iPSC/2 változatban (1987) 80386-os processzorok vannak a csúcspontokban, a mellette lévő 80387-es társprocesszor önálló működésre képes. Ebben a rendszerben a csúcsok közötti kommunikációt is felgyorsították, a *Direct-Connect* elnevezésű hálózat lehetővé teszi, hogy bármely csúcsból bármely csúcsba lényegében azonos idő alatt lehessen üzenetet küldeni (*wormhole routing*). Az üzenettovábbítás sebessége 2,8 Mbyte/sec, ha a kapcsolat már létrejött két csúcs között. A legújabb, iPSC/860 jelű, lényegében változatlan felépítésű gépen i860-as, 80 Mflop/s sebességűnek hirdetett processzorok vannak, amivel a legnagyobb kiépítettségű, 128 processzoros gép elméletileg elérhető legnagyobb sebessége 7,6 Gflop/s, melyhez 2 GByte főmemória és 100 GByte lemezen lévő memória csatolható. A rendszer nagyon költséghatékony, vagyis egy Gflop ára itt csak a tizede a CRAY-YMP gépen kapható 1 Gflop árának.

5.4. Miniszuper számítógépek

A szuper- és párhuzamos számítógépek piacának legújabb vonása, hogy olyan viszonylag olcsó gépek jelennek meg, melyek ára néhány százezer dollár, egy felhasználóra vannak tervezve és egy nagyobb íróasztalnyi helyet igényelnek csupán. Az első ilyen a *Convex társaság* C-1 (1985) gépe volt, amely egy olcsó, léghűtéses vektorprocesszor 20 Mflop/s csúcssebességgel. A *Convex* azóta egy teljes sorozatot hozott ki a piacra 1988-ban, a C 120-tól a C 240-ig; a sorozat utolsó gépének maximális sebessége 200 Mflop/s.

Az *Ardent* cég léghűtéses *Titan* számítógépe a grafikus rendszerek felhasználóinak alkalmas eszköze, a fél négyzetméternél kisebb alapterületű, 120 cm magas gép könnyen elhelyezhető. A képernyője 1280×1024 pixeles és 600.000 háromdimenziós vektort tud rajzolni egy másodperc alatt. A számítógépnek négy darab vektorprocesszora van, és egy egész műveleteket végrehajtó processzora. A vektorprocesszorok működését vektorregiszterek gyorsítják.

Az *Apolló* 10000-es gépe RISC rendszerű (*reduced instruction set*) 64 bites gép, amely egy gépi ciklus alatt 1,2–1,3 utasítást képes végrehajtani. A rendszert négy független CPU-val felszerelve árulják, mindegyik CPU-hoz egy egész és egy lebegőpontos processzor, gyorstár és egy, a memóriát szervező egység tartozik. A vektorműveletek végrehajtását egy nagy regiszterfájl segítségével gyorsítják fel.

5.5. A Connection Machine

A *Thinking Machines Corp.* által kifejlesztett gép erősen eltér az eddig ismertetettektől [9]. A maximális kiépítettségben 65536 (64 K) processzorral rendelkező gép nagyon nagy mértékű párhuzamoságot tesz lehetővé. Akármelyik processzor elérheti akármelyik másik processzort vagy memóriarészt, így a gyakori kommunikációnak nincs akadálya.

Konkrét alkalmazások (szövegvisszakeresések, ill. képfeldolgozás) esetén a gép mért sebessége 6000 Mip/s, illetőleg 1 Gflop/s volt. A rendszert többféle méretben szállítják, 16 K processzor elemtől 64 K processzor elemig.

Egy processzor elemnek csak egy apró feladata lehet, mivel egy ilyen elem csak egy darab 1 bites ALU-t (*Arithmetic and Logical Unit*), valamint 4096 bit helyi memóriát tartalmaz. A processzor elemeket 16-os csoportokba szervezik és ezek a csoportok a legnagyobb kiépítettségű gépen egy 12 dimenziós kocka csúcsain helyezkednek el. Az egyes műveletek végzését bizonyos processzorokra lehet korlátozni a kihagyandó processzorok megjelölésével (*masking*). Lebegőpontos műveleteket csak speciális programok segítségével lehet végezni; két 32 bites szám összeadásához 750 ns idő szükséges.

A *Connection Machine* elsősorban logikai és egész műveletek végrehajtására alkalmas, a processzorok és a memória közötti jó kommunikáció pedig olyan feladatok megoldására teszi alkalmassá, amelyek a programrészek közötti sűrű érintkezést igénylik. Az összekötő hálózat a csomagkapcsolt és áramkör váltó üzenetközvetítés (*packet and circuit switching*) jó oldalait egyszerre tartalmazza. A gép összekötő hálózatára jellemző, hogy a teljes 32 Mbyte memóriát 30 ms alatt sorba lehet rendezni (*sorting*), egy 160 Mbyte-os adatbázisból pedig 50 ms alatt kiemelhető az a 20 dokumentum, amely a megadott 200 szavas kereső mintához a legközelebb volt (hírügynökségi archív anyagok közötti keresés adott témában).

A számítógéptípus újabb generációja a CM-2 jelű gép (1987), amelynek lényeges felépítésbeli változása az első nemzedékbeli CM-1-hez képest az, hogy minden 32 processzorelemhez egy *Weitek lebegőpontos processzort* adtak hozzá; összesen 2048-at. Ezek a *Weitek processzorok* 7,5 MHz-en működnek, 24 Gflop elméleti csúcsteljesítménnyel, amiáltal a CM-1-nél mintegy 30-szor gyorsabb lett a gép. A memóriát 16-szorosára bővítették, vagyis most 8 Kbyte van minden egyes processzorelemnél, összesen 512 Mbyte. Az eredeti processzorok közötti összeköttetést meghagyva a gép processzorait *hypercube* kapcsolatokkal is ellátták.

5.6 A T-series és a transzputerek

A *Floating Point System* által 1986-ban kihozott *T-series gépcsalád* a *Connection Machine*-hez hasonló nagymértékű párhuzamosságot tesz lehetővé. A processzor elemek száma a különböző modellekben 1 K és 16 K között változik. Egy processzor elem egy az *Inmos Inc.* által kifejlesztett 32 bites *transzputert* és 1

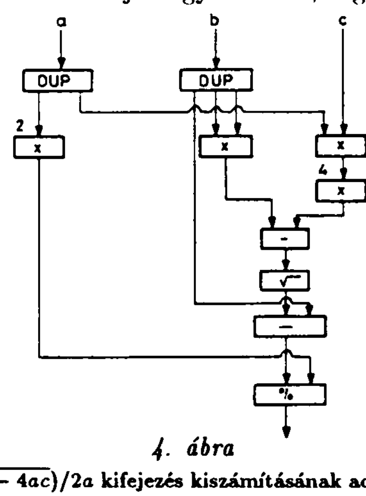
Mbyte helyi memóriát tartalmaz [5]. Egy *transzputer* önmagában képes 16 Mflop csúcsteljesítményre, a teljes rendszer számítási erejét 2–4 Gflop/s-ra becsülik. A *transzputereket* processzor modulokba szervezik, egy modul nyolc *transzputert* tartalmaz. A modulok egy n-dimenziós kocka ($7 \leq n \leq 11$) élei mentén vannak összekötve. A teljesítmény növelését külön lemezegységek segítik elő, minden processzor modulnak van egy 256 Mbyte nagyságú tárral rendelkező lemezegysége. Így a *T-series* gépei inkább lebegőpontos számítások elvégzésére és nem túl sok kommunikációt igénylő feladatok megoldására használhatók. (Míg a *Connection Machine multiprocesszornak*, a *T-series* gépei *multikomputernek* tekinthetők inkább).

A *transzputerek* néhány tulajdonságát említjük meg még kiegészítésként. Ennek a processzorcsaládnak a legfontosabb jellemzője, hogy négy darab kétirányú soros kapcsolata van mindkét irányban 1 Mbyte/s sebességű adatátvitellel, amelyek a *transzputerek* összekapcsolását nagyon egyszerűvé teszik (nincs szinkronizálás és buffer). Az adatátvitel ezeken a csatornákon keresztül a processzor működésétől függetlenül történik. Programozási nyelvük a kevéssé elterjedt *occam*.

A T 414 számú *transzputer* belső ciklusideje 50 ns, valamint van egy 2 Kbyte-os 50 ns ciklusidejű munkaterülete. Gyakorlati feladatokon egy *transzputer* körülbelül kétszer olyen gyors, mint egy VAX 11/780, egy egyetlen kártyán elhelyezhető 17 *transzputer* körülbelül 30 darab VAX 11/780 teljesítményét adja egyetlen VAX gép árának töredékéért [1].

5.7. *A Manchester adatfolyam számítógép*

Az adatfolyam (*data flow*) gépek mind a konvencionális, mind a párhuzamos számítógépektől eltérő alapelven működnek. Röviden szólva utasítások vannak a gépben raktározva és amint minden szükséges adat megérkezett egy utasításhoz, akkor az utasítás végrehajtódik. Ezért ezeket a gépeket adatmeghajtású (*data driven*) gépeknek is nevezik, szemben a többi számítógéppel, amelyek utasításmeghajtásúak (*istruction driven*), hiszen amint bejön egy utasítás, végrehajtódik.



4. ábra

A $(-b \pm \sqrt{b^2 - 4ac})/2a$ **kifejezés kiszámításának adatfolyam gráfja**

Egy programban a szükséges matematikai műveleteket adatfolyam gráffal ábrázoljuk. Példaként tekintsük a $(-b \pm \sqrt{b^2 - 4ac})/2a$ kifejezést kiszámító gráfot (4. ábra). A gráf ívein az adatoknak megfelelő jelképek (*token*) mozognak, minden egyes adatnak külön jelképe van. Amint a gráf egy pontjában lévő művelethez megérkezik az összes, az adott művelet elvégzéséhez szükséges jelkép, a számítás végrehajtódik és a számítás eredményeként egy új jelkép keletkezik. A DUP elnevezésű művelet csak több példányban állítja elő a bemenetként szereplő adatot. Mivel a számítás több ágon haladhat, ezért vigyázni kell, hogy egy adott változó minden szükséges helyen azonos értéket kapjon (vagy a kellőképpen megváltoztatott értéket). Ezt azáltal lehet elérni, hogy minden adat minden egyes változata új azonosítót kap a programban.

A *Manchester adatfolyam gépnek* [14] három része van: a processzor gyűrű (jelkép-párosító, utasítás tár, processzor egység), a szerkezeti tár és az összekötő hálózat. Az adatok csomagkapcsolt módon kerülnek egyik pontból a másikba. A processzor gyűrűkből több darabot is rá lehet kapcsolni a gépre, ezek aztán egymástól függetlenül, párhuzamosan működhetnek. A processzor gyűrű egyes részei csövesítve vannak.

A processzor gyűrű először is a jelkép-párosító (*token matching*) egységben felismeri az odaérkező input jelképeket. Ha egy művelet elvégzéséhez szükséges másik (többi) jelkép már rendelkezésre áll, akkor mindegyik szükséges jelképet (ez unikális, nincs belőle több!) továbbküldik, egyébként tárolják a már várakozókkal együtt. Az utasítástárból kiveszik a megérkezett jelképeket felhasználó utasítást, és végrehajtásra a processzor egységbe küldik.

Az adatfolyam gépeknek még csak kisérleti példányai vannak. Mind hardver, mind szoftver problémák hátráltatják elterjedését. Nyelve a SISAL (egy kifejezésorientált nyelv), amely kevéssé ismert. A problémák közül megemlítendő, hogy a gép időnként óriási mennyiségű elemi utasítást generál, amelyek közül aztán nehezen tudja kiválasztani azokat, amelyeket a feladat elvégzése érdekében elsősorban kellene végrehajtania.

Bár nem kifejezetten adatfolyam gép, de nagyon sok hasonlóságot mutat az adatfolyam gépekkel a HEP (*Heterogenous Element Processor*) számítógép.

5.8. A szisztolikus tömbök

A szisztolikus tömbök elnevezés egyfajta, speciális feladatok ellátására kifejlesztett egyedi felépítésű processzorcsoportot takar. A tömb csomópontjai processzorok, amelyek az aritmetikai műveleteket végzik. Az egyes processzorok által kiszámított részeredményeket általában valamely szomszédos processzornak továbbítják, és csak a végeredményt (például egy mátrixnak egy másik mátrixszal képzett szorzatát) helyezik el a memóriában. Ilyen szisztolikus tömb a WARP és az iWARP.

Megjegyzés: A szerző köszönettel tartozik a lektoroknak, akik értékes megjegyzéseikkel segítették elő a cikk elkészültét.

IRODALOM

[1] AMDAHL, G.M., *Validity of the single-processor approach to achieving large scale computing capabilities* in: AFIPS Conf.Proc., vol. 30 (AFIPS Press,, Reston, 1967), pp. 483–485.

[2] ASKEW, C.R., CARPENTER, D.B., CHALKER, J.T., HEY, A.J.G., MOORE, M., NICOLE, D.A., PRITCHARD, D.J., "Monte Carlo simulation on Transputer arrays", *Parallel Computing* **6** (1988), 247–258.

[3] BERTSEKAS, D.P., TSITSIKLIS, J.N., *Parallel and distributed programming* (Prentice Hall, New York, 1989), pp. 715.

[4] DEÁK, I., "Uniform random number generators for parallel computers", *Parallel Computing* **15** (1990), 155–164.

[5] DEÁK, I., "Random number generators and simulation", *Mathematical methods of operations research* ed. A. Prékopa, (Akadémiai Kiadó, Budapest, 1990), pp. 324.

[6] DEÁK, I., "Procedures to solve STABIL on a parallel computer,", Technical Report, *Industrial Engineering 89/10* (1989), 26, Informatica (leadva).

[7] DONGARRA, J.J., ed., "Experimental parallel computing architectures", *Special Topics in supercomputing* (North Holland, 1986), pp. 302.

[8] FLYNN, J.J., "Very high speed computing systems", *Proc. of the IEEE* **54** (12) (1966), 1901–1909.

[9] FRENKEL, K.A., "Evaluating two massively parallel machines", *Comm. ACM* **29** (1986), 752–758.

[10] GUSTAFSON, J.L., "Reevaluating Amdahl's law", *Comm. ACM* **31** (1988), 532–533.

[11] HOCKNEY, R.W. JESSHOPE, C.R., *Parallel computers: architecture, programming and algorithms* (Adam Hilger, Bristol, England, 1981).

[12] MANUEL, T. et al., "Supercomputers: the proliferation begins", *Electronics* (March 3, 1988), 51–77.

[13] MCGREGOR, J.D. et al., "Support for multiprocessing", *Comm. ACM* **32** (1989), 1062–1101.

[14] QUINN, M.J., *Designing efficient algorithms for parallel computers* (McGraw-Hill Book Comp., New York, 1987).

(Beérkezett: 1989. szeptember 15)

(Átdolgozva beérkezett: 1991. május 22)

DEÁK ISTVÁN
BME VILLAMOSMÉRNÖKI KAR MATEMATIKA TANSZÉK
1111 BUDAPEST, STOCZEK U. H ÉP. III. E.

ARCHITECTURE OF PARALLEL COMPUTERS

DEÁK, I.

Parallel computers using several processors and closely related computers based on recent advanced techniques are summarized and discussed according to their architecture and operating principles. After some history and basic definitions the problems related to their architecture and efficiency are discussed. The most important (widely used, fast, or specially interesting) computers are shortly described.

KÖNYVISMERTETÉS

MARCEL F. NEUTS: *Structured Stochastic Matrices of M/G/1 Type and Their Applications* (Marcel Dekker Inc., 1989)

Az algoritmikus valószínűségszámítás egyik úttörőjének számító M.F. NEUTS ezen könyve a mintegy 10 évvel ezelőtt megjelent másiknak (M.F. NEUTS: *Matrix-Geometric Solutions in Stochastic Models — An Algorithmic Approach,* John Hopkins Univ. Press, 1981) testvére. Mindkét könyv speciális szerkezetű *Markov-láncokkal* ill. *Markov felújítási folyamatokkal* foglalkozik. Ilyen folyamatok gyakran fordulnak elő különböző alkalmazott valószínűségszámítási problémákban, így például sorbanállási feladatokban is.

Az első könyv olyan *Markov-láncokat* tárgyal, melyek átmenetvalószínűség — ill. intenzitásmátrixa (1) szerkezetű (G/M/1 típusú), míg a későbbi tárgyalja a (2) típusú esetet

$$(1)\quad \begin{pmatrix} \mathbf{B}_0 & \mathbf{A}_0 & & & \\ \mathbf{B}_1 & \mathbf{A}_1 & \mathbf{A}_0 & & \\ \mathbf{B}_2 & \mathbf{A}_2 & \mathbf{A}_1 & \mathbf{A}_0 & \\ \vdots & & & & \ddots \end{pmatrix}, \qquad (2)\quad \begin{pmatrix} \mathbf{B}_0 & \mathbf{B}_1 & \mathbf{B}_2 & \mathbf{B}_3 & \\ \mathbf{C}_0 & \mathbf{A}_1 & \mathbf{A}_2 & \mathbf{A}_3 & \dots \\ & \mathbf{A}_0 & \mathbf{A}_1 & \mathbf{A}_2 & \\ & & \ddots & & \end{pmatrix},$$

$\mathbf{A}_i, \mathbf{B}_i, \mathbf{C}_i$ négyzetes mátrixok

($M/G/1$ típusú mátrixok). A két osztály között azonban jelentős különbségek vannak: míg például a $G/M/1$ esetben a stacionárius eloszlás egyszerű szerkezetű (ún. mátrix-geometrikus eloszlást követ), addig az $M/G/1$ esetben az csak sokkal bonyolultabban adható meg. Persze hasonló probléma lép fel már a klasszikus $G/M/1$ és $M/G/1$ sorok *beágyazott Markov-láncainak* vizsgálatakor is. Mindkét könyv nagy előnye, hogy mindvégig igyekszik a gyakorlati (számítógépes) megvalósítást szem előtt tartani.

Nézzük meg most részletesen, milyen témákat ölel fel az utóbbi könyv. Az 512 oldalas könyv, melynek irodalomjegyzéke kb. 1400 hivatkozást tartalmaz, 6 fejezetre tagolódik.

Az 1. fejezet (*The $M/G/1$ Queue and Some of Its Variants*) a klasszikus $M/G/1$ sorra vonatkozó eredményeket adja meg (ergodikusság feltétele, stacionárius eloszlás, várakozási idők, stb.) olyan tárgyalásban, melyhez hasonlóan írható le az általános eset is.

A 2. fejezet (*Markov Chains of $M/G/1$ Type*) definiálja azt a problémakört, mellyel lényegében az egész könyv foglalkozik. Néhány fontos példa után a fejezet főleg az ún. alapperiódus tulajdonságait tárgyalja, melyek az ergodikusság szempontjából lényegesek. Az ergodikusság, a stacionárius eloszlás és momentumainak megadása a 3. fejezet (*Positive Recurrence*) témája. A hátralevő fejezetekben az itt megadott eredmények alapján speciális, alkalmazási kérdések kerülnek sorra.

A 4. fejezetben (*Applications to Queues With Poisson Arrivals*) olyan példák szerepelnek, ahol a beérkezési folyamat*Poisson*. Így az $M/SM/1$ sor (kiszolgálási idők *Markov felújítási folyamatot* alkotnak) és variánsai, valamint az $M/G/1$ sor csoportos kiszolgálással.

Az 5. fejezet (*Versatile Arrival Processes*) olyan általánosításokat tárgyal, ahol a beérkezési folyamat már nem feltétlenül *Poisson*, sőt nem is feltétlenül felújítási, hanem például egy *Markov-lánc* által modulált (változó intenzitású) *Poisson-folyamat*. Az ilyen típusú legáltalánosabb beérkezési folyamat a NEUTS által "*versatile Markov point process*"-nak nevezett, amely számos korábbi esetet (változó intenzitású vagy csoportos beérkezések) magába foglal.

A 6. fejezet (*Selected Special Models*) négy olyan modell leírását adja meg, melyek egyrészt az alkalmazásokban gyakran előfordulnak, másrészt a tárgyalt elmélet különböző kiegészítéseit adják.

A könyv minden fejezete után irodalmi megjegyzéseket és feladatokat találunk. A feladatok között vannak alkalmazási problémák, elméleti kiegészítések, melyek további munkára adhatnak ösztönzést, de van amit szakdolgozati feladatként is felhasználhatunk.

A gondosan összeállított és szerintem hamarosan — testvéréhez hasonlóan — klasszikussá váló könyvet haszonnal forgathatják az alkalmazott valószínűségszámítás szerelmesei: matematikusok, programtervezők, mérnökök egyaránt.

Szirtes András

GOTTFRIED JETSCHKE: *Mathematik der Selbstorganization. Qualitative Theorie nichtlinearer dynamischer Systeme und gleichgewichtsferner Strukturen in Physik, Chemie und Biologie* (Friedr. Vieweg und Sohn Braunschweig–Wiesbaden, 1989).

A könyv valódi célja, hogy (a matematikai statisztika kivételével) összegyűjtse mindazt a matematikai anyagot, amit manapság a természettudományokban alkalmazni szokás. Ezt a célt az alcím szűkíti, a cím pedig még ennél is jobban korlátozza. A jelenség magyarázata az, hogy habár az önszerveződés elmélete (lévén egy számos tudományágat magába sűrítő terület) az összes, itt bemutatott matematikai eszközt használja, ugyanezek az eszközök (esetleg mélyebben, de nem nagyobb számban) alkalmasak arra, hogy a fizika, kémia és matematika számos más problémáját is megoldják. Így azután egyetérthetünk a hátsó borító állításával, amely szerint az olvasó számos, a lineáris és nemlineáris differenciálegyenletekre, a bifurkációkra, a katasztrófaelméletre és a sztochasztikus folyamatokra vonatkozó könyv elolvasását megtakaríthatja magának, ha helyettük ezt a művet olvassa el (illetve tanulmányozza át!). Ezután ugyanis képes lesz eredeti cikkek olvasására. Más szavakkal a könyvet úgy is jellemezhetjük, hogy alig van olyan fogalom (definíció vagy állítás), amelyet a könyv ne tárgyalna, vagy legalább ne érintene. Ennek bizonyítására felsoroljuk a könyv részletes tartalomjegyzékét. Bevezetés, 1. Determinisztikus dinamikai rendszerek, 2–4. Egy/Kettő/Kettőnél több szabadságfokú rendszerek, 5. Kaotikus attraktorok, 6. Bifurkácóelmélet, 7. Katasztrófaelmélet, 8. Reakció–diffúzó rendszerek, 9. Sztochasztikus dinamikai rendszerek, 10. Sztochasztikus differenciálegyenletek, 11. Születési–halálozási folyamatok, 12. Diszkrét idejű zajos rendszerek, 13. Sztochasztikus parciális differenciálegyenletek. A Függelék beszél a modellalkotás folyamatáról általánosságban; azután rövid bevezetőt ad néhány tudományba (mechanika, elektromosságtan, kémia, biológia) oly módon, hogy vázolja fő problémáikat és gondolkodási sémáikat. A Függelék 3. szakasza a klasszikus egyensúlyi és nemegyensúlyi termodinamikába és a termodinamikai stabilitáselméletbe akar matematikailag megalapozott bevezetést adni, ami természetesen szerzőnknek sem sikerülhet a racionális termodinamika keretein kívül. Végül a szinergetika népszerűsítéséről olvashatunk.

149 jól kigondolt, részben eredeti ábra segít a szöveg megértésében. A feladatok megoldása, vagy ötletek megoldásukra szintén megtalálhatók a könyv végén. A könyv tartalmaz egy válogatott hivatkozási listát (*Weiterführende Literatur!*) is, amely valóban továbbvezeti az olvasót, hiszen túlnyomórészt közismert, bevált tankönyveket és klasszikus cikkeket tartalmaz. Az írást tárgymutató zárja.

A könyv külső alakja is igen kellemes, néhány apróságot leszámítva. A betűk és az ábrák vonala inkább szürke, mint fekete, továbbá a borítót a festékréteg nem borítja be teljesen.

Összefoglalva megállapítható, hogy a természettudományok folyamatainak matematikai modellezésével foglalkozó laboratóriumok bizonyára szívesen fogadják szerte a világon azokat a fiatalembereket, akik levizsgáztak ebből a könyvből. A könyv magyarra való lefordítása (ami manapság nyilván csak ábránd) komoly nyeresége lenne a magyar szakkönyvirodalomnak, ugyanis messze felülmúlja HAKEN hasonló célú, magyarul megjelent művét.

Tóth János

IRANPOUR, R. és CHACON, P., *Basic Stochastic Processes. The Mark Kac Lectures* (MacMillan Publishing Company, New York és Collier MacMillan Publishers, London, 1988).

Ismét itt egy könyv abból a sorozatból, amelyet BAILEY, BARTLETT, BHARUCHA-REID, KARLIN és TAYLOR, KEMENY és SNELL, LAHRES és mások indítottak el néhány évtizeddel ezelőtt: egy bevezető a sztochasztikus folyamatok elméletébe, alkalmazók számára.

A könyv a néhai MARK KAC professzor előadásain alapul, amelyeket mérnök-, matematikus és fizikus hallgatók számára tartott a *Dél-Karolinai Egyetemen*. A könyv alapvető célkitűzése, hogy ne használjon absztrakt matematikai fogalmakat; különösen kerülje el a mértékelmélet alkalmazását. A könyv szövege önmagában érthető. Ezt segíti elő a valószínűségszámításra vonatkozó bevezető Nulladik fejezet, és a lineáris algebrai ismereteket nyújtó Függelék. Még az Első (*Fourier-transzformáció és az inverziós formula*) és a Második (*Fourier-sorok*) fejezet is bevezetőnek tekinthető. A könyvben mindvégig nyilvánvaló ennek a megközelítésnek az ereje és egyszerűsége.

A könyv további részének tartalma:

Harmadik fejezet: *Poisson-folyamatok* (több definícióval; kapcsolat az exponenciális és az egyenletes eloszlással, rendezett mintákkal; születési-halálozási folyamat; felújítási folyamatok)

Negyedik fejezet: *Shot Noise* (a véletlen zaj spektrális sűrűsége)

Ötödik fejezet: *Gauss-folyamatok* (a karakterisztikus függvény; a kovariancia-mátrix és a magasabbrendű korrelácók; a *Wiener-folyamat;* stacionárius folyamatok, a spektrális sűrűség, *Bochner tétele*)

Hatodik fejezet: *Gauss-Markov-folyamatok* (*Doob tétele; a Chapman-Kolmogorov-egyenlet, az Ornstein-Uhlenbeck-folyamat*)

Hetedik fejezet: *A Brown-mozgás* (*Einstein megközelítése; a Brown-mozgást* végző részecske sebessége; a *Fokker-Planck-egyenlet;* sztochasztikus differenciálegyenletek)

Nyolcadik fejezet: *Markov-láncok* (az *Ehrenfest-modell;* sztochasztikus mátrixok; véletlen bolyongás; a stacionárius eloszlás, elágazó folyamatok; ergodicitás)

A Függeléket kivéve minden fejezethez számos feladat csatlakozik, amelyeket helyesebb lenne inkább gyakorlatoknak nevezni.

A szerzők intuitív megközelítése lehetővé teszi, hogy igen nagy anyagot tekintsenek át (például még sztochasztikus differenciálegyenletekről is esik szó).

Habár a könyv igazi nyeresége az alkalmazott matematikai irodalomnak, engedtessék meg néhány kritikai megjegyzés. Először is: egyetlen hivatkozás sincs a könyvben. Nem is az eredeti dolgozatokra való utalás hiányzik, hanem a matematikai előzményeket (lineáris algebra, valószínűségszámítás, vagy komplex függvénytan) tartalmazó, illetve a hasonló célú könyvek (például a fent említett szerzőktől) megemlítése. Másodszor: az intuitív megközelítés miatt a könyv nem alkalmas arra, hogy matematikus hallgatók ebből kezdjék az ismerkedést a sztochasztikus folyamatok elméletével. Harmadszor: talán nem ártott volna néhány kémiai és biológiai példát is megemlíteni az alkalmazások között.

A szerzők szándéka volt, hogy megőrizzék MARC KAC előadói stílusát, amely különösen rövid és elegáns bizonyításokban nyilvánul meg.

Összegezve: Érdemei miatt a könyv melegen ajánlható mérnök- és fizikus hallgatók számára, valamint a sztochasztikus folyamatokat oktató tanárok számára. Semmiképpen nem ajánlható azonban első könyvnek matematikus hallgatók számára. Aki pedig ezzel végzett, annak megoldandó tudományos problémákat bőségesen kínál a következő két könyv:

KAMPEN, N. G. VAN, *Stochastic Processes in Physics and Chemistry* (North Holland, Amsterdam, 1981) és GARDINER, G. W., *Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences* (Springer Series in Synergetics, Vol. 13, Springer Verlag, Berlin, 1983);

az alkalmazók számára legfontosabb eredmények pedig könnyen érthető formában leginkább az alábbi műben találhatók meg:

IOSIFESCU, M. és TAUTU, P., *Stochastic Processes and Applications in Biology and Medicine* (Ed. Acad. RSR, Bucuresti és Springer Verlag, Berlin – Heidelberg – New York, 1973).

Ha több lehetősége lenne manapság matematikai tárgyú könyvek megjelenésének Magyarországon, akkor javasolnám, hogy az ismertetett könyv kerüljön fel a lefordítandók listájára.

Tóth János

HELYREIGAZÍTÁS

A Szerkesztőbizottság közli az Olvasóval, hogy a **14.** kötet (1989) 1–2-es számában megjelent MORRIS W. HIRSCH: Dinamikai rendszerek és differenciálegyenletek című dolgozat első oldalán a lábjegyzet sajnálatos szerkesztési hiba miatt lemaradt.

A lábjegyzet szövegét az alábbiakban pótoljuk:

A fordítás M.W. HIRSCH, "The Dinamical Systems Approach to Differential equations", Bulletin of American Mathematical Society (1984), Volume 11, pp. 1–64 dolgozatából az American Mathematical Society engedélyével készült.

A kiadásért felelős az ELTE TTK dékánja
Szedte a KLTE Informatikai és Számítóközpont Kiadvány Szerkesztő Csoportja
és nyomta az ELTE Sokszorosító Üzeme
Felelős vezető: Arató Tamás
Budapest, 1992. – ELTE 92303
Megjelent: 18,55 (A/5) ív terjedelemben
350 példányban
HU ISSN 0133-3399

ÚTMUTATÁS A SZERZŐKNEK

Az Alkalmazott Matematikai Lapok csak magyar nyelvű dolgozatokat közöl. A kéziratok gépelését, olyan formában kérjük, hogy minden gépelt oldal 25, egyenként átlag 50 betűhelyes sort tartalmazzon. A közlésre szánt dolgozatokat három példányban kell beküldeni. Előnyben részesülnek a TEX-ben elkészített dolgozatok. Ezeket két kinyomtatott példány kíséretében diszketten kérjük beadni.

A kéziratok szerkezeti felépítésének a következő követelményeket kell kielégíteni. A fejlécnek tartalmaznia kell a dolgozat címét, a szerző teljes nevét, valamint annak a városnak a nevét, ahol a szerző dolgozik. A fejléc után egy, képletet nem tartalmazó, legfeljebb 200 szóból álló kivonatot kell minden esetben megadni. A dolgozatot címmel ellátott szakaszokra kell bontani, és az egyes szakaszokat arab sorszámmal kell ellátni. Az esetleges bevezetésnek mindig az első szakaszt kell alkotnia. Az irodalomjegyzék mindig az utolsó szakasz kell, hogy legyen, és azt nem kell sorszámmal ellátni. Az irodalomjegyzék után, a kézirat befejezéseképpen fel kell tüntetni a szerző teljes nevét és a munkahelye (illetve lakása) pontos postai címét. A dolgozatban előforduló képleteket szakaszonként újrakezdődően, a képlet előtt két zárójel közé írt kettős számozással kell azonosítani. Természetesen nem szükséges minden képletet számozással ellátni. Az esetleges definíciókat és tételeket (segédtételeket és lemmákat) ugyancsak szakaszonként újrakezdődő, kettős számozással kell ellátni. Kérjük a szerzőket, hogy ezeket, valamint a tételek bizonyítását a szövegben kellő módon emeljék ki. Minden dolgozathoz csatolni kell egy angol, német, francia vagy orosz nyelvű, külön oldalra gépelt összefoglalót. Amennyiben lehetésges, kérjük a nyomtatás számára különösen nehézkes matematikai jelölések használatának az elkerülését.

A dolgozatok ábráit és az esetleges lábjegyzeteket a dolgozat végén, különálló lapokon kérjük beküldeni. Mind az ábrákat, mind a lábjegyzeteket a dolgozat szakaszokra bontásától független, folytatólagos arab sorszámozással kell ellátni. Az ábrák elhelyezését a dolgozat megfelelő helyén, széljegyzetként feltüntetett, ábraazonosító sorszámokkal kell megadni. A lábjegyzetekre a dolgozaton belül az azonosító sorszám felső indexkénti használatával lehet hivatkozni.

Az irodalmi hivatkozások formája a következő. Minden hivatkozást fel kell sorolni a dolgozat végén található irodalomjegyzékben, a szerzők, illetve társszerők esetén az első szerző neve szerint alfabetikus sorrendben úgy, hogy a cirill betűs szerzők nevét a Mathematical Reviews átírási szabályai szerint latin betűsre kell átírni. A folyóiratban megjelent cikkekre [1], a könyvekre [5], a kötetben megjelent dolgozatokra [4], a disszertációkra [3] és a gépi program leírásokra [2] a következő minta szerint kell hivatkozni:

[1] Farkas, J., Über die Theorie der einfachen Ungleichungen, *Journal für die reine und angewandte Mathematik* **124** (1902) 1-27.

[2] Kéri, G., „DUALSIMP", rutin a CDC 3300-as gépekre (Magyar Tudományos Akadémia Számítástechnikai és Automatizálási Kutató Intézete, CDC 3300 felhasználói ismertetők 2. 1973. május) 19-20.

[3] Prékopa, A., „Sztochasztikus rendszerek optimalizálási problémáiról", doktori értekezés. Magyar Tudományos Akadémia, Budapest, 1970.

[4] Prabhu, N. U., „Recent research on the ruin problem of collective risk theory", in: *Inventory Control and Water Storage* Ed. A. Prékopa (János Bolyai Mathematical Society and North-Holland Publishing Company, Amsterdam-London, 1973) 221-228.

[5] Zoutendijk, G., *Methods of Feasible Directions* (Elseiver Publishing Company, Amsterdam and New York, 1960).

A dolgozatok szövegében az irodalmi hivatkozás számait szögletes zárójelben kell megadni, mint például [5] vagy [4, 76-78]. A szerzők a dolgozatukról 50 darab ingyenes különlenyomatot kapnak. A dolgozatok után szerzői díjat az Alkalmazott Matematikai Lapok nem fizet.

Ára: 96 Ft

TARTALOMJEGYZÉK

INDEX