Alkalmazott matematikai lapok

1993/1–2

A MAGYAR TUDOMÁNYOS AKADÉMIA
MATEMATIKAI ÉS FIZIKAI TUDOMÁNYOK
OSZTÁLYÁNAK KÖZLEMÉNYEI

17.
KÖTET

ALKALMAZOTT MATEMATIKAI LAPOK

A MAGYAR TUDOMÁNYOS AKADÉMIA MATEMATIKAI ÉS FIZIKAI TUDOMÁNYOK OSZTÁLYÁNAK KÖZLEMÉNYEI

XVII. kötet 1–2. szám

Szerkesztőség és kiadóhivatal: 1088 Budapest, Múzeum krt. 6–8.

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
1088 Budapest, Múzeum krt. 6–8.

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 ara kötetenként 850 forint. Megrendelések a szerkesztőség címén lehetségesek.

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 17 (1993) 1–18

PÁRHUZAMOS SZÁMÍTÓGÉPEK: OPTIMALIZÁLÁSI PROGRAMOK

BÁLINT ERZSÉBET ÉS DEÁK ISTVÁN

Budapest

A jelenleg létező párhuzamos számítógépes architektúrák rövid leírása után az optimalizálási programcsomagok és algoritmusok párhuzamos változatait tekintjük át. A cikk lényegi részében három fő témával foglalkozunk: a szimplex algoritmus változatai, a nemlineáris programozás (beleértve a hálózati folyamatokat) és a diszkrét programozás szoftverjei. A cikket egy 40-nél több tételt tartalmazó irodalomjegyzék egészíti ki.

1. Bevezetés

A párhuzamos algoritmusok elméletének tanulmányozása a hatvanas évek elején, még a párhuzamos számítógépek, a többprocesszoros rendszerek tényleges megjelenése előtt megkezdődött. A különböző típusú párhuzamos rendszerek gyakorlati megvalósulása tovább fokozta a kutatók érdeklődését. Az utóbbi években egyre többen foglalkoznak az ilyen algoritmusok implementációjának kérdésével. A kutatások nagy része arra irányul, hogyan lehet az egyes algoritmusoknak minél jobb (gyorsabb, hatékonyabb) párhuzamos változatát létrehozni már meglévő párhuzamos számítógépeken. Sokan foglalkoznak azzal a kérdéssel is, hogyan lehetne egy adott algoritmus végrehajtása szempontjából minél jobb párhuzamos rendszert létrehozni.

A operációkutatás területe az egyik első olyan terület, ahol már érezhető ennek a kutatásnak a hatása: számos olyan probléma, algoritmus van itt, amely nagy hasznát veszi a párhuzamos végrehajtási technika alkalmazásának — ilyenek például a nagy méretű operációkutatási feladatok, a fán kereső algoritmusok, szinte minden dinamikus programozási probléma, stb. A párhuzamos feldolgozás lehetővé teszi egyrészt azt, hogy nagyméretű feladatokat az eddiginél gyorsabban oldjunk meg, másrészt azt, hogy olyan komplex feladatokra, amelyek megoldása eddig túl költséges, esetleg lehetetlen volt, most gazdaságos megoldásokat adjunk, kiszélesítve így az operációkutatás számára megközelíthető problémák körét.

Néhány operációkutatási algoritmusnak már több párhuzamos változatát is kidolgozták. Ezek persze számos jellemzőjükben különböznek egymástól: másképp történik bennük az algoritmus egymástól független feladatokra osztása, a feladatmodulok együttműködését, az algoritmus helyes végrehajtását biztosító vezérlés, eltérnek a kommunikáció geometriájában. Jellemző azonban, hogy egy algoritmus különböző párhuzamos implementációiban a modulok szemcsézettsége nagyjából azonos. (A párhuzamos algoritmus moduljainak szemcsézettsége azt a maximális

számítási feladatot jelenti, amelyet egy feladatmodul úgy végezhet el, hogy közben nem kommunikál más feladatmodulokkal: ez a jellemző tehát elsősorban azt tükrözi, hogy mekkora a kommunikáció szerepe az algoritmusban). Egy adott algoritmus gyakran természetes módon osztható fel függetlenül végrehajtható részfeladatokra, s a felosztás meghatározza, milyen lesz a feladatmodulok közötti kommunikáció. Ez persze nem azt jelenti, hogy az algoritmus minden párhuzamos változata szükségszerűen azonos modulszemcsézettségű: néhány algoritmusnak olyan párhuzamos implementációját is kidolgozták már, melyben a feladatmodulok szemcsézettsége eltér az általánostól. Azt, hogy ezek közül melyik jobb, elméleti és gyakorlati vizsgálatoknak kell eldöntenie.

Persze az elvi és gyakorlati összehasonlítás gyakran más-más eredményhez vezet. Egyes párhuzamos algoritmusok például hiába gyorsak, hatékonyak, csak ideális számítógépen implementálhatók — melyekben nincs sem memória-elérési, sem kommunikációs korlátozás —, vagy az implementációhoz speciális többprocesszoros rendszer felépítésére van szükség (általában a szisztolés rendszereket felhasználó algoritmusok ilyenek), esetleg a felhasznált processzorok száma függ a megoldandó feladat méretétől, így a már létező gépek csak kis feladatok megoldására alkalmazhatók. Egy létező gépre tervezett párhuzamos algoritmus pedig éppen azért lesz az elméletileg elvártnál kevésbé gyors vagy hatékony, mert alkalmazkodik az adott gép korlátaihoz. A párhuzamos algoritmusok vizsgálata persze minden irányba kiterjed, a létrehozott algoritmusok között minden változatra találunk példákat.

A párhuzamos számítógépek felépítéséről és működéséről Hockney 1981, Manuel 1988, Deák 1991 könyve illetőleg összeállítása nyújt áttekintést, az általános számítógépes algoritmusokat pedig Bertsekas 1989 és Quinn könyve foglalja össze.

A cikkben az utóbbi években megjelent eredményekről nyújtunk áttekintést: a párhuzamos számítógépeken alkalmazható operációkutatási, optimalizálási algoritmusokról, az ezekkel elérhető számítógépes eredményekről adunk összefoglalást. A következő szakaszban a lineáris programozás szimplex módszerére vonatkozó eredményeket foglaljuk össze. A harmadik szakaszban a kombinatorikus optimalizálás branch and bound módszerének párhuzamos változatait tekintjük át. A negyedik részben a dekompozíciós, az ötödik részben pedig a relaxációs módszerekkel foglalkozunk, míg az utolsó szakaszban egyéb algoritmusokra vonatkozó eredményeket írunk le.

2. A szimplex algoritmus

A lineáris programozási feladatok megoldására széles körben alkalmazott szimplex algoritmussal, gyakorlati és elvi jelentősége miatt, már többen foglalkoztak, többen próbáltak különböző szempontok szerint és különböző párhuzamos környezetben implementálni. A kidolgozott párhuzamos szimplex algoritmusok közül a legtöbb kis modulszemcsézettségű: a szimplex iterációk három lépésén (pivot oszlop kiválasztása, pivot sor meghatározása, pivotálás) belül osztják fel az algoritmust párhuzamosan végrehajtható feladatokra. Az így kapott részfeladatok között gya-

kori a kommunikáció: ezek az algoritmusok az adatokat osztják el a processzorok között, és a számításokat a processzorok együtt végzik el. Vannak már durván szemcsézett gépekre kidolgozott változatok is, de ezek is csak a három lépésen belül párhuzamosítanak, és az implementációs eredmények szerint kevésbé jók, mint a kis modulszemcsézettségű algoritmusok.

A szimplex algoritmus első párhuzamos változatai esetén szisztolés rendszereket terveztek az algoritmus végrehajtására. Az első rendszer (K. Onaga és H. Nagayasu (1984)) egy VLSI wavefront array processor implementáció volt, melyben az algoritmus végrehajtásához szükséges processzorok száma a megoldandó feladat méretétől függött.

Egy másik szisztolés rendszerben (A.A. Bertossi, M.A. Bonucelli, 1987) a processzorok egy $m \times n$-es fa-hálózatot alkotnak (ahol $m-1$ a feltételek, $n-1$ a változók száma), azaz a feldolgozóegységek közül mn egy $m \times n$-es négyzetrácson helyezkedik el, és a többi csúccsal úgy van összekötve, hogy az i-edik sorban lévők is illetve a j-edik oszlopban lévők is egy teljes bináris fa leveleit alkotják ($i = 0, \ldots, m-1,\ j = 0, \ldots, n-1$). A négyzetrácson elhelyezkedő processzorok között vannak elosztva a feladat adatai, a többi processzor a levelek és a fák gyökereit összekötő, általános célú host processzor közötti kommunikációt biztosítja. Az egyes feldolgozóegységek működése egy közös órajelhez van szinkronizálva. A processzorok felépítése egyszerű: néhány regiszterből, egy aritmetikai és logikai egységből és egy vezérlőegységből állnak. A processzorok közötti kommunikáció kétirányú síneken keresztül zajlik.

Bertossi és Bonucelli, alsó becslést adva arra az időre, amely alatt egy $p \times q$-as ($p \times q = O(mn)$) tömbben illetve egy bináris fa-hálózatban összekapcsolt többprocesszoros rendszer végre tud hajtani egy pivot lépést, azt is megmutatta, hogy ezen a rendszeren jobb eredmények érhetők el, mint bármely tömbprocesszoron.

A chip felépítésével viszont az a gond, hogy bár a VLSI technológiával az ilyen — sok, viszonylag egyszerű processzorból álló — szisztolés rendszert egyetlen, vagy legfeljebb néhány áramköri tokban meg lehet valósítani, előfordulhat (minél nagyobb integráltságú egy áramkör, annál gyakrabban), hogy néhány feldolgozóegység vagy sín meghibásodik: ilyen esetben a rendszeren implementált algoritmusok eredményei sem megbízhatók. Széles körben foglalkoznak már azzal a kérdéssel, hogyan lehet olyan hibatűrő redszereket tervezni, amelyek a hibákat felismerik és, önmagukat például újrakonfigurálva, más, működőképes struktúrát alakítanak ki.

Ilyen rendszer az A.A. Bertossi és M.A. Bonucelli által a szimplex algoritmus végrehajtására tervezett másmilyen felépítésű VLSI chip (A.A. Bertossi, M.A. Bonucelli, 1989). Ebben a PE-k egy ún. cousin-connected tree-t (CCT) alkotnak, azaz egy olyan teljes bináris fát, amelyben egy tetszőleges csúcs bal (jobb) utóda a csúcs bátyjának bal (jobb) utódával is össze van kapcsolva. A rendszerben ki kell jelölni a hibás csúcsokat; azok a csúcsok, amelyek a fa gyökeréből nem érhetők el nem hibás csúcsokon keresztül haladó egyszerű úton, az algoritmus szempontjából szintén használhatatlanok: ezek lesznek a „halott" csúcsok. A többi csúcs „élő". Az újrakonfigurált topológia egy hármas fa lesz (reconfigured ternary tree, RTT): csúcsai az élő csúcsok, élei pedig a CCT azon élei, melyeknek nincs hibás vagy halott

szomszédja, és nem két olyan unokatestvért kötnek össze, melyek szülői is élnek.

E hálózat választását az indokolta, hogy a CCT-ben a csúcsok közötti kommunikációs idő illetve a CCT felépítéséhez szükséges hely csak egy konstans faktorban tér el attól, amely egy teljes bináris fa esetén (ez a legelterjedtebb VLSI hálózat) szükséges. Ugyanakkor implementációs eredmények igazolják, hogy ugyanannyi hibás csúcs esetén a CCT-ben sokkal több csúcs marad használható, mint egy teljes bináris fában.

Az újrakonfigurált rendszerben az $m \times n$-es tömbbe rendezett adatok úgy vannak elhelyezve, hogy e tömb minden oszlopát az RTT egy részfája tárolja (minden csúcs legfeljebb egy elemet). A tároló fák „feletti" processzorok feladata az, hogy az adatokat továbbítsa a CCT gyökere és a tároló fák gyökerei között.

A fa-hálózatra kidolgozott párhuzamos szimplex algoritmus alkalmazásával a soros változathoz képest elérhető (elméleti) sebességnövekedés $O(mn/\log n)$, ennek alapján az átlagos processzor-kihasználtság (mivel $O(mn)$ processzorra van szükség) $O(1/\log n)$. A második esetben pedig $O(m)$ a sebességnövekedés, $O(m/N)$ a processzor-kihasználtság, ahol N a processzorok száma (és természetesen $N \geq mn$).

E három implementációra jellemző tehát, hogy az algoritmus végrehajtásához speciális hardware-re (szisztolés rendszerre) van szükség, melyben a szükséges processzorok száma a megoldandó feladat méretétől függ, és a tervezett algoritmusokat nehéz úgy módosítani, hogy olyan esetekben is alkalmazni lehessen őket, melyekben csak korlátozott számú processzor áll rendelkezésre.

Ezeket a problémákat akarta megoldani G.H. Chen, H.F. Ho, S.H. Lin és J.P. Sheu (1990). Olyan párhuzamos szimplex algoritmust dolgoztak ki, amellyel nagyméretű LP feladatokat lehet megoldani a megoldandó feladatoktól független méretű hypercube multicomputereken. Algoritmusuk alapja az, hogy a feladat adatait szétosztják az egyes processzorok között, úgy, hogy a számításokat az adatok elrendezése miatt könnyű legyen elvégezni. Ha N jelöli a megoldandó feladatban a változók, M a feltételek számát, és $MN > p$ (a processzorok száma $p = 2^h$, ahol h a kocka dimenziója), akkor az adatokat a következőképpen osztják el processzorok között: feltehető, hogy $M = k_1 2^m$, $N = k_2 2^n$ $(m+n = h)$ alakú (ha ez nem teljesül, segédsorok és oszlopok vezethetők be). Az A együttható-mátrixot felosztják $2^m 2^n$ $k_1 \times k_2$ dimenziós tömbre $(A_{ij} - k)$, és a d jobboldal-vektort valamint a c költség-vektort is ennek megfelelően osztják részvektorokra. A_{ij}-t az a processzor tárolja, amely címének első m bitje az i, az utolsó n pedig a j bináris reprezentációja, c_j-t az, amely címének első m bitje 0, az utolsó n a j bináris reprezentációja, d_i-t az, amelynek első m bitje az i reprezentációja, utolsó n pedig 1, végül z-t a $(0, \ldots, 0, 1, \ldots, 1)$ című processzor. (A h-dimenziós kocka minden csúcsának van egy h-bites címe, úgy, hogy két processzor pontosan akkor van összekapcsolva, ha a címük egy bit-helyen tér el egymástól).

Az adatoknak ez az elrendezése a szimplex algoritmus egy gyors implementációját teszi lehetővé. Például ha az algoritmus egy iterációjában a pivot oszlopát már meghatároztuk (indexe u), akkor a pivot sora a következőképpen kereshető meg: ehhez a lépéshez d elemeire és a pivot oszlopra van szükség. Először minden i-re $(0 \leq i \leq 2^m - 1)$ a d_i részvektort el kell küldeni annak a processzornak, amely

A_{iy}-t tárolja, ahol $y = [u/k_2]$. Mivel A_{iy} és d_i ugyanabban a részkockában van, és ezek a részkockák függetlenek, ezért ez az átvitel párhuzamosan hajtható végre. Ekkor minden processzor meghatározza a d_i/a_{ik}-k minimumát (f.h. $a_{ik} > 0$), majd a kapott 2^m minimális érték közül kell a legkisebbet kiválasztani (ezek a minimumok mind azonos részkockában vannak). Ezzel meghatároztuk a pivot sor indexét. Hasonlóan hajtható végre az algoritmus többi lépése is.

A teljes algoritmust megvizsgálva kiszámolható, hogy a végrehajtáshoz szükséges kommunikációs és műveleti lépések száma k_1-től, k_2-től és h-tól függ, méghozzá úgy, hogy az elért sebességnövekedés aszimptotikusan lineáris lesz. A gyakorlati esetekben a sebességnövekedés viszont jelentősen függ az adatok partíciójától: az optimális k_1 és k_2 értékek egy optimalizációs feladatból számolhatók ki. E feladat megoldása speciális esetektől eltekintve nem könnyű, bár segítségével legalább „szuboptimális" megoldás meghatározható.

A szimplex algoritmusnak ezeket a párhuzamos változatait még csak elméletileg vizsgálták meg, gyakorlati eredmények nincsenek róluk, bár ez utóbbi algoritmus végrehajtásához szükséges megfelelő hypercube felépítésű gép legalább már létezik, kereskedelmileg is elérhető. A futási eredmények hiánya miatt azonban ezek a párhuzamos szimplex algoritmusok gyakorlati alkalmazhatóság és gyorsaság szempontjából még nem összehasonlíthatók.

R. Marciano és T. Rus (1988) vizsgálta meg azt, hogyan lehet felhasználni különböző típusú, létező párhuzamos számítógépeket a szimplex algoritmus párhuzamos implementálására, milyen eredmények érhetők el ezeken, melyik típus felel meg legjobban a szimplex algoritmusnak, azaz melyiken lesznek legjobbak különböző teszt-feladatok futási eredményei.

Három, különböző párhuzamos számítógép-osztályokat képviselő számítógépen végeztek implementációkat: az egyik az MPP gép, egy VAX-11/78 ellenőrzése alatt futó tömbprocesszor volt, melynek egyik fő része a kétdimenziós, 128×128-as processzor-tömb, mely egy nagyobb tárhoz, és egy ellenőrző-egységhez kapcsolódik. Ez utóbbi tárolja a skalár adatokat, irányítja a tömbprocesszor működését és végrehajtja a skalár műveleteket. Az implementációban az algoritmus lépései közül a szimplex tábla elemei új értékének kiszámítását végezte a processzor-tömb. A teljes szimplex táblát a host gép 128×128-as részmátrixokra bontotta fel. Az elemek új értékeinek kiszámításához a tömbprocesszor a tárból egymás után beolvasta az egyes részmátrixokat, a processzorok kiszámolták az elemek új értékeit, majd következett az újabb részmátrix beolvasása. Az új értékek meghatározásakor minden processzor megkapta a pivot sornak és oszlopnak azt az elemét, amelyre az általa tárolt elem transzformációjához szüksége volt, ezután a processzorok egyszerre végezték el a megfelelő műveleteket.

A másik vizsgált gép egy aszinkron osztott tárú többprocesszoros rendszer, az Encore Multimax volt. Ebben a rendszerben az algoritmust párhuzamos C nyelven írt programmal hajtották végre: a program saját stack-kel ellátott, így párhuzamosan végrehajtható függvényekből állt: az iterációk mindhárom lépésének végrehajtása ilyen párhuzamosan végrehajtható függvényekkel történt. Egy másik implementáció olyan program volt, amely párhuzamos programfolyamok létrehozását

tette lehetővé a felhasználói programban.

A harmadik az Alliant FX-8 vektor-számítógép volt: ez a memórián és az I/O műveleteket végző processzorokon kívül 8 vektorprocesszorból áll. A párhuzamos működést itt a vektor-műveletek pipeline végrehajtása és a 8 vektorprocesszorral való párhuzamos feldolgozás jelentette. A programot a Fortran nyelv olyan változatával írták, amely a tömbökkel való műveleteket támogatja.

A szimplex algoritmus implementációinak összehasonlításához mindhárom gép esetén ugyanazokat a feladatokat oldották meg. Az eredmények lényegében megfeleltek a várakozásnak: mivel a szimplex algoritmusban mátrix elemein kell műveleteket végrehajtani, ezért a tömbprocesszorok a leghatékonyabbak az algoritmus végrehajtásában, és a vektorprocesszorok használatával is jobb eredmény érhető el, mint általános többprocesszoros gépek esetén. Ugyanakkor kiderült, hogy nagy feladatok esetén a vektorprocesszor rendszer sebessége megközelíti a tömbprocesszorét. Ez annak a következménye volt, hogy nagy méretű feladatok esetén az MPP-ben sok adat-átvitelre volt szükség a processzor-tömb és a tár között, és ez a sebesség csökkenéséhez vezetett.

3. A branch and bound módszer

Az operációkutatásnak a párhuzamosítás szempontjából egyik legígéretesebb és legtöbbet vizsgált területe a kombinatorikus optimalizálás. Ezen a területen nagy szükség van arra, hogy kihasználjuk a párhuzamos végrehajtás előnyeit: az ebbe a körbe tartozó feladatok közül ugyanis sok NP-nehéz, a legrosszabb esetben csak exponenciális időben oldható meg, és bár párhuzamos számítógépek segítségével sem oldhatók meg polinomiális időben (legfeljebb akkor, ha exponenciálisan sok processzor áll rendelkezésre), mégis ezek használatával egyrészt felgyorsítható a feladatok megoldása, másrészt növelhető a megoldható feladatok mérete, kiszélesíthető az elfogadható idő alatt megoldható feladatok köre.

Az egyik legáltalánosabb, számos kombinatorikus optimalizálási probléma megoldására alkalmazható módszer a diszkrét optimalizálásban a branch and bound (BB) algoritmus. Széles körű alkalmazhatósága miatt e módszer párhuzamosításának kérdésével már nagyon sokat foglalkoztak.

A BB módszer párhuzamos implementációjának kérdése alapvetően különbözik a szimplex algoritmusétól. Míg az utóbbiban a hatékony, gyors algoritmusok kis modulszemcsézettségűek, és fő probléma az, hogyan lehetne az adatokat jól elrendezni, megfelelően szétosztani a processzorok között, addig a BB módszer esetén szinte minden implementációban kevesebb, de bonyolultabb felépítésű processzorra van szükség és a processzorok önállóan dolgoznak, kevesebb közöttük a kommunikáció.

Az algoritmus párhuzamosítására már többféle módszert is kidolgoztak. Ezek lényegében abban különböznek egymástól, hogy az algoritmus lépései közül (a szétválasztandó csúcs kiválasztása, a feladat részfeladatokra bontása és a részfeladatok célfüggvény-korlátjának kiszámolása, azoknak a csúcsoknak a meghatározása és elhagyása, amelyek bizonyíthatóan nem vezetnek optimális megoldáshoz) mely(ek)nek

a végrehajtása történik párhuzamosan.

Az egyik leggyakrabban kihasznált lehetőség az, hogy a rendszer minden iterációban egynél több csúcsot választ szét egyszerre, pontosabban annyit, ahány processzor van a rendszerben (ha van egyáltalán annyi szétválasztandó csúcs). Ahhoz, hogy egy ilyen algoritmus hatékonyan működjön, biztosítani kell, hogy a szabaddá vált processzorok megfelelő munkához jussanak, valamint biztosítani kell a korlát-teszt, a megengedettség-teszt és a dominancia-teszt hatékony alkalmazását. A munka elosztása a processzorok között általában a szétválasztandó csúcsok csoportokba — sorokba — rendezésével történik: a csúcsok sorrendjét az határozza meg, hogy milyen szabály alapján történik az algoritmusban a következő szétválasztandó részfeladat kiválasztása. Ha egy processzor szabaddá válik, akkor ezekből a sorokból vesz ki magának munkát, ha pedig valamelyik processzor új feladatot generál, akkor azt ezen sorok valamelyikébe illeszti. Minden sort a processzorok egy meghatározott részhalmaza érhet el. Ha valamelyik sorból minden munka elfogy, akkor azt valamelyik szomszédjából újra lehet tölteni, ha pedig minden sor kiürül, akkor az algoritmus befejeződik. A két leggyakrabban alkalmazott eset az, amikor egyetlen centralizált sor van, melyet minden processzor elérhet, és amikor minden processzornak saját feladat-sora van.

Az egyetlen centralizált sor használatának az előnye az, hogy így a teljes rendszernek jó áttekintése van a még megoldandó munkáról, és könnyű a processzorokat jó részfeladatokkal ellátni. Hátránya viszont, hogy mivel a közös sort egyszerre csak egy processzor érheti el, ezért az algoritmus gyakran szűk keresztmetszetet jelent. Azzal, hogy minden processzornak saját feladat-csoportja van, ez elkerülhető, viszont akkor nem biztos, hogy minden processzor jó részfeladatot dolgoz fel, hiszen lehet, hogy nincs megfelelő részfeladat a saját sorában. Ezen kívül a dominanciateszt is csak korlátozottan alkalmazható a részfeladatok eliminálására, hiszen azok szét vannak szórva az egyes processzorok között. Ezen ugyan teljes információcserével lehet segíteni, de ehhez minden processzornak az összes szerzett ismeretét el kell küldenie a többi processzornak, ez azonban nagyon megnöveli a párhuzamos algoritmus kommunikációs bonyolultságát.

A BB algoritmust először E.A. Pruul implementálta párhuzamosan, 1975-ben (E.A. Pruul, G.L. Nemhauser, R.A. Rushmeier, 1988), az utazó ügynök probléma megoldására. Valódi párhuzamos számítógép akkor még nem állt rendelkezésre, így soros gépen szimulált osztott tárú rendszert. Rendszerében a „processorok" működését egy host processzor irányította, ez tartotta nyilván az aktív részfeladatokat, ez határozta meg, hogy melyik részfeladatot melyik processzor mikor bontsa fel. A slave processzorok végezték a részfeladatok szétválasztását, és az így generált új részfeladatokra a célfüggvény korlátjának kiszámítását. Mivel azonban a szimulációban a processzorok a közös tárat egyszerre is elérhették, memória-konfliktusok nélkül, ezért a szimulált rendszer idealizált párhuzamos számítógépnek felelt meg.

Ehhez az algoritmushoz hasonló párhuzamos algoritmust írt le H.W.J.M. Trienekens (1986) is, az ő rendszerében azonban a slave-processzorok már nem kommunikálhattak egyszerre a master processzorral. Ennek az algoritmusnak a működését elemezve vizsgálta meg A. Bruin, H.G. Rinnooy Kan és H.W.J.M. Trienekens (1988)

azt, hogy a párhuzamos számítógépek speciális felépítése, nevezetesen a processzorok száma és a kommunikációs költség/műveleti költség arány változása hogyan befolyásolja az implementáció eredményeit. Ahhoz, hogy a különböző tényezők hatásait elemezhessék, soros gépen szimuláltak különböző párhuzamos számítógépeket, és ezeken implementáltak egy 75-pontú, a kétdimenziós térben generált euklideszi utazó ügynök problémát BB módszerrel megoldó algoritmust. A szimulált rendszerekben a processzorok száma 2^k volt, ahol $k = 0, 1, \ldots, 6$; bármely két processzor kommunikálhatott egymással, a kommunikációs költség pedig vagy 0 volt, vagy (nem nulla) konstans, vagy az átvitt byte-ok számának lineáris függvénye.

A szimuláció eredményei megerősítették azt, amit az algoritmus elemzése is megjósolt: ha a processzorok száma illetve a kommunikációs költség megnő, akkor a master processzor már nem tudja elég gyorsan új munkával ellátni a slave-processzorokat, így ezek a paraméterek, egy ponton túl, szűk keresztmetszetet okoznak. Ezen az értéken túl a slave-processzorok egyre hosszabb ideig voltak tétlenek és a futási idő is újra nőtt.

M.J. Quinn (1990) azt vizsgálta teszt-feladatok megoldásával, hogy mennyire befolyásolja a párhuzamos algoritmus jellemzőit (a sebességnövekedést, a processzorok kihasználtságát) a centralizált illetve decentralizált sor használata. Az implementációkat NCUBE/7, 64 processzort tartalmazó számítógépen végezte.

Az algoritmus lazán szinkronizált, centralizált sort használó változatában egy kijelölt processzor tárolta a részfeladatokat, ez küldte el a $p-1$ legjobb célfüggvény-korlátú vizsgálatlan részfeladatot a $p-1$ processzornak, majd minden iteráció végén összegyűjtötte az új részfeladatokat, és beillesztette azokat a többi közé.

A másik változat egy aszinkron, decentralizált sorokat használó algoritmus volt: az eredeti feladatot egy kijelölt processzor kapta meg, egy általános iterációban pedig minden processzor, melynek a részfeladatokat tartalmazó sora nem volt üres, kiválasztotta abból a legjobb célfüggvény-korlátú részfeladatot, szétválasztotta, a kapott részfeladatokat beillesztette a sorba, majd adott számú (de a generált részfeladatoknál kevesebbet) kiválasztott innen, és elküldte azokat a szomszédos processzoroknak: ez biztosította azt, hogy a processzorok kihasználtsága lehetőleg egyensúlyban legyen. (Persze az egyes processzorok iterációi nem voltak szinkronizálva.)

A megoldandó teszt-feladat egy 30-pontú utazó ügynök probléma volt, melyben az élsúlyok asszimetrikusak és 0 és 90 közötti egyenletes eloszlású véletlen egész számok voltak. A processzorok száma az implementációkban mindkét teszt-feladat esetén 2^d volt, $d = 0, 1, \ldots, 6$ mellett.

A processzorok számának a növekedése, ahogy várható volt, a megoldási sebesség növekedéséhez vezetett: míg két processzor esetén a sebességnövekedés (a soros implementációhoz képest) 1 és 2 között volt, addig 32 processzor esetén 8 és 12.9 között. Ugyanakkor megfigyelhető volt, hogy a processzorok számának további növekedése általában már a kommunikációs igények túlzott megnövekedéséhez és a sebességnövekedés csökkenéséhez vezetett. Az eredmények azt is jelezték, hogy a centralizált sor használata valóban szűk keresztmetszetet jelent, így nem érhető el vele akkora sebességnövekedés, mint decentralizált sor alkalmazásával.

Számos egyéb implementáció is van már a BB algoritmusnak erre a párhuzamos változatára: különösen a durván szemcsézett hypercube többprocesszoros számítógépeken vizsgálták meg különféle változatait (ld. M.J. Quinn (1987), S. Anderson és M.C. Chen (1987), T.S. Abdelraham és T.N. Mudge (1988), E.W. Felten (1988), F.S. Tsung és M.H. Ma (1988), R.P. Pargas és D.E. Wooster (1988)). Van már olyan általános diszkrét feladatok megoldására kidolgozott párhuzamos algoritmus is, amely szintén ezen a párhuzamos BB módszeren alapszik (R.L. Boehning, R.M. Butler, B.E. Gillett (1988)): ebben a csúcsok felbontása vágás-módszerrel történik, a célfüggvény-korlátok meghatározása pedig szimplex algoritmussal.

Ezeknek a párhuzamos BB algoritmusoknak a vizsgálatai különös szabálytalanságokat mutattak a futási sebességben. Az algoritmus párhuzamos végrehajtásától azt várnánk, hogy több processzor gyorsabban old meg egy feladatot, mint kevesebb, az elért sebességnövekedés viszont a processzorok számának arányánál kevesebb. Előfordulhat azonban, hogy n_2 processzor lassabban old meg egy feladatot, mint n_1, ahol $n_2 > n_1$ (káros anomália), és az is lehetséges, hogy a sebességnövekedés n_2/n_1-nél is nagyobb (gyorsulási anomália). A káros anomáliák a prosszorok közötti teljes információcsere hiányával magyarázhatók: az egyes processzorok kevesebb és „rosszabb" adatokkal rendelkezhetnek, mint a soros esetben egyetlen processzor, emiatt több feladatot vizsgálnak meg feleslegesen. Ugyanakkor több processzor egyszerre több feladatot tud megvizsgálni, hamarabb tud jó információkhoz jutni, s ha ezek az információk jókor kerülnek megfelelő processzorokhoz, akkor kevesebb részfeladat megvizsgálására lehet szükség, az algoritmus a processzorok számának arányával nagyobb sebességnövekedést is elérhet.

Sokan vizsgálták már a BB-módszernek ezt a tulajdonságát, s kerestek olyan feltételeket, amelyek mellett elkerülhetők a káros, állandósíthatók a gyorsulást okozó anomáliák (pl. T.-H. Lai és S. Sahni (1983), A. Sprague és T.-H. Lai (1985), G.-J. Li és B.W. Wah (1986)).

A BB algoritmus párhuzamosításának egy másik lehetősége az, hogy az algoritmus lépései közül csak egynek a végrehajtása (pl. a szétválasztandó csúcs meghatározása) történik párhuzamosan, a többi lépést egyetlen processzor végzi. Ez a módszer azonban kevésbé ígéretes, mint az előző: ezt igazolták J. Mohan (1983) eredményei is. Mohan egy olyan változatot implementált és vizsgált meg, amelyben egy master processzor választja ki a szétválasztandó csúcsot, és a kiválasztott csúcs részfeladatokra bontása történik párhuzamosan: a slave-processzorok egy-egy új csúcsot generálnak a kiválasztott csúcsból, és kiszámolják ezekre a célfüggvény korlátját. Az algoritmus többi lépését a master processzor végzi. Mohan az algoritmust utazó ügynök problémák megoldására alkalmazta, az implementációt Cm^* többprocesszoros gépen végezte.

Az implementáció eredményei azonban azt mutatják, hogy az elért sebességnövekedés csak 2 és 4 processzor mellett elfogadható (4 processzor esetén 2.8 volt), nagyobb processzor-szám esetén viszont a sebességnövekedés még csökkent is (8–16 processzor mellett 2.6 volt). A processzorok számának növekedésével ugyanis nőtt a szükséges számítások mennyisége is (hiszen a processzorok számával egyenlő számú részfeladatra bontott fel a rendszer minden csúcsot), a nagy mennyiségű egyszerre

végrehajtott számítás pedig szűk keresztmetszetet okozott.

Az eddig áttekintett párhuzamos BB-módszerek nagy modulszemcsézettségűek és csak durván szemcsézett többprocesszoros rendszeren implementálhatók: végrehajtásukhoz viszonylag kis számú (1000-nél kevesebb), de bonyolult utasítások végrehajtására is képes és jelentős memóriával rendelkező processzorra van szükség (több processzor használata esetén ugyanis vagy túlzottan megnőne a feleslegesen megvizsgált csúcsok száma, vagy túl sok információcserére lenne szükség a processzorok között, és ez az algoritmus hatékonyságát rontaná).

F. Dehne, A.G. Ferreira és A. Rau-Chaplin (1990) dolgozott ki olyan párhuzamos BB algoritmust, amely finoman szemcsézett hypercube felépítésű többprocesszoros rendszeren implementálható: módszerük lényege az, hogy a rendszer együttesen tárolja a BB-fa élő (generált, de még nem eliminált) csúcsait úgy, hogy minden processzor legfeljebb egy csúcsot kezel. Minden iterációban az új csúcsokat tartalmazó processzorok kiszámolják ezek célfüggvény-korlátait (az így kapott globális információkat minden processzor megkapja), majd minden processzor megvizsgálja, törölhető-e az általa tárolt csúcs (és persze azzal együtt annak utódai is), és megállapítja, hogy hány csúcsot kell létrehoznia az általa tárolt részfeladat szétválasztásakor. Ezután a rendszer az addig generált információk alapján létrehozza az új BB-fát (törli a megfelelő csúcsokat és helyet készít az újaknak), és annak a csúcsait újra elosztja a processzorok között: végül elvégzi a megfelelő csúcsok felbontását, elhelyezi az új csúcsokat, és kezdődhet az új iteráció.

Erről az algoritmusról nincsenek implementációs eredmények, az viszont így is látszik, hogy csak olyan feladatokra alkalmazható, melyeknél biztos, hogy az egyszerre létező élő csúcsok száma nem haladja meg a processzorok számát (illetve ha meghaladja, akkor módosítani kell az algoritmust úgy, hogy egy processzor több csúcsot is kezelni tudjon egyszerre).

4. A dekompozíciós módszer

A gyakorlati alkalmazásokban felmerülő nagy méretű optimalizációs feladatok között számos olyan van, mely felbontható kvázi-független részfeladatokra (ezek például különböző időszakoknak, földrajzi területeknek, áruknak felelnek meg). Az ilyen feladatok megoldására kidolgozott dekompozíciós módszerek természetes módon, magas szinten párhuzamosíthatók: a többprocesszoros rendszerekre kidolgozott algoritmusokban a részfeladatokat független slave-processzorok oldják meg, a master feladat megoldását, a koordinációs lépést, a slave processzorok irányítását pedig egy host processzor végzi.

R.J. Chen és R.R. Meyer például hálózati feladatok, nevezetesen konvex célfüggvényű többtermékes folyam feladatok megoldására alkalmazott dekompozíciós módszert. Algoritmusaik a blokk Gauss–Seidel algoritmus párhuzamos változatai voltak. Ebben az eredeti nemlineáris konvex célfüggvényt minden nagyobb, ún. major iterációban egy szeparábilis szakaszonként lineáris függvénnyel közelítik, s a feltételek blokk-szerkezetét kihasználva a kapott feladatra már alkalmazható a de-

kompozíciós elv: a kisebb (minor) iterációkban minden részfeladat a többi $k-1$ termék legújabb értékének felhasználásával oldható meg.

Az egyik párhuzamos eljárásban (Chen és Meyer (1986)) a processzorok minden major iterációban az összes részfeladatot egyszerre, párhuzamosan oldják meg, és a koordinációs lépés az így egyszerre előállított új változó-értékeken alapul. A módszer hátránya azonban, hogy így a szükséges processzorok száma a megoldandó feladattól függ (a processzorok számának meg kell haladnia a részfeladatok számát), másrészt pedig az, hogy a párhuzamos algoritmus konvergenciája lassabb: nem tudja úgy felhasználni a termékek új értékeiből származó információkat, mint a soros változat (hiszen az minden részfeladat megoldásakor a másik $k-1$ termék legújabb értékeit használta fel).

Ezt próbálta kiküszöbölni a másik változat (Chen és Meyer (1988)). Ez az algoritmus annyi blokkot választ ki, ahány processzor van, és az ezekhez tartozó folyam vektorok új értékeit a processzorok párhuzamosan számolják ki a minor iterációkban. Ezután egy master processzor az eredeti célfüggvénynek megfelelően ellenőrzi az új értékek elfogadhatóságát, mialatt a többi processzor a következő blokk-csoportban dolgozik, a régi értékeket használva. Amikor ez a csoport kész, az előző csoport koordinációs ellenőrzése befejeződött, akkor az új értékek felhasználhatók a következő blokk-csoport feldolgozásánál. Látható, hogy egy processzor mellett a soros változatot kapjuk vissza, míg ha annyi processzor van, ahány blokk, akkor az előző párhuzamos algoritmust. A CRYSTAL számítógépen végzett implementációk megerősítették azt, hogy kevesebb processzor használatával gyorsabb a konvergencia (kevesebb a szükséges összes iterációk száma). A módszerrel elérhető sebességnövekedés, a közölt eredmények szerint, körülbelül $\frac{1}{2}p$ (ahol p a processzorok száma), feltéve, hogy minden esetben ugyanannyi iterációt kell végrehajtani. (Persze ha a sebességet az mérte volna, hogy mennyi idő múlva ér el az eredmény adott pontosságot, akkor a konvergencia-tulajdonságok miatt a sebességnövekedés nyilván alacsonyabb lett volna nagyobb processzor-szám mellett).

S.-P. Han és G. Lou (1988) olyan általános konvex programozási feladatok megoldására alkalmazta a dekompozíciós elvet, melynek a célfüggvénye $\mathbb{R}^n$-en differenciálható és egyenletesen konvex, a megengedett pontok halmaza pedig konvex, zárt halmazok metszete. Az ilyen feladatok megoldására egy iteratív eljárás adható, amely a feladatok megoldását kvadratikus programozási részfeladatok megoldására vezeti vissza: a részfeladatok függetlenek, így ezeket több processzorral egyszerre lehet megoldani. S.-P. Han és G. Lou ezt a párhuzamos algoritmust elemezte: megvizsgálták az algoritmus konvergenciáját, és a módszert általánosították lineáris egyenletrendszerek és LP feladatok megoldására is.

Nagy méretű LP feladatok párhuzamos implementálására J.K. Ho, T.C. Lee és R.P. Sundarraj (1988) alkalmazott dekompozíciós algoritmust: ők a Dantzig-Wolfe algoritmus block-angular felépítésű LP feladatokat megoldó párhuzamos változatait dolgozták ki. A teszt-feladatokat CRYSTAL számítógépen oldották meg.

A párhuzamos változatok alapvető jellemzője itt is az, hogy a részfeladatokat a processzorok egyszerre oldják meg: minden részfeladatot más processzor kap meg, a master feladatot pedig egy host gép kezeli. Ehhez persze itt is fel kell tenni, hogy

legalább annyi processzor van, mint amennyi a részfeladatok száma.

A párhuzamos algoritmus standard változatában a host gép küldi el a processzoroknak a megfelelő részfeladatok adatait: ezek minden iterációban megoldják a részfeladatokat, ezzel megállapíthatják, hogy melyik oszlopot kellene a master fealdataiba illeszteni. A host gép megvárja, hogy minden processzor befejezze a munkáját, felállítja a master feladatokat, megoldja és elküldi az új árakat a processzoroknak, azok pedig újra megoldják a részfeladatokat.

Erre a változatra, tíz tesztfeladat megoldása alapján, a következő eredményeket kapták:

a processzorok száma átlagosan	7,4 (5 és 11 között)
a host processzor kihasználtsága	67,1% (42,8 és 92,5 között)
a csúcsok kihasználtsága	15,6% (4,28 és 27,9 között)
a sebességnövekedés	5,12 (2,45 és 7,8 között)

volt.

Mivel az adatokból kitűnt, hogy a processzorok kihasználtsága nem túl magas, ezért különböző stratégiák kidolgozásával próbáltak javítani ezen.

Az egyik a gyorsított feedback stratégia volt: ennek a lényege, hogy amikor a master feladat tétlen, ellenőriz egy megfelelő buffert, tud-e már új oszlopot illeszteni a master feladatba. Amint talál (egy vagy több) oszlopot, azokat beilleszti a master feladatba, és megoldja azt. Közben a részfeladatok további, a master feladattól kapott legutolsó áraknak megfelelő oszlopokat generálnak, és ezeket a bufferben tárolják. Ha a master feladat optimális megoldást talál, az új árakat elküldi a csúcsoknak, melyek a régit azonnal újra cserélik. A master feladat újra megnézi a buffer tartalmát, és új iteráció kezdődik.

Az eredmények megmutatták, hogy ennek a stratégiának az alkalmazásával mind a processzorok kihasználtsága, mind a sebesség nőtt: átlagosan 1,31-szer volt gyorsabb a standard változatnál, a host átlagos kihasználtsága 67,1%-ról 95,8%-ra, a többi processzoré 15,6%-ról 21%-ra nőtt.

Ezek az eredmények is jelzik azonban, hogy az algoritmuson lehetne még javítani: a csúcsok kihasználtsága még így is alacsony, vagyis a részfeladatok a masterhez képest viszonylag „könnyűek". Ezt ki lehetne használni, egy csúcs kezelhetne egyszerre több részfeladatot is: ezzel egyúttal olyan feladatok is megoldhatóvá válnának, melyekben több részfeladat van, mint amennyi a slave-processzorok száma.

5. A relaxációs módszer

A hálózati folyam feladatok megoldására számos módszer született már: általános, széles körben vizsgált kérdés az utóbbi években, hogyan használható ki a módszerekben rejlő párhuzamosság.

A hálózati folyam feladatok egy bő osztályának, a konvex szeparábilis hálózati folyam feladatoknak a megoldására egy Gauss–Seidel típusú relaxációs módszer alkalmazható: a duál feladat ugyanis nemkorlátozott, a célfüggvénye szeparábilis, differenciálható. A relaxációs algoritmus minden iterációjában a p ár-vektorhoz

(duál-változóhoz) egy olyan $\hat{p}_1$ pontot kell meghatározni valamely i-re, amely a duál célfüggvényt a p kezdőpontból a P_i ár mentén minimalizálja.

Az algoritmusnak a párhuzamos implementálására két alapvetően eltérő módszer született: D.P. Bertsekas és D. El Baz (1987) egy aszinkron, S.A. Zenios és J.M. Mulvey (1988) pedig egy szinkron párhuzamos eljárást dolgozott ki.

Az aszinkron algoritmus lényege az, hogy minden p_i árat egy külön processzor ellenőriz. Minden processzor külön-külön bufferekben tárolja minden duál változó legújabb értékét: ha egy processzor a hozzá tartozó árnak új közelítő értékét határozza meg, akkor azt elküldi az összes többi processzornak, azok pedig a régit az újabbra cserélik. Minden processzor a rendelkezésére álló legutolsó ár-vektor felhasználásával számolja ki a hozzá tartozó duál-változó új értékét (vagyis az új pontból kiindulva minimalizálja a duál célfüggvényt a megfelelő irány mentén). Az algoritmusban sem az egyes processzorok számításait, sem pedig a processzorok közötti kommunikációt nem kell szinkronizálni: az egyes processzorokban lévő, más processzorok áraira vonatkozó információ tetszőlegesen „régi" lehet, csak azt kell feltenni, hogy nem marad állandó: minden processzornak kell új árakat kiszámolnia és kommunikálnia az összes többi processzorral. Az is megengedett, hogy egyes processzorok gyakrabban iteráljanak, mint a többiek.

Az algoritmusról bebizonyították, hogy ha lényegében egyetlen optimális ár-vektor létezik, akkor a közelítő értékek ehhez fognak konvergálni. Ha az optimális megoldás nem egyértelmű, akkor a konvergencia a kiindulóponttól függően teljesül.

S.A. Zenios és J.M. Mulvey a relaxációs algoritmus soros implementációjának azt a tulajdonságát használta ki párhuzamos környezetben való implementálásra, hogy az algoritmus egyszerre egy csúcs árát módosítja, úgy, hogy közben csak a szomszédos csúcsoktól kap információkat: így azoknak a csúcsoknak az árai, amelyek nincsenek közvetlenül összekapcsolva, párhuzamosan módosíthatók.

A párhuzamos algoritmusban tehát először a csúcsok olyan részhalmazait kell meghatározni, amelyeken belül nincs él, és az egyes részhalmazok számossága közel P, ahol P a processzorok száma: így az ugyanabban a részhalmazban lévő csúcsokat egyszerre lehet feldolgozni (és a részhalmazok mérete miatt a processzorok kihasználtsága egyensúlyban van). A partíciós probléma megoldására módosított gráfszínezési heurisztika alkalmazható (hiszen az ilyen részhalmazok legkisebb száma a hálózat kromatikus számával egyenlő): a módosított algoritmus úgy választja ki a halmazokat, hogy mindegyik minél több, de P-nél kevesebb csúcsot tartalmazzon. Az implementációban ezt a gráfszínező algoritmust sorosan hajtották végre.

A feladatok elosztása a processzorok között úgy történik, hogy egy részhalmazon belül minden csúcsot minden iterációban más processzorhoz rendelünk, ciklikus módon. Ennek ugyan hátránya, hogy valahányszor új csúcsot dolgoz fel a processzor, minden odavágó adatot be kell olvasni a közös tárból, előnye viszont, hogy az algoritmus még akkor is konvergens lesz, ha csak egy processzor működik.

Az algoritmust egy soros gépen szimulált osztott környezetben implementálták, lényegében egyetlen processzorral futott. Az eredmények azt mutatták, hogy az eredményesen használható processzorok száma felülről korlátozott: nagy processzorszám esetén a gráfszínező algoritmus kevesebb számú nagyobb halmazt állít elő, egy

ponton azonban a rendszer telítődik, és a további processzorok tétlenek maradnak. Ez a szám persze függ a feladattól, a feladat gráfjának a ritkaságától. Megfelelő processzorszám esetén az elért sebességnövekedés megközelítette az Amdahl törvényéből kiszámolható elméleti értéket.

Zenios és Mulvey néhány alapvető jellemző alapján összehasonlította a szinkron és aszinkron algoritmust. Az aszinkron változat egyik hátránya a szinkronnal szemben nyilván az, hogy a konvergenciához ebben a kezőpontnak bizonyos feltételeket ki kell elégíteniük; a másik hátrány az, hogy csak akkor működik jól, ha megfelelő számú processzor áll rendelkezésre, továbbá ha csak egy processzor is meghibásodik, működésképtelenné válik, akkor már nem teljesül a konvergencia. Ezzel szemben a szinkron vátozat esetén akár egy processzoron is helyesen lefut az algoritmus. A szinkron változat viszont nem tudja úgy kihasználni a több processzor nyújtotta lehetőségeket, korlátozott számú processzor esetén tud hatékonyabban működni, és az optimális processzor-szám itt is a feladattól függ; az elért sebesség sem lehet akkora, mint az aszinkron változaté.

6. További párhuzamos algoritmusok

Az eddig áttekintett operációkutatási módszerek, algoritmusok olyanok voltak, amelyeknek több párhuzamos változatát is kidolgozták már: van azonban számos olyan algoritmus, melynek párhuzamosítására eddig csak egyetlen eljárást adtak. Ezek az eljárások nagyon sokféleképpen használják ki az egyes algoritmusokban rejlő párhuzamosságot.

D.L. Miller, J.F. Pekny és G.L. Thompson (1990) például a szállítási feladat megoldására alkalmazott párhuzamos primál algoritmust: ezt egy durván szemcsézett számítógépre, a BBN Butterfly Plus-ra tervezték. A primál algoritmus két fő lépése a pivot elem meghatározása és a pivot-transzformáció: e két lépés közül a pivot elem megkeresését végezte a rendszer párhuzamosan, a pivotálást pedig, a többitől függetlenül, minden processzor elvégezte. Ezzel felgyorsult az első lépés végrehajtása, és csökkent az algoritmus második lépésének kommunikációs bonyolultsága, mert szükségtelenné vált, hogy a pivotálást végző processzornak közölnie kelljen az eredményeket a többi processzorral. Mivel a megoldandó feladat méretének növekedésével a pivot meghatározása válik az algoritmus domináns részévé, ezért nagy feladatok megoldása esetén várható nagy sebességnövekedés.

Deák István (1989) a lineáris célfüggvényű konvex programozási feladatokat megoldó támaszsík-módszerre adott párhuzamos algoritmust: az algoritmus alapja az, hogy több processzor alkalmazásával felgyorsítja a módszer konvergenciáját, azáltal, hogy minden iterációban a megengedett pontok halmazának több támaszsíkját állítja elő. A processzorok közül az egyik a soros változatnak megfelelően működik, előállítja a megfelelő támaszsíkok generálásához szükséges adatokat, ezeket eljuttatja a többi processzornak, azok pedig további támaszsíkokat állítanak elő, amelyet az első proceszor felhasznál a következő LP feladat felállításakor.

G.-H. Chen, M.-S. Chern és J.-H. Jang (1990) az egytermékes bináris hátizsák

feladat megoldására tervezett egy pipeline architektúrát. Az n-változós hátizsák feladat megoldására dinamikus programozási módszerrel n lépésben oldható meg: rendszerükben a lineáris tömb alakban összekapcsolt processzorok mindegyike egy-egy lépésért felelős, és a szükséges számítások pipeline módon hajthatók végre. (A párhuzamos rendszert úgy dolgozták ki, hogy a szükséges processzorok száma n-től független legyen.)

V. Pan és J. Reif (1986) a lineáris legkisebb négyzetek probléma megoldásának párhuzamosítását vizsgálta. Felhasználva, hogy a Karmarkar algoritmus minden iterációjában meg kell oldani egy ilyen feladatot, a módszert ennek az algoritmusnak a párhuzamos végrehajtására is alkalmazták.

A nemlineáris programozási és hálózati folyam-feladatok körében születtek olyan eredmények is, amelyek egyes algoritmusok esetén a legbelső szintű párhuzamosítást, a vektorizációt valósították meg. S.A. Zenios és J.M. Mulvey (1988) az általánosított hálózati folyam-feladatokat megoldó primál Newton módszernél, L. Grandinetti és D. Conforti (1988) a rekurzív kvadratikus programozási módszerrel megoldható nemlineáris programozási feladatok megoldásánál a szükséges mátrix-vektor szorzások elvégzésére alkalmazott vektorizációt, Cray vektor-számítógépeken.

Vannak olyan kutatások a párhuzamos algoritmusok elméletében, amelyek azt vizsgálják, hogy egyes algoritmusoknak melyek a legjobb párhuzamos változatai. Ezek a kutatások általában feltételezik, hogy tetszőlegesen sok processzor illetve tetszőlegesen nagy tár áll rendelkezésre egy feladat megoldásánál. X. Deng (1990) például azt mutatta meg, hogy kétváltozós lineáris programozási feladatokra van olyan optimális párhuzamos algoritmus, amely $n/\log n$ processzorral $\log n$ idő alatt old meg egy n feltételes feladatot. E.D. Karnin az egyfeltételes bináris hátizsák feladat megoldására adott olyan algoritmust, amely $O(2^{n/2})$ művelettel, $O(2^{n/6})$ processzorral és memória sejttel old meg egy n változós feladatot: algoritmusa a két-lineáris és négy-táblás algoritmusok ötvözetének párhuzamos változata.

Befejezés

A leírt algoritmusok köre természetesen nem teljes, nem is lehet az: már eddig is számos eredmény jelent meg a párhuzamos algoritmusokról, operációkutatási, optimalizálási módszerek párhuzamos változatairól és implementációiról, és az utóbbi időben az ilyen irányú vizsgálatok száma egyre nő: így itt csak példákat mutathattunk be arra, hogyan alkalmazhatók a párhuzamos számítógépek ezen a területen. Számítógépes eredményekről viszonylag kevés cikk született: ma még a párhuzamos alogoritmusokkal kapcsolatos elméleti eredmények a legtöbb területen a gyakorlat előtt járnak. A VLSI technológia fejlődése, az egy chipen megvalósított nagyszámú processzor lehetősége azonban hamarosan nagy távlatot nyit a párhuzamos algoritmusok széles körű alkalmazása előtt.

IRODALOM

[1] T. S. ABDELRAHAM and T. N. MUDGE, „Parallel branch and bound algorithms on hypercube multiprocessors", *Proc. Third. Conf. Hypercube Concurrent Computers and Applications* (ACM Press, New York, 1988), 1492–1499.

[2] G. AMDAHL, „The validity of single processzor approach to achieving large scale computing capabilities", *AFIPS Proceedings* **30** (1967), 783–785.

[3] S. ANDERSON, M. C. CHEN, „Parallel branch and bound algorithms on the hypercube", *Hypercube Multiprocessors* (M. T. Heath, ed.) (SIAM Press, Philadelphia, PA, 1987), 309–317.

[4] A. A. BERTOSSI, M. A. BONUCCELLI, „A VLSI implementation of the simplex algorithm", *IEEE Trans. on Comput.* C-36 **(2)** (1987), 241–247.

[5] A. A. BERTOSSI, M. A. BONUCCELLI, „A gracefully degradable VLSI system for linear programming", *IEEE Trans. on Comput.* **38 (6)** (1989), 853–861.

[6] D. P. BERTSEKAS, D. E. BAZ, „Distributed asynchronous relaxation methods for convex network flow problems", *SIAM J. Control and Optimization* **25 (1)** (1987), 74–85.

[7] D. P. BERTSEKAS, J. N. TSITSIKLIS, *Parallel and distributed programming* (Prentice Hall, New York, 1989), 715.

[8] R. L. BOEHNING, R. M. BUTLER, B. E. GILLETT, „A parallel integer linear programming algorithm", *European J. of Oper. Res.* **34** (1988), 393–398.

[9] BÓNA–ERDÉLYI–VAJDA, *Többmikroprocesszoros rendszerek* (Műszaki Könyvkiadó, Budapest, 1986).

[10] A. DE BRUIN, A. H. G. RINNOOY KAN, H. W. J. M. TRIENEKENS, „A simulation tool for the performance evaluation of parallel branch and bound algorithms", *Math. Progr.* **42** (1988), 245–271.

[11] G.-H. CHEN, H.-F. HO, S.-H. LIN, J.-P. SHEU, „Data mapping of linear programming on fixed size hypercubes", *Parallel Comput.* **13 (2)** (1990), 235–243.

[12] G.-H. CHEN, M.-S. CHERN, J.-H. JANG, „Pipeline architedtures for dynamic programming algorithms", *Parallel Comput.* **13** (1990), 111–117.

[13] R. J. CHEN, R. R. MEYER, „Parallel optimization for traffic assignment", *Math. Progr.* **42** (1988), 327–345.

[14] R. J. CHEN, R. R. MEYER, *A scaled trust region method for a class of convex optimization problems* (University of Wisconsin-Madison Computer Sciences Department Tech. Rpt., 1986).

[15] DEÁK ISTVÁN, „Procedures to solve STABIL on a parallel computer", *Technical Report in Industrial Engineering 89-10* (University of Wisconsin–Madison, 1989).

[16] DEÁK ISTVÁN, „Uniform random number generators for parallel computers", *Parallel Computing* **15** (1990), 155–164.

[17] DEÁK ISTVÁN, STRAZICKY BEÁTA, „Párhuzamos számítógépek felépítése", *Alkalmazott Matematikai Lapok* (1991), (*megjelenés alatt*).

[18] F. DEHNE, A. G. FERREIRA, A. RAU-CHAPLIN, „Parallel branch and bound on fine-grained hypercube multiprocessors", *Parallel Comput.* **15** (1990), 201–209.

[19] X. DENG, „An optimal parallel algorithm for linear programming in the plane", *Inform. Process. Lett.* **35** (1990), 213–217.

[20] M. E. DYER, „Linear time algorithms for two-and-three-variable linear programs", *SIAM J. Comput.* **13** (1984), 1–18.

[21] E. W. FELTEN, „Best-first branch and bound on a hypercube", *Proc Third. Conf. on Hypercube Concurrent Computers and Applications* (ACM Press, New York, 1988), 1500–1504.

[22] M. J. FLYNN, „Very high speed computing systems", *Proc. IEEE* **54 (12)** (1966), 1271–1277.

[23] L. GRANDINETTI, D. CONFORTI, „Numerical comparision of nonlinear programming algorithms on serial and vector processors using automatic differentiation", *Math. Progr.* **42** (1988), 375–389.

[24] R. GURKE, „The approximate solution of the Euclidean traveling salesman problem on a CRAY X-MP", *Parallel Computing* **8** (1988), 177–183.

[25] S. P. HAN, G. LOU, „A parallel algorithm for a class of convex programs", *SIAM J. Control and Optim.* **26** (1988), 345–355.

[26] J. K. HO, T. C. LEE, R. P. SUNDARRAJ, „Decomposition of linear programs using parallel computation", *Math. Program.* **42** (1988), 391–405.

[27] R. W. HOCKNEY, C. R. JESSHOPE, *Parallel computers: architecture, programming and algorithms* (Adam Hilger, Bristol, England, 1981).

[28] R. W. HOCKNEY, „Parallel computers: architecture and performance", in Parallel computing 85, (M. Feilmeier, G. Joubert, U. Schendel, eds.) (Elsevier, 1986), 33–69.

[29] E. D. KARNIN, „A parallel algorithm for the knapsack problem", *IEEE Trans. on Computer* **C-33 (5)**, (1984), 404–408.

[30] T. H. LAI, S. SAHNI, „Anomalies in parallel branch-and-bound algorithms", *Proc. 1983 Int. Conf. Parallel Processing, IEEE* (1983), 183–190.

[31] T. H. LAI, A. SPRAGUE, „Performance of parallel branch-and-bound algorithms", *IEEE Trans. on Comput.* **C-34 (10)** (1985), 962–964.

[32] G. J. LI, B. W. WAH, „Coping with anomalies in parallel branch-and-bound algorithms", *IEEE Trans. on Comput.* **35** (1986), 568–573.

[33] T. MANUEL, „Supercomputers: the proliferation begins", *Electronics,* March **3** (1988), 51–56.

[34] R. MARCIANO, T. RUS, „Parallel implementation of the simplex algorithm", *Proceedings. The 2-nd Symposium on the Frontiers of Massively Parallel Computation* (Fairfax, USA, 1988), 85–92.

[35] N. MEGIDDO, „Linear-time algorithm for linear programming in R^3 and related problems", *SIAM J. Comput.* **12** (1983), 759–776.

[36] D. L. MILLER, J. F. PEKNY, G. L THOMPSON, „Solution of large dense transportation problems using a parallel primal algorithm", *Oper. Res. Lett.* **9** (1990), 319–324.

[37] J. MOHAN, „Experience with two parallel programs solving the traveling salesman problem", *Proc. 1983 Int. Conf. Parallel Processing,* (IEEE, 1983), 191–193.

[38] K. ONAGA, H. NAGAYASU, „A wavefront-driven algorithm for linear programming on dataflow processor-arrays", *Proc. International Computer Symposium* (1984), 739–746.

[39] V. PAN, J. REIF, „Eifficient parallel linear programming", *Oper. Resh. Lett.* **5** (1986), 127–135.

[40] R. P. PARGAS, D. E. WOOSTER, „Branch and bound algorithms on a hypercube", *Proc. Third. Conf. Hypercube Concurrent Computers and Applications* (ACM Press, New York, 1988), 1514–1519.

[41] E. A. PRULL, G. L. NEMHAUSER, R. A. RUSHMEIER, „Branch-and-bound and parallel computation: a historical note", *Oper. Res. Lett.* **7 (2)** (1988), 65–69.

[42] M. J. QUINN, *Designing efficient algorithms for parallel computers* (McGraw-Hill Book Co., New York, 1987).

[43] M. J. QUINN, „Implementing best-first branch and bound algorithms on hypercube multicomputers", *Hypercube Multiprocessors* (M. T. Heath, ed.) (SIAM Press, Philadelphia, PA, 1987), 318–326.

[44] M. J. QUINN, „Analysis and implementation of branch-and-bound algorithms on a Hypercube Multicomputer", *IEEE Trans. on Comput.* **39** (1990), 384–387.

[45] H. W. J. M. TRIENEKENS, „Parallel branch and bound on an MIMD system", *Report 8640/A* (Economic Institute Erasmus University Rotterdam, 1986).

[46] R. R. TRIPPI, E. TURBAN, „Parallel processing and OR/MS", *Computers Ops. Res.,* **18 (2)** (1991), 199-210.

[47] F. S. TSUNG, M. H. MA, „A dynamic load Balancer for parallel branch and bound algorithm", *Proc. Third. Conf. Hypercube Corcurrent Computers and Applications* (ACM Press, New York, 1988), 1505–1513.

[48] B. W. WAH, Y. W. E. MA, „MANIP – A multicomputer architecture for solving extremum-search problems", *IEEE Trans. Comput.* **33 (5)** (1984), 377–390.

[49] B. F. WANG, G. H. CHEN, „Two-dimensional processor array with a reconfigurable bus system is at least as powerful as CRCW model", *Inform. Process. Lett.* **36** (1990), 31–36.

[50] S. A. ZENIOS, J. M. MULVEY, „A distributed algorithm for convex network optimization problems", *Parallel Computing* **6** (1988), 45-56.

[51] S. A. ZENIOS, J. M. MULVEY, „Vectorization and multitasking of nonlinear network programming algorithms", *Math. Progr.* **42** (1988), 449-470.

(Beérkezett: 1991. július 16.)

DEÁK ISTVÁN ÉS BÁLINT ERZSÉBET
BME VILLAMOSMÉRNÖKI KAR MATEMATIKA TANSZÉK
1111 BUDAPEST, MŰEGYETEM RKP. 3.

PARALLEL COMPUTERS: OPTIMIZATION SOFTWARE

E. BÁLINT and I. DEÁK

After a short description of existing architectures a survey of existing optimization software is presented. Three topics constitute the main body of the paper: variants of the simplex algorithm, nonlinear programming (and network flows) and discrete programming. A bibliography of more than 40 papers completes the survey.

Alkalmazott Matematikai Lapok 17 (1993) 19–40

EGY INTERVALLUM-ARITMETIKÁN ALAPULÓ ALGORITMUS A SZINTHALMAZOK KORLÁTAINAK MEGKERESÉSÉRE*

CSENDES TIBOR

Szeged

Egy új, intervallum-aritmetikán alapuló algoritmust mutatunk be, amely alkalmas adott szinthalmazhoz egy azt szorosan tartalmazó n-dimenziós intervallum megkeresésére. Az implementálás során olyan módosításokat vezettünk be az eljáráson, amelyek lényegesen javították a hatékonyságát. A numerikus hatékonyságot standard globális optimalizálási feladatokkal teszteltük. Ezen vizsgálatok szerint annak ellenére, hogy algoritmusunk garantált megbízhatóságú, az eljárás versenyképes a hagyományos módszerekkel a felhasznált CPU-időt és a függvényhívások számát tekintve.

Bevezetés

A paraméterbecslési feladat szokásos alakja a következő:

$$\min_{x \in X} f(x) \tag{1}$$

ahol $f(x) : \mathbb{R}^n \to \mathbb{R}$,

$$f(x) = \sqrt{\sum_{i=1}^{m} (f_i - f_{\mathrm{mod}}(i, x))^2}$$

$f_i \in \mathbb{R}$ adatpont; $f_{\mathrm{mod}}(i, x)$ pedig a modell-függvény, $i = 1, 2, \ldots, m$; $m > 0$ egész. $X \subseteq \mathbb{R}^n$ kompakt halmaz, a lehetséges megoldások halmaza. X a legtöbb esetben egy n-dimenziós intervallum: $X = \{x \in \mathbb{R}^n : a_j \leq x_j \leq b_j\}$; $a_j, b_j \in \mathbb{R}$; $j = 1, 2, \ldots, n$. Ebben az esetben a lehetséges megoldások halmazát megadó feltételek egyszerű korlátok a paraméterekre. Az (1) feladat célfüggvénye a modellfüggvény optimális legkisebb négyzetes értelmű illesztését adja meg az f_i adatpontokhoz.

A paraméterbecslési feladatot a numerikus matematikához, és azon belül az optimalizáláshoz, vagy más megközelítésben az operációkutatáson belül a matematikai programozáshoz szokás sorolni. Ezen területeket hazánkban is kiterjedten kutatják, számos értékes könyv, dolgozat jelent meg magyar nyelven is [1, 9, 10, 15, 17, 21]. Ahogy korábban megmutattuk [4], a paraméterbecslési feladat négyzetösszeg alakja

*A kutatások anyagi feltételeit részben az 1074/1987 és 2879/1991 sz. OTKA pályázat és a 314/108/004/8 sz. DAAD ösztöndíj biztosították.

nem jelent megszorítást a helyi minimumpontok halmazának szerkezetére, tehát a nemlineáris optimalizálás számos általános eredménye közvetlenül alkalmazható feladatunkra is.

A paraméterbecslési feladat fontosságát az is jelzi, hogy a 100 legtöbb hivatkozást kapott természettudományi közlemény között az egyetlen matematikai cikk D. W. MARQUARDT publikációja a nemlineáris paraméterbecslésről [20]. A feladat nehézségét pedig az jellemzi, hogy például a Fermat-sejtést is meg lehet fogalmazni paraméterbecslési feladatként [22].

A csak a célfüggvényt és annak deriváltjait kiszámító szubrutinokra támaszkodó algoritmusok nem abszolut megbízhatók [4]. Ezek hatékonysága a gyorsan növekvő számítógépes kapacitások miatt leértékelődik, és egyre fontosabbá válnak a lassúbb, de abszolút megbízható eljárások. Ezek olyan további, a feladatokra vonatkozó információkra támaszkodnak, mint a célfüggvényhez tartozó Lipschitz-konstans [23], az intervallum-aritmetika koncepciójára alapozott befoglaló függvények [26], illetve a célfüggvény explicit képlete [7, 24].

Doldozatunkban olyan algoritmust tárgyalunk, amely intervallum-aritmetikával számított befoglaló függvényeket használ. Jóllehet a valós műveletek kiterjesztése intervallumokra a hatvanas évektől ismert a numerikus matematikában, az ilyen programfejlesztést támogató programozási nyelvek fejletlensége és a magyar nyelvű szakirodalom hiánya miatt az intervallum-aritmetika koncepciója hazánkban jórészt ismeretlen. Az ezzel kapcsolatos alapvető fogalmakat a Függelékben foglaltuk össze; ezek ismerete hasznos, de nem nélkülözhetetlen. Az ismertetett eljárások mind tárgyalhatók kizárólag a befoglaló függvény fogalmára támaszkodva, amelynek egy lehetséges implementációs módszere használja az intervallum-aritmetikát.

Legyen I^n az n-dimenziós kompakt intervallumok tere. Ekkor az $F(X) : I^n \to I$ az $f(x)$ n-változós valós függvény befoglaló függvénye, ha $f(x) \in F(X)$ érvényes minden $x \in X$ pontra és $X \in I^n$ intervallumra [25]. Másszóval, $F(X)$ akkor befoglaló függvénye $f(x)$-nek, ha az $F(X)$ intervallum tartalmazza az $f(x)$ értékkészletét (amit $\overline{f}(X)$-szel jelölünk) X-en. A nagybetű az illető függvény befoglaló függvényét jelöli, pl. $F_j'(X)$ lesz a $\partial f(x)/\partial x_j$ parciális derivált befoglaló függvénye. A továbbiakban feltételezzük, hogy minden célfüggvény és deriváltjai befoglaló függvénye izoton (azaz $X \subseteq Y$-ból következik $F(X) \subseteq F(Y)$). Azt mondjuk, hogy az $F(X)$ befoglaló függvény $\alpha > 0$ rendű, ha létezik olyan c valós konstans, hogy $w(F(X)) - w(\overline{f}(X)) \leq cw(X)^\alpha$ teljesül minden $X \in I^n$-re, ahol $w(X)$ az X intervallum szélessége [25].

Ebben a dolgozatban a naiv intervallum-aritmetikát, vagy másszóval a természetes intervallum-kiterjesztést használjuk a befoglaló függvények előállítására. A befoglaló függvényeket tehát úgy építjük fel, hogy minden, az illető valós függvényben előforduló művelet vagy standard függvény (pl. $\sin(x)$) helyett a megfelelő intervallum-műveletet, illetve intervallum-függvényt alkalmazzuk. Az így előálló befoglaló függvények izotonok, és $\alpha = 1$ rendűek [25].

1. Szinthalmaz korlátainak megkeresése

Az (1) paraméterbecslési feladat sajátossága, hogy jól azonosítható modell esetén, és ha a mért adatok pontossága közelítőleg ismert, az optimum értéke jól becsülhető. Ezt a tulajdonságot hagyományos paraméterbecslési eljárások alkalmazásakor úgy szokás kihasználni, hogy a becsült optimumtól lényegesen eltérő minimumértéket (mint a valószínűleg nem globálisat) nem fogadják el, és új keresést indítanak.

A globális minimumpont ismeretén túl, különösen paraméterbecslési feladatoknál fontos ismerni a célfüggvény érzékenységét a paraméterek változtatására a globális minimumpont környezetében. A szokásos érzékenységvizsgálati eljárás az egyes paraméterekre konfidencia-intervallumokat ad meg olyan helyi információk alapján, mint például a Hesse-mátrix közelítő értéke a globális minimumpontban [19]. Ennek a módszernek viszont az a hátránya, hogy a felhasználó nem kap igazi képet az S_{f_ε} szinthalmazról (ami olyan pontok halmaza, amelyekre $f(x) \leq f_\varepsilon$, ahol $f_\varepsilon > f(x^*)$ egy, az $f(x^*)$ globális minimumhoz közeli érték).

Egy ilyen halmaz általában természetesen nem jellemezhető véges sok valós számmal, tehát pontos megadása sem lehetséges egy véges algoritmussal. Intervallum-aritmetika és befoglaló függvények segítségével viszont lehet garantált korlátokat adni ehhez a szinthalmazhoz. A következőekben egy olyan eljárást ismertetünk és vizsgálunk, amely a szinthalmazt korlátozó X^* intervallumot határozza meg a célfüggvény és gradiense befoglaló függvénye felhasználásával [5, 6].

A bemutatandó módszer egy X^0 kiindulási intervallumban keresi meg azt a lehető legszűkebb intervallumot, amely még tartalmazza a keresett S_{f_ε} szinthalmazt. Az eljárás által használt befoglaló függvényekről feltesszük, hogy folytonosak [25] és izotonok. Legyen $F(X)$ az $f(x)$ függvény befoglaló függvénye, $F'_j(X)$ az $f'_j(x) = \partial f(x)/\partial j$ parciális derivált befoglaló függvénye, f_ε $(> f(x^*))$ pedig a transzformáció paramétere. Jelöljük az X intervallum lapjait a következőek szerint:

$$\underline{X}_j = X_1 \times X_2 \times \cdots \times X_{j-1} \times \min X_j \times X_{j+1} \times \cdots \times X_n$$

és

$$\overline{X}_j = X_1 \times X_2 \times \cdots \times X_{j-1} \times \max X_j \times X_{j+1} \times \cdots \times X_n$$

$j \in (1, 2, \ldots, n)$. Tekintsük a következő transzformációkat:

$$t_j(X) = \min X_j - \frac{\min F(\underline{X}_j) - f_\varepsilon}{\min F'_j(X)},$$

$$t^j(X) = \max X_j - \frac{\min F(\overline{X}_j) - f_\varepsilon}{\max F'_j(X)}.$$

Jelöljön $T_j(X)$ egy olyan X' intervallumot, hogy

$$\min X'_j = t_j(X), \qquad \max X'_j = \max X_j$$

és

$$\min X'_i = \min X_i, \quad \max X'_i = \max X_i \quad i = 1, 2, \ldots, j-1, j+1, \ldots, n.$$

Hasonló módon legyen $T^j(X)$ egy olyan X' intervallum, hogy

$$\min X'_j = t^j(X), \qquad \max X'_j = \max X_j$$

és

$$\min X'_i = \min X_i, \quad \max X'_i = \max X_i \quad i = 1, 2, \ldots, j-1, j+1, \ldots, n.$$

Ezek a transzformációk arra valók, hogy a kiindulási X^0 intervallumot olyan X^* intervallumba alakítsák, hogy $f_\epsilon \in F(\underline{X}_j^*)$, $f_\epsilon \in F(\overline{X}_j^*)$ és $S_{f_\epsilon} \subseteq X^*$. Az X^* intervallum tehát olyan, hogy minden lapján az $F(X)$ befoglaló függvény értéke tartalmazza f_ϵ-t, és az S_{f_ϵ} szinthalmaz mégis teljes egészében benne van X^*-ban. Ennél többet nem várhatunk a transzformációktól, hiszen $f(x)$-nek és deriváltjainak befoglaló függvénye áll csak a rendelkezésükre. Másrészt ha $F(X) = \overline{f}(X)$ minden X-re, akkor a fenti relációk biztosítják, hogy X^* lapjai érintik az S_{f_ϵ} halmazt.

Az alábbi feltételek együttes teljesülése mellett a T_j és T^j transzformációk csökkentik az X argumentum-intervallum méretét.

A1. $f_\epsilon < \min F(\underline{X}_j)$, és $f_\epsilon < \min F(\overline{X}_j)$,
A2. $\min F'_j(X) < 0$, és $\max F'_j(X) > 0$,
A3. $f_\epsilon \geq f(x^*)$,

ahol x^* egy globális minimumpont X belsejében: $\forall x \in X \quad f(x) \geq f(x^*)$ és $\min X_i < x_i^* < \max X_i \quad i = 1, 2, \ldots, n$.

Vegyük észre, hogy $\min F(\underline{X}_j)$ és $\min F(\overline{X}_j)$ lényegében ugyanaz a függvény, az egyetlen különbség, hogy a $\min F(\underline{X}_j)$ függvényben szereplő $\min X_j$ változó helyett $\max X_j$ van $\min F(\overline{X}_j)$-ben. Ezt a tulajdonságot hangsúlyozandó, egységes jelölést használhatunk mindkettőre:

$$m_{j,X}(y) = \min F(Y)$$

ahol $Y = X_1 \times X_2 \times \cdots \times X_{j-1} \times y \times X_{j+1} \times \cdots \times X_n$ egy $\underline{X}_j$-vel párhuzamos intervallum. Rögzített j-re és X-re az $m_{j,X}(y) : \mathbb{R} \to \mathbb{R}$ egy szokásos valós függvény. Erre természetesen érvényes $\min F(\underline{X}_j) = m_{j,X}(\min X_j)$, és $\min F(\overline{X}_j) = m_{j,X}(\max X_j)$. A továbbiakban feltesszük, hogy $m_{j,X}(y)$, mint y függvénye, egyenletesen Lipschitz-folytonos minden X és $j = 1, 2, \ldots, n$ értékre. Feltesszük továbbá, hogy

A4. léteznek olyan $\underline{L}_j(X)$ és $\overline{L}_j(X)$ valós számok, hogy minden $y', y'' \in X_j$-re, amelyre $y' < y''$ teljesül,

$$\underline{L}_j(X)(y'' - y') \leq m_{j,X}(y'') - m_{j,X}(y') \leq \overline{L}_j(X)(y'' - y'),$$

és

$$\underline{L}_j(X) \in F'_j(X), \quad \overline{L}_j(X) \in F'_j(X)$$

érvényes minden intervallumra a továbbiakban. Mivel az algoritmusunk által generált intervallum-sorozat monoton csökkenő, ezért az A4 tulajdonság öröklődő. A T_j és T^j transzformációkat akkor is végre lehet hajtani, ha A4 nem teljesül, de — mint ahogy látni fogjuk — szükséges ez a tulajdonság ahhoz, hogy $f_\epsilon = \min F(\underline{X}^*_j)$ és $f_\epsilon = \min F(\overline{X}^*_j)$ igaz legyen az X^* határérték-intervallumra. Az A4 feltételnek az a jelentése, hogy az $F(X)$ és $F'(X)$ befoglaló függvények nem lehetnek teljesen függetlenek. Minden Lipschitz-folytonos $m_{j,X}(y)$ függvényre az $\underline{L}_j(X)$ és $\overline{L}_j(X)$ konstansok végesek, tehát minden $F'_j(X)$ megnövelhető úgy, hogy tartalmazza őket. Az $F'_j(X)$ ilyen megváltoztatása nem befolyásolja az eredmény-intervallumot, csak a konvergencia-sebességet csökkenti.

Az alábbiakban mutatunk egy eljárást, amellyel a természetes intervallum-kiterjesztéssel létrehozott $F(X)$-hez lehet olyan $F'(X)$ befoglaló függvényt összeállítani, amely kielégíti az A4 feltételt. Tekintsük először a következő példát: $f(x) = x_1 \sin(x_1 x_2)$, és legyen ennek befoglaló függvénye $F(X) = X_1 \mathrm{SIN}(X_1 X_2)$. Ekkor

$$m_{1,X}(y) = \min(y\ \mathrm{SIN}(yX_2)) = y \min(\mathrm{SIN}(yX_2)).$$

Itt $\min(\mathrm{SIN}(yX_2))$ értéke -1, $\sin(y \min X_2)$, vagy $\sin(y \max X_2)$ lehet. Mindegyik esetben lehet adni egy olyan $C \in X_2$ valós konstanst, hogy $m_{1,X}(y) = y \sin(yC)$. Bár C értéke változhat y függvényében, $C \in X_2$ mindig teljesül. Található olyan C konstans is, amely egy y-t tartalmazó pozitív szélességű intervallumra érvényes, ha — mint feltételeztük — $f(x)$ folytonosan differenciálható. Bonyolultabb függvény esetén előfordulhat, hogy egy x_k változót a képletben különböző helyeken különböző konstansokkal kell pótolni, mégis mindegyik X_k-ba kell, hogy tartozzon.

Ezekután az $F'_j(X)$-et összeállító eljárás a következő: pótoljuk az $x_1, \ldots, x_{j-1}$, $x_{j+1}, \ldots, x_n$ változók minden előfordulását különböző C_i konstansokkal. Számítsuk ki a parciális deriváltat formálisan az x_j változó szerint, és írjuk be a C_i konstansok helyett az X_k intervallumot, ha a C_i az x_k változót helyettesítette az előző lépésben. A konstansok használata meggátolja, hogy az intervallum-aritmetikában nem megengedett egyszerűsítést hajtsunk végre. Könnyen belátható, hogy ez az eljárás az $\underline{L}_j(X)$-et és $\overline{L}_j(X)$-et tartalmazó, szűk intervallumot szolgáltatja.

A transzformációs lépésekből összeállított algoritmus vizsgálata előtt tanulmányozzuk a T_j és T^j transzformációk tulajdonságait. A következő állítások főleg a T_j tanszformációra vonatkoznak majd, és csak a fontos eltéréseket említjük meg zárójelben a T^j esetére. Az utóbbira a bizonyítások hasonlóak, a megfelelő módosításokkal.

1. LEMMA. *Ha az* A1–A4 *feltételek teljesülnek valamely* $j \in (1, 2, \ldots, n)$*-re* X*-ben, akkor az* $X' = T_j(X)$ *(vagy az* $X' = T^j(X)$*) intervallum tartalmazza az összes* $x \in X$ *pontot, amelyre* $f(x) \leq f_\epsilon$*, és érvényes továbbá* $f_\epsilon \leq \min F(\underline{X}'_j)$ *és* $f_\epsilon \leq \min F(\overline{X}'_j)$.

Bizonyítás. Tegyük fel, hogy $x' \in X$, $f(x') \leq f_\epsilon$ és x' nincs X'-ben. Ekkor

$$(x'_j - \min X_j) < \frac{f_\epsilon - \min F(\underline{X}_j)}{\min F'_j(X)},$$

és

$$f(x') \geq \min F(\underline{X}_j) + \min F'_j(X)(x'_j - \min X_j)$$

miatt, mivel $\min F'_j(X) \leq 0$, azt kapjuk, hogy $f(x') > f_\epsilon$, ami ellentmondás. Felhasználva az A4 feltételt, rövid számolással adódik, hogy

$$\min F(\underline{X}'_j) \geq \min F(\underline{X}_j) - \frac{\min F(\underline{X}_j) - f_\epsilon}{\min F'_j(X)} \underline{L}_j(X) \geq f_\epsilon,$$

és

$$\min F(\overline{X}'_j) \geq \min F(\overline{X}_j) - \frac{\min F(\overline{X}_j) - f_\epsilon}{\max F'_j(X)} \overline{L}_j(X) \geq f_\epsilon. \qquad \square$$

2. LEMMA. *Ha az* A1–A3 *feltételek teljesülnek egy* X *n-dimenziós intervallumban egy* $j \in (1, 2, \ldots, n)$*-re, akkor a* T_j *(vagy a* T^j*) transzformációval kapott* X' *intervallumra az* $F'_i(X')$ *intervallum tartalmazza a nullát minden* $i = 1, 2, \ldots, n$*-re.*

Bizonyítás. Az 1. Lemma alapján az X azon pontjai, amelyekre az $f(x)$ függvényérték nem nagyobb, mint f_ϵ, benne vannak az X' intervallumban. Másrészt, $f(x)$-nek legalább egy x^1 helyi minimumpontja van X'-ben (pl. a globális minimumpont). Erre $F(x^1) \leq f_\epsilon$, és nyilván $f'_i(x^1) = 0$, ahol $i = 1, 2, \ldots, n$. Ennek befoglaló függvényére pedig

$$\min F'_i(X') \leq 0 \leq \max F'_i(X')$$

minden $i = 1, 2, \ldots, n$-re. $\square$

Az az eset, amikor $\min F'_i(X') = 0$ (vagy $\max F'_i(X') = 0$) valamely $i \in (1, 2, \ldots, n)$-re a T_j (illetve a T^j) transzformáció végrehajtása után, nyilván csak akkor fordulhat elő, ha f_ϵ egyenlő az $f(x^*)$ globális minimummal, $\min F'_i(X') = \min \overline{f}'_i(X')$ ($\max F'_i(X') = \max \overline{f}'_i(X')$), $f(x)$ konstans vagy monoton növekvő (csökkenő) az x_i változó szerint X'-ben, és így $f_\epsilon \in F(\underline{X}'_i)$ (illetve $f_\epsilon \in F(\overline{X}'_i)$).

Legyen X^+ olyan intervallum, hogy $X^+_i = X^0_i \quad i = 1, 2, \ldots, j-1, j+1, \ldots, n$; $\max X^+_j = \max X^0_j$ (a T^j esetén $\min X^+_j = \min X^0_j$) és $\min X^+_j$ a legkisebb olyan érték, hogy $f_\epsilon \in F(\underline{X}^+_j)$ ($\max X^+_j$ a legnagyobb olyan érték, hogy $f_\epsilon \in F(\overline{X}^+_j)$).

1. TÉTEL. *Ha az* A1–A4 *feltételek teljesülnek valamely* $j \in (1, 2, \ldots, n)$, X^0, f_ε*-ra, és az* $F(X)$, $F'(X)$ *befoglaló függvényekre, akkor az* $X^{k+1} = T_j(X^k)$ *(illetve az* $X^{k+1} = T^j(X^k)$*)* $k = 0, 1, 2, \ldots$ *intervallum-sorozat konvergál az* X^+ *intervallumhoz.*

Bizonyítás. Az 1. és 2. Lemma szerint az A1–A3 feltételek érvényesek maradnak az X^k sorozat minden intervallumára, kivéve amikor $X^\ell = X^+$ valamely ℓ egészre, és így az intervallum-sorozat véges. Emiatt elegendő azt az esetet vizsgálni, ha a sorozat végtelen. Tekintsünk egy olyan x'-t, amelyre $\min X_j^0 < x' < \min X_j^+$. A $J = [\min X_j^0, x']$ intervallumban

$$\min F'(X_1^0 \times X_2^0 \times \cdots \times X_{j-1}^0 \times x' \times X_{j+1}^0 \times \cdots \times X_n^0),$$

mint az x' függvénye egy negatív δ_j maximummal (egy δ^j pozitív minimummal) rendelkezik. Hasonlóan,

$$\min F(X_1^0 \times X_2^0 \times \cdots \times X_{j-1}^0 \times x' \times X_{j+1}^0 \times \cdots \times X_n^0),$$

lévén x'-nek folytonos függvénye, egy γ_j (γ^j) minimummal rendelkezik, amely nagyobb, mint f_ε. Ebből következik, hogy a T_j transzformáció lépésköze pozitív J-ben:

$$\frac{f_\varepsilon - \min F(\underline{X}_j^k)}{\min F_j'(X^k)} \geq \frac{f_\varepsilon - \gamma_j}{\delta_j} > 0,$$

illetve a T^j transzformációra:

$$\frac{f_\varepsilon - \min F(\overline{X}_j^k)}{\max F_j'(X^k)} \leq \frac{f_\varepsilon - \gamma^j}{\delta^j} < 0.$$

Mivel ez igaz minden olyan J intervallumra, amelyre $\min X_i^0 \leq \min J$ és $\max J < \min X_i^+$, továbbá X^+ nyilvánvalóan fix-intervalluma a T_j transzformációnak, ezért a tétel állítása bizonyított. Az igazolás hasonlóan történik a T^j transzformációra is. □

A bizonyítás szerint az $X^{k+1} = T_j(X^k)$ (illetve az $X^{k+1} = T^j(X^k)$) $k = 0, 1, 2, \ldots$ intervallum-sorozat monoton: $X^{k+1} \geq X^k$ (illetve $X^{k+1} \leq X^k$). $X^{k+1} = X^k$ akkor és csak akkor teljesül, ha X^k egyenlő a T_j (a T^j) transzformáció egy X^+ fix-intervallumával. Ezért a monotonitást biztosító szokásos kifejezés [16]: $X^{k+1} = X^k \cap T_j(X^k)$ (vagy $X^{k+1} = X^k \cap T^j(X^k)$) nem szükséges.

Az intervallum-sorozat mindig konvergens az A1–A4 feltételeknek megfelelő befoglaló függvényekre. Az X^* eredmény-intervallum viszont függ a befoglaló függvények minőségétől (rend, befoglalási izotonitás stb.). Ezért fontos a befoglaló függvények típusának helyes megválasztása [25], különben X^* lényegesen nagyobb lehet, mint S_{f_ε}.

A T_j és T^j transzformációkra számos algoritmust lehet felépíteni, amely az S_{f_ϵ} szinthalmazt korlátozó, lehető legszűkebb intervallumot keresi meg; először a legegyszerűbb ilyen algoritmust vizsgáljuk:

0. lépés: Legyen $k = 0$, és $X^0 = X$.

1. lépés: Legyen $X^{k+1} = T_1(X^k)$, $X^{k+2} = T^1(X^{k+1})$, $X^{k+3} = T_2(X^{k+2}), \ldots,$ $X^{k+2n} = T^n(X^{k+2n-1})$.

2. lépés: Ha a megállási feltétel teljesül: STOP, különben $k = k+2n$ és folytassa az 1. lépésnél.

A megállási feltétel például

$$\mathrm{vol}(X^k) - \mathrm{vol}(X^{k+2n}) < \eta$$

lehet, ahol $\mathrm{vol}(X)$ az X intervallum térfogatát jelöli, η pedig a felhasználó által meghatározott nem-negatív konstans. Az $\eta = 0$ választás esetén az algoritmus nem áll meg, csak ha az X^* eredmény-intervallumot véges számú lépésben sikerült elérni: $X^k = X^*$ valamely k pozitív egészre.

Algoritmusunk eredmény-intervalluma általában más, mint az 1. Tétel előtt definiált X^+ intervallum, mivel az algoritmus a transzformációkat felváltva minden oldalra hajtja végre. Az algoritmus az A1–A4 feltételek teljesülését az X^0 intervallumban minden $j = 1, 2, \ldots, n$-re megköveteli. Az 1. Tétel bizonyítása szerint egy X intervallum nem lehet a T_j és T^j transzformációk fix-intervalluma, ha az A1–A3 feltételek érvényesek X-re és j-re. Az 1. és 2. Lemma szerint az A1–A3 feltételek érvényesek maradnak véges sok transzformációs lépés után, kivéve azt az esetet, ha $f_\epsilon = \min F(\underline{X}_j^\ell)$, vagy $f_\epsilon = \min F(\overline{X}_j^\ell)$ valamely $j \in (1, 2, \ldots, n)$ és $\ell > 0$ egészekre. A T_j és T^j transzformációkat az utóbbi esetre is lehet definiálni: $\underline{X}_j^{\ell+1} = T_j(X^\ell) = \underline{X}_j^\ell$ és $\overline{X}_j^{\ell+1} = T^j(X^\ell) = \overline{X}_j^\ell$.

Jelöljük az 1. lépésben végrehajtott $2n$ transzformációt együttesen T-vel. Ez folytonos operátor [16], mert minden egyes T_j és T^j transzformáció folytonos. Krawczyk tétele alapján ([16], 2.1 Tétel), az $X^{k+1} = T(X^k)$ sorozat $k = 1, 2, \ldots$ konvergens, és $\lim_{k\to\infty} X^k = X^\infty$ pszeudo-fix intervallum (amelyre $X^\infty \subseteq T(X^\infty)$). Másrészt, ha $\min F(\underline{X}_j^\infty) > f_\epsilon$, vagy $\min F(\overline{X}_j^\infty) > f_\epsilon$ valamely $j \in (1, 2, \ldots, n)$-re, akkor (hasonlóan, mint az 1. Tétel bizonyításában) egyszerű megmutatni, hogy a $\lim_{k\to\infty} X^k \neq X^\infty$ ellentmondáshoz jutunk. Mivel az A4 feltétel miatt $F(\underline{X}_j^k) \geq f_\epsilon$, és $F(\overline{X}_j^k) \geq f_\epsilon$ minden $j \in (1, 2, \ldots, n)$ és $k = 1, 2, \ldots$ esetén, ezért érvényes a

2. TÉTEL. *Ha az $F(X)$ és $F'(X)$ befoglaló függvények folytonosak és izotonok X^0-ban, és az A1–A4 feltételek teljesülnek X^0-ban minden $j \in (1, 2, \ldots, n)$-re, akkor az $X^{k+1} = T(X^k)$ $k = 0, 1, 2, \ldots$ intervallum-sorozat konvergens, $\lim_{k\to\infty} X^k = X^\infty$ olyan intervallum, hogy $\min F(\underline{X}_j^\infty) = \min F(\overline{X}_j^\infty) = f_\epsilon$ érvényes minden $j \in (1, 2, \ldots, n)$-re, és $S_{f_\epsilon} \subseteq X^\infty$.*

2. A konvergencia sebessége és példák

Ebben a szakaszban a T_j transzformáció konvergenciájának sebességét vizsgáljuk (a gondolatmenet könnyen megismételhető T^j-re is). Tegyük fel, hogy a T_j-re teljesülnek az A1–A4 feltételek az X intervallumban. Tekintsük az $X^0 = X$, $X^1 = T_j(X^0)$, $X^2 = T_j(X^1), \ldots$ intervallum-sorozatot, és jelöljük a $\lim_{k\to\infty} X^k$ intervallumot a korábbiaknak megfelelően ismét X^+-szal.

Ha a $\min F(\underline{X}_j)$, mint a $\min X_j$ függvénye, lineáris egy $J=[\min X_j^+ - \delta, \min X_j^+]$ intervallumban ($\delta > 0$), azaz $\min F(\underline{X}_j) = a(\min X_j) + b$, és $\min F_j'(X) = a$ a J intervallumban, akkor létezik olyan k pozitív egész, hogy $\min F(\underline{X}_j^k) = f_\epsilon$, tehát az X^+ határ-intervallumot véges számú lépésben eléri a sorozat. Egy nemlineáris célfüggvény esetén az előző eset csak durva befoglaló függvénnyel érhető el, és ilyenkor X^+ meglehetősen távol kerülhet a szinthalmaztól. A numerikus vizsgálatok során nem találkoztunk ezzel a jelenséggel. A továbbiakban ezért feltesszük, hogy minden X^k intervallumra $\min F(\underline{X}_j^k) > f_\epsilon$. Az előző szakaszban megmutattuk, hogy a $\min X_j^k$, $k = 1, 2, \ldots$ sorozat monoton növő, tehát

$$\frac{\min X_j^+ - \min X_j^{k+1}}{\min X_j^+ - \min X_j^k} = \frac{|\min X_j^+ - \min X_j^{k+1}|}{|\min X_j^+ - \min X_j^k|} < 1 .$$

Ebből következik, hogy a konvergencia sebessége legalább lineáris. Az $X^{k+1} = T_j(X^k)$ általános lépésre adódik, hogy:

$$\min X_j^+ - \min X_j^{k+1} = \min X_j^+ - \min X_j^k + \frac{\min F(\underline{X}_j^k) - f_\epsilon}{\min F_j'(X^k)} .$$

Itt $\min F_j'(X^k) < 0$ az 1. szakasz szerint, tehát

$$(2) \qquad |\min X_j^+ - \min X_j^{k+1}| = |\min X_j^+ - \min X_j^k| - \left| \frac{\min F(\underline{X}_j^k) - f_\epsilon}{\min F_j'(X^k)} \right| ,$$

és az utolsó tag határozza meg a konvergencia sebességét.

Tegyük fel, hogy $\min F(\underline{X}_j)$, mint a $\min X_j$ függvénye, differenciálható a $\min X_j^+$ pontban, legalábbis balról, és hasonlóan, a $\min F(\overline{X}_j)$ differenciálható jobbról a $\max X_j^+$ pontban. Jelöljük ezeket a derivált-értékeket $\underline{f}_j^+$-szal, illetve $\overline{f}_j^+$-szal. Tekintsük először azt az esetet, amikor $\underline{f}_j^+ < 0$. Ekkor az A4 feltétel miatt $\min F_j'(X^+)$ is negatív, és így

$$\lim_{k\to\infty} \left| \frac{\min F(\underline{X}_j^k) - f_\epsilon}{\min F_j'(X^k)(\min X_j^+ - \min X_j^k)} \right| = \frac{\underline{f}_j^+}{\min F_j'(X^+)} \leq 1 .$$

Ezt a (2) egyenletbe helyettesítve kapjuk, hogy

$$C = \lim_{k\to\infty} \frac{|\min X_j^+ - \min X_j^{k+1}|}{|\min X_j^+ - \min X_j^k|} = 1 - \frac{\underline{f}_j^+}{\min F_j'(X^+)}. \tag{3}$$

Az X^k intervallum-sorozat tehát akkor és csak akkor konvergál szuperlineárisan az X^+ intervallumhoz, ha $\underline{f}_j^+ = \min F_j'(X^+) \neq 0$. Különben a konvergencia lineáris, és $0 < C < 1$ érvényes a (3)-ban szereplő C konstansra.

A (3) egyenlet alapján a konvergencia sebessége akkor is lineáris, ha $\underline{f}_j^+ = 0$ és $\min F_j'(X^+) \neq 0$, de ekkor $C = 1$, és ezért az X^k intervallum-sorozat lelassul X^+ közelében.

Az utolsó eset az, amikor $\min F_j'(X^+) = 0$, és így $\underline{f}_j^+ = 0$. Ekkor az intervallum-sorozat konvergencia-sebessége azon múlik, hogy milyen gyorsan tart $\min F(\underline{X}_j^k)$ az f_ε-hoz, és $\min F_j'(X^k)$ nullához, míg X^k tart X^+-hoz. Tegyük fel, hogy

$$\left| \frac{\min F(\underline{X}_j^k) - f_\varepsilon}{\min X_j^+ - \min X_j^k} \right| \leq C_1 (\min X_j^+ - \min X_j^k)^\ell$$

és

$$|\min F_j'(X^k)| \leq C_2 (\min X_j^+ - \min X_j^k)^m$$

minden $k = 0, 1, 2, \ldots$-ra, ahol C_1, C_2 pozitív konstansok, m, ℓ pozitív egészek. Az A4 feltételből következik, hogy $\ell \geq m$. Ha ℓ egyenlő m-mel, akkor a konvergencia lineáris, és $C_1 \leq C_2$ miatt

$$C = \lim_{k\to\infty} \frac{|\min X_j^+ - \min X_j^{k+1}|}{|\min X_j^+ - \min X_j^k|} = 1 - \frac{C_1}{C_2} \leq 1. \tag{4}$$

Különben a konvergencia rendje $\ell - m + 1$.

Ezek a számítások a T^j transzformációra megismételhetők az $\overline{f}_j^+$ konstanssal. Eredményeinket összefoglalva, érvényes a

3. TÉTEL. *Ha az* A1–A4 *feltételek érvényesek* X^0*-ban valamely* $j \in (1, 2, \ldots, n)$*-re, akkor az* X^k *intervallum-sorozat konvergenciájának sebessége a következő:*

1. *Ha* $\min F_j'(X^+)$ *negatív és egyenlő* $\underline{f}_j^+$*-al, akkor a konvergencia szuperlineáris.*
2. *Ha* $\min F_j'(X^+)$ *negatív, de nem egyenlő* $\underline{f}_j^+$*-al, akkor a konvergencia lineáris a* (3) *egyenletben megadott konstanssal.*
3. *Ha* $\min F_j'(X^+) = 0$*, és így* $\underline{f}_j^+ = 0$*, akkor*
   (a) *ha* $\ell = m$*, akkor a konvergencia lineáris a* (4)*-beli konstanssal,*
   (b) *ha* $\ell \neq m$*, akkor a konvergencia rendje* $\ell - m + 1$.

Gyakorlati feladatokban a 2. eset a leggyakoribb, a többi kevésbé valószínű.

A tárgyalt algoritmusunk általános nemlineáris függvényekre (tehát nem csak a négyzetösszeg alakúakra) is használható, ha a megfelelő f_ϵ szintet megadjuk. Az eljárás működését a nagyon egyszerű $f(x) = (x_1 - 1)^2 + (x_2 - 1)^2$ függvénnyel mutatjuk be. Ennek nyilván csak egy helyi minimumpontja van $\mathbb{R}^2$-ben: $x^* = (1,1)^T$, és a globális minimum értéke $f(x^*) = 0$. A természetes intervallum-kiterjesztéssel adódó befoglaló függvénye $F(X) = (X_1 - 1)^2 + (X_2 - 1)^2$. Egy Y intervallum négyzetét úgy definiálhatjuk, hogy legyen

$$\min Y^2 = 0\,,$$

ha $0 \in Y$, és különben

$$\min Y^2 = \min((\min Y)^2,\ (\max Y)^2)\,,$$

és

$$\max Y^2 = \max((\min Y)^2,\ (\max Y)^2)\,.$$

Ez az intervallum-művelet pontos abban az értelemben, hogy minden Y intervallumra $\bar{s}(Y) = Y^2 \subseteq YY$, ahol $\bar{s}(Y) = \{s(y) : y \in Y\}$ az $s(y) = y^2$ függvény értékkészlete (cf. [26]). Az algoritmusunk számára csak $\min F(\underline{X}_j)$ és $\min F(\overline{X}_j)$ szükséges ($j = 1,2$):

$$\min F(\underline{X}_j) = (\min X_j - 1)^2 + \min(X_{3-j} - 1)^2\,,$$
$$\min F(\overline{X}_j) = (\max X_j - 1)^2 + \min(X_{3-j} - 1)^2\,,$$

valamint a gradiens befoglaló függvénye, $F_j'(X) = 2(X_j - 1)$. A transzformációs lépések ezekkel:

$$t_j(X) = \min X_j - \frac{(\min X_j - 1)^2 + \min(X_{3-j} - 1)^2 - f_\epsilon}{2(\min X_j - 1)}$$

és

$$t^j(X) = \max X_j - \frac{(\max X_j - 1)^2 + \min(X_{3-j} - 1)^2 - f_\epsilon}{2(\max X_j - 1)}\,,$$

mivel $\min F_j'(X) = \min(2(X_j - 1)) = 2(\min X_j - 1)$, és hasonlóan a $\max F_j'(X)$-re. Ha az A2 feltétel teljesül X-ben F'-re, akkor $0 \in F_j'(X)$, és ezért $\min(X_j - 1)^2 = 0$ ($j = 1,2$). Ennek megfelelően

$$t_j = \min X_j - \frac{(\min X_j - 1)^2 - f_\epsilon}{2(\min X_j - 1)}\,,$$

és

$$t^j = \max X_j - \frac{(\max X_j - 1)^2 - f_\epsilon}{2(\max X_j - 1)}.$$

Az A4 feltétel teljesül $F(X)$-re és $F'(X)$-re, mert $\underline{L}_j(X) = 2(\min X_j - 1) = \min F'_j(X)$ és $\overline{L}_j(X) = 2(\max X_j - 1) = F'_j(X)$ $j = 1, 2$. Mivel a befoglaló függvények pontosak (azaz értéküik megegyezik a megfelelő értékkészlettel), meg tudjuk határozni az X^* határ-intervallumot: $X_j^* = [1 - \sqrt{f_\epsilon}, 1 + \sqrt{f_\epsilon}]$ $j = 1, 2$.

Legyen $X_j^0 = [0,\ 10]$ $j = 1, 2$, és $f_\epsilon = 0$, ekkor $j = 1, 2$-re

$$X_j^1 = [0{,}500,\ 5{,}500],$$
$$X_j^2 = [0{,}750,\ 3{,}250],$$
$$X_j^3 = [0{,}875,\ 2{,}125],$$

és így tovább. A konvergencia lineáris a $C = 1/2$ konstanssal. A sorozat intervallumainak térfogata rendre $\mathrm{vol}(X^0) = 100$, $\mathrm{vol}(X^1) = 25$, ..., $\mathrm{vol}(X^k) = 100(1/2)^{2k}$.

Abban az esetben, ha $X_j^0 = [0,\ 10]$ $j = 1, 2$ és $f_\epsilon = 1$, $X_j^* = [0,\ 2]$ lesz, és

$$X_j^1 = [0{,}000,\ 5{,}556],$$
$$X_j^2 = [0{,}000,\ 3{,}388],$$
$$X_j^3 = [0{,}000,\ 2{,}403],$$

$j = 1, 2$. A konvergencia most szuperlineáris, és az első néhány intervallum térfogata $\mathrm{vol}(X^0) = 100$, $\mathrm{vol}(X^1) = 30{,}864$, $\mathrm{vol}(X^2) = 11{,}475$, és $\mathrm{vol}(X^3) = 5{,}775$.

A valósághoz közelebbi példa az (1) feladat az $f_{\mathrm{mod}}(i, x) = e^{-x_1 i} + e^{x_2 i}$ modellfüggvénnyel, és az $f_i = e^{-i} + e^i$ $i = -1, 0, 1$ adattal. A globális minimumpont $\mathbb{R}^2$-ben nyilván $x_1^* = x_2^* = 1$, és $f(x^*) = 0$. A természetes intervallum-kiterjesztéssel felépített befoglaló függvények

$$F(X) = \sum_{i=-1}^{1} ((e^{-i} + e^i) - (e^{-X_1 i} + e^{X_2 i}))^2$$

és

$$F'_1(X) = 2 \sum_{i=-1}^{1} i e^{X_1} ((e^{-i} + e^i) - (e^{-X_1 i} + e^{X_2 i})),$$

$$F'_2(X) = -2 \sum_{i=-1}^{1} i e^{X_2} ((e^{-i} + e^i) - (e^{-X_1 i} + e^{X_2 i})).$$

Itt $e^X = [e^{\min X}, e^{\max X}]$. A transzformációs lépések:

$$t_1(X) = \min X_1 - \frac{\min \sum_{i=-1}^{1}((e^{-i}+e^{i}) - (e^{-\min X_1 i} + e^{X_2 i}))^2 - f_\varepsilon}{2\min \sum_{i=-1}^{1} i e^{X_1}((e^{-i}+e^{i}) - (e^{-X_1 i} + e^{X_2 i}))},$$

$$t^1(X) = \max X_1 - \frac{\min \sum_{i=-1}^{1}((e^{-i}+e^{i}) - (e^{-\max X_1 i} + e^{X_2 i}))^2 - f_\varepsilon}{2\max \sum_{i=-1}^{1} i e^{X_1}((e^{-i}+e^{i}) - (e^{-X_1 i} + e^{X_2 i}))},$$

$$t_2(X) = \min X_2 - \frac{\min \sum_{i=-1}^{1}((e^{-i}+e^{i}) - (e^{-X_1 i} + e^{\min X_2 i}))^2 - f_\varepsilon}{-2\max \sum_{i=-1}^{1} i e^{X_2}((e^{-i}+e^{i}) - (e^{-X_1 i} + e^{X_2 i}))},$$

$$t^2(X) = \max X_2 - \frac{\min \sum_{i=-1}^{1}((e^{-i}+e^{i}) - (e^{-X_1 i} + e^{\max X_2 i}))^2 - f_\varepsilon}{-2\min \sum_{i=-1}^{1} i e^{X_2}((e^{-i}+e^{i}) - (e^{-X_1 i} + e^{X_2 i}))}.$$

Tekintsük az $X_j^0 = [0,\ 10]$ $j = 1,2$ kiindulási intervallumot. Az $F(X)$ befoglaló függvény minimuma X^0 oldallapjain

$$\min F(\underline{X}_j^0) = 1,180$$
$$\min F(\overline{X}_j^0) = 4,850 \cdot 10^8$$

$j = 1,2$. A gradiens befoglaló függvényének értéke $F_j'(X^0) = [-9,703 \cdot 10^8,\ 9,703 \cdot 10^8]$. $f_\varepsilon = 0$ esetén az $X^1 = T(X^0)$ intervallum:

$$X_1^1 = [1{,}215 \cdot 10^{-9},\ 9{,}500],$$
$$X_2^1 = [2{,}004 \cdot 10^{-9},\ 9{,}500].$$

Ennek térfogata körülbelül 10 százalékkal kisebb, mint X^0-é. Néhány intervallum a generált X^k $k = 1,2,\ldots$ sorozatból:

$$X_1^{10} = [1{,}511 \cdot 10^{-5},\ 5{,}031],$$
$$X_2^{10} = [2{,}467 \cdot 10^{-5},\ 5{,}031],$$

és

$$X_1^{50} = [0{,}989,\ 1{,}010],$$
$$X_2^{50} = [0{,}990,\ 1{,}009].$$

Természetesen az algoritmus által számított intervallum-sorozat és az X^* határintervallum is függ a számábrázolás pontosságától.

3. Implementáció és numerikus teljesítmény

Nagy kiindulási intervallum és bonyolult célfüggvény esetén az $F(X)$ és $F'(X)$ befoglaló függvények meglehetősen durva becslései is lehetnek a megfelelő értékkészleteknek. Emiatt a transzformációs lépések a lehetségesnél lényegesen kisebbek lesznek, és az eredmény-intervallum is távol lesz az illető szinthalmaztól. Algoritmusunk nyilván akkor konvergál a lehető leggyorsabban, ha befoglaló függvényeink megegyeznek a megfelelő értékkészlet-függvényekkel. Ez utóbbi kiszámítása viszont általános esetben nem lehetséges.

A természetes intervallum-aritmetikával előállított egyszerű befoglaló függvényeket viszont lehet javítani, például az intervallum-felosztási (Moore-Skelboe) algoritmussal [25, 26]. Ez az eljárás eredetileg a globális minimum megbízható alsóbecslésére szolgál, tehát közvetlenül a $\min \overline{f}(\underline{X}_j)$, a $\min \overline{f}(\overline{X}_j)$ és a $\min \overline{f}'_j(X)$ becslésének javítására használható ($j = 1, 2, \ldots, n$). (A felülvonás az f felett ismét a megfelelő értékkészlet-függvényt jelöli.) A $\max f'_j(X)$ becslésének javításához az eljárást a $-f'_j(X)$ függvényre kell alkalmazni, és az eredmény előjelét az ellenkezőjére váltani.

Minden említett részprobléma azonban azonos bonyolultságú, mint az eredeti szinthalmaz-korlátozási feladat, tehát nem szabad ezeket teljesen megoldani, csak a befoglaló függvények becslésének javítását érdemes célul tűzni. Emiatt az intervallum-felosztási módszer megállási feltételeit meg kell változtatni, úgy, hogy közel optimális arányt érjünk el a javított befoglaló függvényekkel elért lépéshossz és a teljes eljárás által használt függvényhívások száma (NFEV) között. Az $F(X)$ befoglaló függvényre az intervallum-felosztási módszer álljon meg, ha

1. a természetes intervallum-kiterjesztéssel kapott $\min F(\underline{X}_j)$ nagyobb, mint f_ε,
2. az $\overline{f}$ előző iterációban kapott F_{old} becslése nagyobb, mint f_ε, és az F_{act} aktuális becslésre

$$F_{\text{act}} < \frac{(F_{\text{old}} - f_\varepsilon)\,\text{NFEV}}{\text{NFEV} - 2} + f_\varepsilon\,,$$

vagy, ha

3. az intervallum-felosztási algoritmus listája betelt.

Az $\overline{f}'(X)$ becslései másként hatnak az iterációs lépés hosszára, és ennek tükröződnie kell a megállási feltételekben. Az iterációkat akkor állítja le az algoritmus, ha

a. a $\min f'_j(X)$ (vagy a $-\max f'_j(X)$) természetes intervallum-kiterjesztéssel kapott aktuális becslése, F'_{act} nagyobb, mint $(f_\varepsilon - F_{\text{act}})/(0,01w(X_j))$,
b. az aktuális becslés nagyobb, mint nulla (azaz $f(x)$ monoton),
c. az aktuális becslés kisebb, mint $F'_{\text{old}}(\text{NFEV} - 2)/\,\text{NFEV}$, vagy ha
d. az intervallum-felosztási algoritmus listája betelt.

Itt $w(X_j)$ az illető intervallum szélességét jelöli, F_{act} az $\overline{f}$ végső becslését, NFEV pedig az aktuális transzformációs lépés meghatározásához használt célfüggvény- és deriválthívások számának összegét. Ezt a befoglaló függvények javítására

szolgáló eljárást a továbbiakban az optimális pontosságú befoglaló függvény módszerének nevezzük.

Tekintsük az algoritmusnak azt a fázisát, amikor az $\underline{X_j}$ lapot mozgatjuk befelé. Az előző, 2. szakaszban az F' befoglaló függvényt a teljes X aktuális intervallumra határoztuk meg. Mivel a lépésköz általában lényegesen kisebb, mint az X_j szélessége, ezért előnyösebb F'_j kiértékelését egy kisebb X' intervallumon elvégezni, amelyre $X'_i = X_i$, $i = 1, 2, \ldots, j-1, j+1, \ldots, n$; $\min X'_j = \min X_j$ és $w(X'_j) < w(X_j)$. Természetesen $\max X_j$ mozgatásakor mindez hasonlóan történik $\max X'_j$-vel.

Az X' intervallum $w(X'_j)$ szélességét a megfelelő irányban úgy módosítjuk adaptív módon, hogy az a számított lépésközt minél jobban megközelítse. Ez konkrétan a következőt jelenti: induláskor $w(X'_j)$ a $w(X_j)/10$ értéket kapja. Ha egy lépésköz kisebb volt, mint a számításához használt $w(X'_j)$, akkor $w(X'_j)$ új értéke a régi érték és a tényleges lépésköz átlaga lesz. Különben a ténylegesen megtett lépés hosszát a régi $w(X'_j)$-re korlátozzuk, és X'_j új szélességét a régi szélesség 1,5-szerese és X_j szélessége közül a kisebbre állítjuk be.

Az 1. szakaszban a konvergencia-tulajdonságok vizsgálata céljából mutattunk egy egyszerű megállási feltételt. A numerikus hatékonyság javítása érdekében ezt megváltoztattuk: az algoritmus most akkor áll meg, ha

I. az aktuális intervallum szélessége kisebb, mint egy előre adott δ konstans,

II. a célfüggvény- és deriválthívások együttes száma nagyobb, mint 100 000, vagy ha

III. mind az egy iteráción belül végrehajtott $2n$ transzformációs lépés hossza, kisebb, mint $\delta/100$.

Az utolsó módosítás, amit az alap-algoritmuson végrehajtunk, azt a numerikus tesztelés során tapasztalt jelenséget igyekszik kiküszöbölni, hogy az algoritmus hajlamos bizonyos koordináták mentén több nagyságrenddel több CPU-időt használni, mint a többi mentén. Ezért minden lépés során kiszámítunk egy E hatékonysági mutatót, amely egyenlő a lépésköz osztva a célfüggvény- és deriválthívások számának összegével. Minden iterációs lépésben meghatározzuk a 2n irányban E maximumát. A következő iterációban kihagyjuk azokat az irányokat, amelyekben a hatékonysági mutató kisebb volt, mint a maximum 10 százaléka. Minden 100 iteráció után viszont ismét az összes irányban végrehajtjuk a transzformációs lépést, tekintet nélkül a hatékonysági mutatóra. Ez biztosítja, hogy a befoglaló függvények pontosságában időközben bekövetkezett javulás felismerhető legyen, és így semelyik irány se marad ki túl hosszú ideig indokolatlanul.

Az 1. szakaszban kapott konvergencia-eredmények rögzített befoglaló függvényekre vonatkoztak. A bevezetett változtatásokkal ezek a függvények az algoritmus aktuális állapotától is függenek. Az 1. Lemmát és az 1. Tételt erre az esetre módosítva adódik a

4. TÉTEL. *Minden $F(X)$ és $F'(X)$ befoglaló függvényre, ha $f_\varepsilon < \min F(\underline{X}_j)$, és $\min F'_j(X) < 0$ valamely $j \in (1, 2, \ldots, n)$-re, akkor az iterációs lépés $t_j(X) - \min X_j$ hossza pozitív, és minden $x \in X$ pontra, amelyre $f(x) \leq f_\varepsilon$, teljesül, hogy $t_j(X) \leq x_j$.*

Hasonló módon, minden $F(X)$ és $F'(X)$ befoglaló függvényre, ha $f_\varepsilon < \min F(\overline{X}_j)$, és $0 < \max F'_j(X)$ valamely $j \in (1, 2, \ldots, n)$-re, akkor az iterációs lépés $\max X_j - t^j(X)$ hossza pozitív, és minden $x \in X$ pontra, amelyre $f(x) \leq f_\varepsilon$, teljesül, hogy $x_j \leq t^j(X)$.

Bizonyítás. Tegyük fel, hogy $f(x) < f_\varepsilon$ és $\min X_j \leq x_j < t_j(X)$. A $t_j(X)$ definíciójából

$$x_j - \min X_j < \frac{f_\varepsilon - \min F(\underline{X}_j)}{\min F'_j(X)} .$$

Az $f_\varepsilon < \min F(\underline{X}_j)$ és a $\min F'_j(X) < 0$ feltételek felhasználásával azt kapjuk, hogy

$$f(x) \geq \min F(\underline{X}_j) + \min F'_j(X)(x_j - \min X_j) > f_\varepsilon ,$$

ami ellentmondás. A lépésköz pozitivitása közvetlenül a transzformációk definíciójából következik. A $t^j(X) < x_j \leq \max X_j$ bizonyítása hasonló. □

A 4. Tétel másszóval azt állítja, hogy minden $F(X)$ és $F'(X)$ befoglaló függvényre az említett feltételek biztosítják, hogy az $X' = (X_1, X_2, \ldots, X_{j-1}, [t_j(X), t^j(X)], X_{j+1}, \ldots, X_n)$ intervallum X-nek az összes olyan pontját tartalmazza, amelyre a célfüggvény értéke kisebb, mint f_ε. Ebből következik, hogy ha X' üres, akkor X-nek nem lehet olyan pontja, amelyre a célfüggvény értéke legfeljebb f_ε. Ez azt jelenti, hogy a módosított algoritmus az I. – III. megállási feltételek nélkül olyan intervallumhoz konvergál, amelynek lapjain az $F(X)$ befoglaló függvény minimumának még az intervallum-felosztási módszerrel (az adott számítógépes korlátok mellet) elérhető legjobb becslése sem nagyobb, mint f_ε.

Tekintsük az 1–3 és az a–d megállási feltételeket a t_j transzformációra. (A t^j transzformációra hasonlóan kell a következő számításokat végrehajtani.) Legyen $\min F'(X)$ rögzített, és legyen F_{old} és F_{act} a $\min \overline{f}(\underline{X}_j)$ két egymást követő becslése. Az ezek meghatározásához szükséges függvényhívások száma NFEV−2, illetve NFEV. Ezen transzformációk hatékonysága az előzőek szerint

$$\frac{f_\varepsilon - F_{\text{old}}}{(\text{NFEV} - 2)\ \min F'(X)} \tag{5}$$

és

$$\frac{f_\varepsilon - F_{\text{act}}}{\text{NFEV}\ \min F'(X)} . \tag{6}$$

Az F_{act} becslés akkor eredményez hatékonyabb transzformációs lépést, ha (6) nagyobb, mint (5). Rövid számolás után kapjuk, hogy ez az eset akkor és csak akkor következik be, ha

$$F_{act} > \frac{(F_{old} - f_\epsilon)\text{NFEV}}{\text{NFEV} - 2} + f_\epsilon .$$

Rögzítsük most az F_{act} aktuális becslést, és tekintsük $\min \overline{f}'(X)$ két egymást követő becslését, legyenek ezek F'_{old} és F'_{act}. A megfelelő hatékonysági mutatók

$$\frac{f_\epsilon - F_{act}}{(\text{NFEV} - 2)\, F'_{old}} \tag{7}$$

és

$$\frac{f_\epsilon - F_{act}}{\text{NFEV}\, F'_{act}} . \tag{8}$$

Az F'_{act} becslés ismét akkor ad hatékonyabb transzformációs lépést, ha (8) nagyobb, mint (7). Ez pontosan akkor teljesül, ha

$$F'_{act} > \frac{F'_{old}(\text{NFEV} - 2)}{\text{NFEV}} .$$

Összegezve eredményeinket, azt állíthatjuk, hogy

5. TÉTEL. *A* 2 *és* c *megállási feltételek akkor állítják meg az intervallum-felosztási eljárást, amikor az aktuális becslés nem eredményez javítást az iterációs lépés hatékonyságán.*

A 2 és c megállási feltételek tehát egylépéses előrenézéses statisztika szerint optimalizálják a befoglaló függvényeket.

A numerikus tesztelést a standard globális optimalizálási tesztfeladatokkal hajtottuk végre. Ennek az az oka, hogy egyrészt a [4] 1. Állítása értelmében ezek minimumainak szerkezete megfeleltethető valamely paraméterbecslési feladatának, másrészt ezek alakja és globális minimumpontjai pontos elhelyezkedése is jól ismert (l. [4] 1. Táblázat). A kiindulási intervallumok megegyeztek az egyes feladatokban megszokottakkal [27]. Az f_ϵ értékeket minden esetben 0,001%-al választottuk nagyobbra, mint a megfelelő globális minimum. Ilyen problémafelvetés esetén könnyű ellenőrizni, hogy a korlátozó intervallumok elég szűkek-e. A megállási feltételek δ paramétere minden feladatra 0,01 volt. Az intervallum-felosztási eljárás egy legfeljebb 200 elemű listát használt a befoglaló függvények javítására. Az eljárás-paraméterek azonosak voltak az összes tesztfeladatra.

A programot ismét standard (tehát az intervallum-aritmetikát nem támogató) FORTRAN nyelven írtuk. Az intervallum-aritmetikát teljes egészében (az ún.

kifelé-kerekítéssel együtt) FORTRAN szubrutinokkal implementáltuk a hordozhatóság kedvéért. Az intervallum-felosztási algoritmusnak az alapváltozatát használtuk [6]. Említésre méltó, hogy a kifelé-kerekítés elhagyásával, illetve hozzávételével adódó teszteredmények az első 7 értékes számjegyben megegyeztek. A kapott hatékonysági mutatókat az 1. Táblázat foglalja össze.

1. Táblázat. A szinthalmaz korlátozási módszer hatékonysági mutatói

	Tesztfeladat								
	$S5$	$S7$	$S10$	$H3^{\dagger}$	$H6^{\dagger}$	$GP^{\dagger}$	$RB^{\dagger}$	$SHCB^{\dagger}$	$RCOS$
STU	3,2	5,5	12,3	118,8	585,5	621,6	49,8	49,8/80,7	39,8/1,1
NFE	170	188	251	13462	35903	82410	13511	18986/29159	8826/267
NDE	176	192	361	603	4099	17644	40327	1190/723	6766/263
NIT	27	25	34	82	322	552	3482	141/196	965/70

STU jelöli a felhasznált CPU-időt a standard egységben mérve, NFE és NDE a szükséges célfüggvény-, illetve deriválthívások számát, NIT pedig a végrehajtott iterációk számát. Egy † jelzi azokat a feladatokat, amelyek nem voltak teljesen megoldva abban az értelemben, hogy az algoritmus a függvényhívások nagy száma, vagy a túl kicsi lépésköz miatt állt le, míg az eredmény-intervallum még nem volt elég pontos. A táblázatban szereplő utolsó két probléma több globális minimumponttal rendelkezik a lehetséges megoldások halmazán. Ezekre a feladatokra két-két szám szerepel minden sorban: ezek az eredeti, illetve egy olyan kiindulási intervallumra vonatkoznak, amely csak egy globális minimumpontot tartalmaz.

A teszteredmények közvetlen összevetése globális optimalizálási módszerek hasonló mutatóival több szempontból is félrevezető lehet. Egyrészt a globális minimum értékének, mint input paraméternek a használata nem szokásos, másrészt a szinthalmaz-korlátozási algoritmus eredményhalmaza összehasonlíthatatlanul több információt nyújt a felhasználónak, mint a hagyományos eljárások. Algoritmusunk azt a nagyon erős állítást igazolja, hogy az $X^0 \setminus X^*$ tartományban nincs olyan pont, amelyre a célfüggvény értéke kisebb lenne f_ε-nál. Ennek ellenére, a teljesen megoldott feladatokra adódó hatékonysági mutatók a szakirodalomban közölt legjobbak közé tartoznak (v.ö. [4, 8, 27]).

A 3. szakaszban bevezetett módosítások lényegesen javítottak az algoritmus hatékonyságán és az eredmény-intervallum pontosságán is. Az optimális pontosságú befoglaló függvények módszere pedig más, intervallum-aritmetikán alapuló algoritmus hatékonyságát is javíthatja, anélkül, hogy az egyszerű, természetes intervallumkiterjesztésről le kellene mondani. Algoritmusunk garantált megbízhatóságú eredményhalmazát, a szinthalmazt korlátozó intervallumot sokféleképp lehet használni. Információt szolgáltat a célfüggvény érzékenységéről egy helyi minimumpont környezetében, és képes igazolni, hogy egy helyi minimum egyben globális minimum-e.

Az algoritmus nyilván azt is képes ellenőrizni, hogy egy paraméterbecslési feladat nem redundáns-e valamely változóban (azaz, van-e olyan x_i változó ($i \in$

$(1, 2, \ldots, n)$), hogy minden $x \in X$-hez és $x_i' \in X_i$-hez létezik olyan $x' \in X$, amire $f(x') = f(x)$), másszóval az illető modell azonosítható-e. Ha ugyanis a kezdeti intervallumot minden koordináta mentén sikerült összenyomni, akkor a feladat nem lehet redundáns. Ez utóbbi jelenség felismerése fontos a paraméterbecslési feladatokban, és az ismertetett algoritmus az egyetlen a szakirodalomban, amely képes ezt a tulajdonságot kimutatni.

FÜGGELÉK
Intervallum-aritmetika és a befoglaló függvények

Legyen I a kompakt valós intervallumok tere. Az intervallum-aritmetika műveletei ezen a halmazon vannak értelmezve. A műveleteket úgy kell definiálni, hogy az $A * B$ eredménye egy olyan C intervallum legyen, amely pontosan azon c valós számok halmaza, amelyekhez léteznek olyan $a \in A$ és $b \in B$ valósok, hogy $c = a * b$. Itt $*$ a négy alapművelet valamelyikét jelöli. Az ilyen aritmetika segítségével követni lehet a kerekítési hibákat, és az adatainkat terhelő bizonytalanság tükröződhet az eredményekben.

Az előző definíció mellett az intervallum-aritmetikát lehet kizárólag a valós aritmetikára támaszkodva is definiálni. Az $[a, b]$ és $[c, d]$ intervallumokra legyen

$$
\begin{aligned}
[a, b] + [c, d] &= [a + c, b + d], \\
[a, b] - [c, d] &= [a - d, b - c], \\
[a, b][c, d] &= [\min(ac, ad, bc, bd),\ \max(ac, ad, bc, bd)], \\
[a, b]/[c, d] &= [a, b][1/d, 1/c].
\end{aligned}
$$

Az osztást csak akkor értelmezzük, ha $0 \notin [c, d]$. Érdemes megjegyezni, hogy ez utóbbi feltétel jól megfogalmazott gyakorlati feladatokban tapasztalataink szerint szinte kivétel nélkül teljesül. A valós műveleteknek ezt a kiterjesztését intervallumokra természetes vagy naiv intervallum-kiterjesztésnek nevezik [25]. Az utóbbi években vizsgálják az olyan intervallum-aritmetikákat is, amelyek nem csak kompakt intervallumokon definiáltak. Ezeken a nullát tartalmazó intervallummal való osztás is értelmezhető.

Bár az alapműveletek pontosak a fenti értelemben, mégis, a velük kiszámított bonyolultabb függvények durva becslései is lehetnek a megfelelő értékkészletnek. A gyakran emlegetett példa [25] a következő: az $x - x^2$ értékkészlete a $[0, 2]$ intervallumon $[-2, 0, 25]$. Ezzel szemben az intervallum-kiterjesztéssel adódó intervallum $[-4, 2]$.

Az intervallum-aritmetika műveleteinek tulajdonságaival foglalkozik az intervallum-algebra. Számos, a valós műveletekre érvényes tulajdonság változatlanul teljesül az intervallum-műveletekre is (pl. a kommutativitás, asszociativitás az összeadásra és a szorzásra), de általában nincs inverz, és érvényes a szubdisztribúciós tulajdonság: $A(B + C) \subseteq AB + AC$.

Az alapműveletekhez hasonlóan könnyen lehet definiálni az elemi függvények intervallum-kiterjesztését is, tehát a számítógépen kiszámítható függvényeket szinte kivétel nélkül meg lehet valósítani természetes intervallum-kiterjesztésben is.

Az intervallum-aritmetika alkalmazása szempontjából alapvető fogalom a befoglaló függvény. Az $F(X) : I^n \to I$ az $f(x)$ n-változós valós függvény befoglaló függvénye, ha $f(x) \in F(X)$ érvényes minden $x \in X$ pontra és $X \in I^n$ intervallumra. Az intervallum-matematika fontos eredménye, hogy az $f(x)$ valós függvényből természetes (vagy naiv) intervallum-kiterjesztéssel adódó $F(X)$ függvény befoglaló függvény.

A befoglaló függvényektől természetes azt elvárni, hogy bővebb argumentum-intervallumra ne adjanak szűkebb eredmény-intervallumot. Ezt a feltételt fogalmazza meg az izotonitás: egy $F(X)$ befoglaló függvény akkor izoton, ha $X \subseteq Y$-ból következik $F(X) \subseteq F(Y)$. Az izotonitás szinte minden intervallum-aritmetika implementációra érvényes.

A befoglaló függvények minőségének fontos mutatója a rend: azt mondjuk, hogy az $F(X)$ befoglaló függvény rendje $\alpha > 0$, ha létezik olyan c valós konstans, hogy $w(F(X)) - w(\overline{f}(X)) \leq cw(X)^\alpha$ teljesül minden $X \in I^n$-re, ahol $w(X)$ az X intervallum szélessége. A természetes intervallum-kiterjesztéssel adódó befoglaló függvények elsőrendűek, de kidolgozott a magasabbrendű befoglaló függvények elmélete is [25]. Az egynél szélesebb intervallumokra a természetes intervallum-kiterjesztést, a kisebbekre pedig a magasabbrendű befoglaló függvényeket szokták ajánlani.

A számítógépes megvalósítás során minden intervallum-művelet végrehajtása után a kapott intervallumot módosítani szokás. Az intervallum alsó határát lefelé, felső határát felfelé kell kerekíteni a legközelebbi ábrázolható számra. Ezzel az úgynevezett kifelé kerekítési eljárással el lehet érni, hogy a befoglalási tulajdonság a kerekítési hibák ellenére is fennmaradjon. Ezen a módon számítógéppel automatizálható a garantált megbízhatóságú befoglaló függvények előállítása.

Az intervallum-aritmetikához használatos speciális kerekítéseket az IEEE szabvány biztosítja, ezért napjaink szinte minden processzora támogatja. A hetvenes évek közepétől elérhetők olyan programozási nyelvek [2,18], amelyek az INTERVAL adattípus használatát támogatják. Ilyen nyelveken még az intervallum-aritmetikát megvalósító szubrutinokat sem kell megírni: a megfelelő befoglaló függvény implementálásához elegendő a függvény kiszámításához használt változók típusát megváltoztatni.

A befoglaló függvényekre támaszkodó numerikus algoritmusok érzékenyek a befoglaló függvény minőségére, pontosságára. A vázolt természetes intervallum-kiterjesztés mellett számos más eljárás is ismert a befoglaló függvények előállítására, például a magasabbrendű deriváltakat is használó ún. középponti alakok, az automatikus deriválásra és monotonitás-vizsgálatra épülő stratégiák a befoglaló függvény javítására, illetve a 3. szakaszban ismertetett optimális pontosságú befoglaló függvényt generáló eljárás. Ezek a módosítások természetesen növelik az egy befoglaló függvény kiértékeléséhez szükséges számítások mennyiségét.

IRODALOM

[1] BAHVALOV, N.Sz., *A gépi matematika numerikus módszerei* (Műszaki Könyvkiadó, Budapest, 1977).

[2] BLEHER, J.H., S.M. RUMP, U. KULISCH, M. METZGER, CH. ULLRICH, W. WALTER, „FORTRAN-SC, A study of a FORTRAN extension for engineering/scientific computation with access to ACRITH", *Computing* **39** (1987), 93–110.

[3] CSENDES T., DARÓCZY B., HANTOS Z., „Nonlinear parameter estimation by global optimization: comparison of local search methods in respiratory system modelling", in: *System Modelling and Optimization*, Lecture Notes in Control and Information Sciences, No. 84, Eds. A. Prékopa and B. Straziczky, (Springer, Berlin, 1986), 188–192.

[4] CSENDES T., „Nonlinear parameter estimation by global optimization — efficiency and reliability", *Acta Cybernetica* **8** (1988), 361–370.

[5] CSENDES T., „An interval method for bounding level sets of parameter estimation problems", *Computing* **41** (1989), 75–86.

[6] CSENDES T., „Interval method for bounding level sets: revisited and tested with global optimization problems", *BIT* **30** (1990), 650–657.

[7] CSENDES T., RAPCSÁK T., „Nonlinear coordinate transformations for unconstrained optimization I. Basic transformations", *J. of Global Optimization* **3** (1993), 213–221.

[8] DIXON, L.C.W., G.P. SZEGŐ (EDS.), *Towards Global Optimisation 2* (North-Holland, Amsterdam, 1978).

[9] FORGÓ FERENC, *Nemkonvex és diszkrét programozás: korszerű ismeretek gazdasági szakemberek számára* (Közgazdasági és Jogi Könyvkiadó, Budapest, 1978).

[10] GERENCSÉR LÁSZLÓ, *Nemlineáris programozási feladatok megoldása szekvenciális módszerekkel*, SZTAKI Tanulmányok 49/1976 (Budapest, 1976).

[11] GILL, P. E., W. MURRAY, M.H. WRIGHT, *Practical Optimization* (Academic Press, London, 1981).

[12] HANSEN, E.R., „Global optimisation using interval analysis: the one-dimensional case", *J. Optim. Theory Appl.* **29** (1979), 331–344.

[13] HANSEN, E.R., „Global optimisation using interval analysis — the multi-dimensional case", *Numer. Math.* **34** (1980), 247–270.

[14] ICHIDA, K., Y. FUJII, „An interval arithmetic method for global optimization", *Computing* **23** (1979), 85–97.

[15] KÓSA ANDRÁS, *Optimumszámítási modellek* (Műszaki Könyvkiadó, Budapest, 1979).

[16] KRAWCZYK, R., „Properties of interval operators", *Computing* **37** (1986), 227–245.

[17] KREKÓ BÉLA, *Optimumszámítás. Nemlineáris programozás* (Közgazdasági és Jogi Könyvkiadó, Budapest, 1972).

[18] KULISCH, U. (ed.), *PASCAL-SC: A PASCAL extension for scientific computation* (Whiley & Sons, Chichester, 1987).

[19] LUTCHEN, K.R., JACKSON, A.C., „Statistical measures of parameter estimates from models fit to respiratory impedance data: emphasis on joint variabilities", *IEEE Trans. Biomed. Eng.* **33** (1986), 1000–1009.

[20] MARQUARDT, D.W., „An algorithm for least-squares estimation of nonlinear parameters", *SIAM J. Appl. Math.* **11** (1963), 431–441.

[21] MARTOS BÉLA, *Nemlineáris Optimalizálás* (Akadémiai Könyvkiadó, Budapest, 1974).

[22] MURTY, K.G., S.N. KABADI, „Some NP-complete problems in quadratic and nonlinear programming", *Math. Programming* **39** (1987), 117–130.

[23] PINTÉR, J., „Extended univariate algorithms for n-dimensional global optimization", *Computing* **36** (1986), 91–103.

[24] RAPCSÁK T., CSENDES T., „Nonlinear coordinate transformations for unconstrained optimization. II. Theoretical background", *J. of Global Optimization* **3** (1993), 359–375.

[25] RATSCHEK, H., J. ROKNE, *Computer Methods for the Range of Functions* (Ellis Horwood, Chichester, 1984).

[26] RATSCHEK, H., „Inclusion functions and global optimization", *Math. Programming* **33** (1985), 300–317.

[27] TÖRN, A., A. ŽILINSKAS, *Global Optimization*, Lecture Notes in Computer Science No. 350, G. Goos and J. Hartmanis, Eds. (Springer, Berlin, 1989).

(Beérkezett: 1991. június 10.)

CSENDES TIBOR
JATE KALMÁR LABORATÓRIUM
6720 SZEGED, ÁRPÁD TÉR 2.

AN INTERVAL METHOD FOR BOUNDING LEVEL SETS

T. CSENDES

An interval method for bounding level sets, modified to increase its efficiency and to get sharper bounding boxes, is presented. The new algorithm was tested with standard global optimization test problems. The test results show that, while the modified method gives a more valuable, guaranteed reliability result set, it is competitive with non-interval methods in terms of CPU time and number of function evaluations.

Alkalmazott Matematikai Lapok 17 (1993) 41–56

POLINOMOK GYÖKSTRUKTÚRÁJÁNAK VIZSGÁLATA A PARAMETRIKUS REPREZENTÁCIÓ MÓDSZERÉVEL

SIMON L. PÉTER ÉS FARKAS HENRIK

Budapest

A dinamikai rendszerek kvalitatív vizsgálata során gyakran felmerül az a probléma, hogy egy adott polinomnak hány valós gyöke van, illetve a gyökei között hány pozitív, negatív és nulla valós részű van. Dolgozatunkban megvizsgáljuk, hogy a polinom együtthatóit változtatva hogyan változhatnak ezek a tulajdonságok, különös tekintettel a polinom stabilitásvesztésére. Ezzel kapcsolatban igazolunk egy összefüggést a Routh-Hurwitz kritérium és a Hopf-féle bifurkáció között. Ezután mutatunk egy módszert, amelynek segítségével felderíthető egy polinom gyökeinek elhelyezkedése és az együtthatóktól való függése a gyökök numerikus meghatározása nélkül. Ezt a módszert részletesen ismertetjük a harmad- és negyedfokú polinomok esetében.

1. Bevezetés

Dinamikai rendszerek stacionárius pontjainak vizsgálatánál gyakran egy polinom gyökeinek elhelyezkedését kell vizsgálni. Ezzel kapcsolatban itt két fontos problémakört vizsgálunk:

1. A stacionárius pontok számát keressük. Ez a probléma egy $f(x) = 0$ egyenlet valós megoldásai számának megkeresésére redukálható.

2. Egy stacionárius pont jellegének megállapításához a Jacobi-mátrix karakterisztikus polinomjának gyökeit vizsgáljuk. Ekkor a negatív, illetve pozitív valós részű gyökök számát szeretnénk megtudni.

Konkrét dinamikai rendszerekből indulva, az 1. problémában szereplő f nem feltétlenül polinom, de igen gyakran az. Másrészről az is előfordulhat, hogy az 1. problémához az eredeti probléma redukciójával vagy transzformációjával jutunk, s így a kapott f polinom együtthatói az eredeti problémában szereplő paraméterektől függenek, de nem feltétlenül azonosak azokkal. Fontos alkalmazás a kémiai reakciókinetika, amikor is egy algebrai egyenletrendszer megoldásaiból adódnak a rendszer stacionárius állapotai. Az algebrai egyenletrendszer — igaz, fáradságos módon — de mindig redukálható egyetlen egyenletre [2].

Mindkét esetben a vizsgált polinom együtthatói az eredeti rendszer paramétereitől függenek, és az is vizsgálat tárgyát képezheti, hogy a paramétereket (és ezzel a polinom együtthatóit) változtatva, a gyökök elhelyezkedése hogyan változik. Ha a paraméterek változtatásakor a valós gyökök számában vagy a gyökök valós része előjelében változás következik be, *bifurkációról* beszélünk. Azokat a paraméterértékeket, amelyeknél a minőségi változás bekövetkezik, *bifurkációs értéknek* nevezzük. Tekintsünk két egyszerű példát a bifurkációra.

1. Példa. Vizsgáljuk az $x \mapsto x^2 + u$ polinom valós gyökeinek számát az $u \in \mathbb{R}$ paraméter különböző értékeinél. Az $u = 0$ paraméterérték bifurkációs érték a valós gyökök számának szempontjából, ugyanis $u < 0$ esetén két valós gyöke van a polinomnak, $u > 0$ esetén viszont nincs valós gyöke.

2. Példa. Vizsgáljuk az $x \mapsto x^2 + ux + 1$ polinom gyökei valós részének előjelét az $u \in \mathbb{R}$ paraméter különböző értékeinél. Az $u = 0$ paraméterérték bifurkációs érték a gyökök valós része előjelének szempontjából, mert $u > 0$ esetén két negatív valós részű, $u = 0$ esetén két nulla valós részű, $u < 0$ esetén pedig két pozitív valós részű gyöke van a polinomnak.

A bifurkáció fogalmának pontos meghatározásához tehát azt kell meghatározni, hogy mikor nevezünk két polinomot kvalitatíve egyformának (ekvivalensnek). Két polinomot akkor tekinthetünk például kvalitatíve egyformának, ha valós gyökeik száma megegyezik (1. példa). Egy másik feladat esetében két polinom akkor ekvivalens, ha pozitív, negatív, illetve nulla valós részű gyökeik száma rendre egyenlő (2. példa).

A dolgozat második szakaszában általánosan definiáljuk a bifurkáció fogalmát, majd ezt alkalmazzuk a polinomok esetére, és meghatározzuk az 1. és 2. példában szereplő tulajdonságokra vonatkozó bifurkációs értékek halmazát (azaz a bifurkációs halmazokat). Ezután egy polinom stabilitásvesztésének lehetséges eseteit vizsgálva kapcsolatot teremtünk a Hopf-féle bifurkáció és a Routh-Hurwitz-féle kritériumok sérülése között.

A dolgozat harmadik szakaszában két együtthatótól (a nullad- és az elsőfokú tag együtthatójától) függő polinomok halmazában vizsgáljuk a bifurkációs halmazokat a parametrikus reprezentáció módszerével.

A dolgozat negyedik és ötödik részében részletesen elvégezzük a fent ismertetett vizsgálatokat a harmad- és negyedfokú polinomok esetében, szemléletes módszert adva ezen polinomok gyökszerkezetének vizsgálatára.

2. Bifurkációs halmazok a polinomok halmazában

Ebben a szakaszban először általánosan ismertetjük a bifurkáció fogalmát. Legyen T egy topologikus tér, $\sim$ pedig egy ekvivalenciareláció a T téren.

1. Definíció. Egy $x \in T$ pontot *reguláris pontnak* nevezünk, ha létezik olyan N környezete, hogy minden $y \in N$ esetén $y \sim x$. Egy nem reguláris pontot *bifurkációs pontnak*, ezek halmazát *bifurkációs halmaznak* nevezzük.

A bifurkáció fogalmát legtöbbször a következőképpen vezetik be [1,5,7,9]. Legyen H egy halmaz, $\approx$ egy ekvivalenciareláció a H halmazon, $T \subset \mathbb{R}^k$ egy nyílt halmaz (paraméterhalmaz), $\lambda : T \to H$ pedig egy függvény (paraméterezés). Az $u_0 \in T$ paraméterértéket *reguláris paraméternek* nevezzük, ha létezik $U \subset T$ környezete, hogy minden $u \in U$ esetén $\lambda(u) \approx \lambda(u_0)$. A nem reguláris paramétereket

bifurkációs paramétereknek nevezzük. Ez a meghatározás megegyezik az 1. Definícióval abban az esetben, amikor a T topologikus tér $\mathbb{R}^k$ egy nyílt részhalmaza. Ugyanis a $\approx$ ekvivalenciareláció a λ függvényen keresztül definiál a T paraméterhalmazon egy $\sim$ ekvivalenciarelációt, és az 1. Definíció értelmében vett, $\sim$ ekvivalenciarelációra vonatkozó regularitás, megegyezik az iménti, $\approx$ ekvivalenciarelációra vonatkozó regularitás fogalommal.

A következőkben az n-ed fokú polinomok halmazában bifurkációkat fogunk vizsgálni. Legyen $a = (a_0, a_1, \ldots, a_{n-1}) \in \mathbb{R}^n$. Ez a vektor meghatároz egy

$$p_a(x) = a_0 + a_1 x + \cdots + a_{n-1}x^{n-1} + x^n$$

n-ed fokú polinomot. Ezért az 1 főegyütthatójú valós együtthatós polinomok halmaza az $\mathbb{R}^n$ térrel azonosítható. Legyen tehát a vizsgált topologikus tér $T := \mathbb{R}^n$. Kétféle ekvivalenciarelációt adunk meg a T topologikus téren a bevezetésben említett 1. és 2. feladatnak megfelelően:

1. Legyen $a \sim b$, ha p_a és p_b valós gyökeinek száma megegyezik.
2. Legyen $a \approx b$, ha p_a és p_b pozitív, negatív és nulla valós részű gyökeinek száma (multiplicitással) rendre egyenlő.

Vizsgáljuk meg, hogy a T térben milyen a bifurkációs pontok halmaza. Legyen

$$D_R = \{a \in \mathbb{R}^n : \exists x \in \mathbb{R},\ p_a(x) = p_a'(x) = 0\}$$
$$N = \{a \in \mathbb{R}^n : \exists x \in \mathbb{R},\ p_a(ix) = 0\}.$$

A D_R halmazhoz tartozó paraméterértékeknél a polinom diszkriminánsa [6] zérus, mert ott a polinomnak legalább kétszeres valós gyöke van. Másrészről a diszkrimináns zérus voltából többszörös gyök léte következik, amely azonban nem feltétlenül valós. Ezért már itt bevezetjük megkülönböztetésként a D diszkrimináns halmazt is:

$$D = \{a \in \mathbb{R}^n; \exists x \in \mathbb{C},\ p_a(x) = p_a'(x) = 0\}.$$

Az N halmaz a nulla valós részű gyökkel rendelkező polinomok halmaza.

Bebizonyítjuk, hogy D_R, illetve N az $\sim$, illetve az $\approx$ ekvivalenciarelációra vonatkozó bifurkációs halmazok.

1. TÉTEL. *Az $\sim$ relációra vonatkozó B_1 bifurkációs halmaz azonos a D_R halmazzal: $B_1 = D_R$.*

2. TÉTEL. *Az $\approx$ relációra vonatkozó B_2 bifurkációs halmazra $B_2 = N$.*

A tételek bizonyítása a függelékben található.

Az 1. Tétel más megfogalmazásban azt állítja, hogy a valós gyökök száma a paraméterek értékének folytonos változtatásakor nem változhat, ha a D_R halmazon kívül maradunk. Ekkor ugyanis mindegyik gyök a paraméterek folytonos függvényeként fejezhető ki az implicit-függvény tétel szerint [2].

Az $\approx$ reláció ekvivalenciaosztályai közül kiemelünk egyet, amelyben a stabilis polinomokhoz tartozó együttható-vektorok vannak:

$$S := \{a \in \mathbb{R}^n; \forall x \in \mathbb{C},\ (p_a(x) = 0 \implies \operatorname{Re} x < 0)\},$$

ezt stabilitási tartománynak nevezzük. A jól ismert Routh-Hurwitz-feltételek meghatározzák a stabilitási tartományt, így kézenfekvő, hogy az N halmaz és a Routh-Hurwitz-feltételek között kapcsolatot keressünk. A p_a polinom Routh-Hurwitz-mátrixa:

$$\begin{pmatrix} a_{n-1} & a_{n-3} & \cdot & \cdot & \cdot & \cdot & 0 \\ 1 & a_{n-2} & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & a_2 & a_0 & 0 \\ \cdot & \cdot & \cdot & \cdot & a_3 & a_1 & 0 \\ \cdot & \cdot & \cdot & \cdot & a_4 & a_2 & a_0 \end{pmatrix}$$

Jelölje $\Delta_i(a)$ ennek i-edig főminorját ($i = 1, 2, \ldots, n$). A Routh-Hurwitz-kritérium szerint [6]:

$$S = \{a \in \mathbb{R}^n : \Delta_i(a) > 0 \quad i = 1, 2, \ldots, n\}.$$

Legyen $a \in \mathbb{R}^n$, jelölje a p_a polinom gyökeit x_{ai} ($i = 1, 2, \ldots, n$) valamilyen sorrendben. (A többszörös gyököket többször soroljuk fel.) Legyen

$$H^{**} = \{a \in \mathbb{R}^n : \exists i \neq j,\ x_{ai} + x_{aj} = 0\}$$

azon együttható-vektorok halmaza, amelyek által meghatározott polinom valamely két gyökének összege nulla. Ez tartalmazza a tiszta képzetes gyökkel rendelkező polinomok együttható-vektorait is. (Megyjegyezzük, hogy ha a p_a polinomnak 0 többszörös gyöke, akkor $a \in H^{**}$.)

A H^{**} halmaz jelentőségét világítja meg az alábbi tétel, amely egyúttal az $(n-1)$-edik Routh-Hurwitz-feltételnek ad szemléletes jelentést. Bizonyítása a függelékben található.

3. TÉTEL. $H^{**} = \{a \in \mathbb{R}^n : \Delta_{n-1}(a) = 0\}$.

3. Kétparaméteres eset

A továbbiakban a polinom a_0 és a_1 együtthatóját tekintjük kontrollparaméternek, a többi együttható értékét rögzítettnek tekintjük. A bifurkációs halmazokat az $\mathbb{R}^2$ síkon a parametrikus reprezentáció módszerével vizsgáljuk. Az e módszer irodalmában [4,2] szokásosabb jelölésekre áttérve legyen

$$g(x) := u_2x^2 + \cdots + u_{n-1}x^{n-1} + x^n \quad u_i \in \mathbb{R} \quad \text{rögzített, ha} \quad 2 \leq i \leq n-1,$$

és

$$p_u(x) = u_0 + u_1x + g(x),$$

ahol most $u = (u_0, u_1) \in \mathbb{R}^2$ a paramétervektor.

A korábban bevezetett diszkrimináns halmaz, illetve D_R halmaz most értelemszerűen:

$$D = \{u \in \mathbb{R}^2 : \exists x \in \mathbb{C},\ p_u(x) = p'_u(x) = 0\},$$
$$D_R = \{u \in \mathbb{R}^2 : \exists x \in \mathbb{R},\ p_u(x) = p'_u(x) = 0\}.$$

A D halmazt meghatározó egyenletekből az x változó eliminálásával az u_0 és u_1 közötti, rendszerint igen bonyolult algebrai egyenletet kapunk. A parametrikus reprezentáció módszerével a D_R halmaz könnyen vizsgálható, most azt is megmutatjuk, hogy a D_R halmaz esetünkben egy görbe. Tekintsük ugyanis a

$$p_u(x) = p'_u(x) = 0$$

egyenletrendszerben az x változót paraméternek, és az u_0, u_1 koordinátákat az x „paraméter" segítségével fejezzük ki:

$$\begin{aligned} u_0(x) &= xg'(x) - g(x) \\ u_1(x) &= -g'(x) \end{aligned} \qquad x \in \mathbb{R}. \tag{1}$$

Tehát valóban egy görbét kaptunk, amelyet a továbbiakban D_R-görbének nevezünk. Az 1. Tétel szerint a D_R-görbe a valós gyökök száma szempontjából jelentős, sőt a valós gyökök értékére vonatkozóan is ad információt [2]. Ugyanis bebizonyítható, hogy [2]:

1. A D_R-görbe olyan tartományokra bontja a paramétersíkot, amelyekben a valós gyökök száma állandó.

2. A D_R-görbe x paraméterű pontjához húzott érintő pontosan azon u értékpárokat tartalmazza, amelyekhez tartozó p_u polinomnak az x gyöke.

Emlékeztetünk arra, hogy a H^{**} halmaz azon paraméterek halmaza, amelyekhez tartozó polinom két gyökének összege nulla. Esetünkben tehát:

$$H^{**} = \{u \in \mathbb{R}^2 : \exists i \neq j,\ x_{ui} + x_{uj} = 0\},$$

ahol x_{uj} $i = 1, 2, \ldots n$ a p_u polinom gyökeit jelöli. A Hopf-féle bifurkációnál a H^{**} egy H részhalmazának van jelentősége:

$$H = \{u \in H^{**} : \exists x \in \mathbb{R} \setminus \{0\},\ p_u(ix) = 0\}.$$

Ugyanis, ha a p_u polinom egy dinamikai rendszer valamely stacionárius pontjában a Jacobi-mátrix karakterisztikus polinomja, akkor $u \in H$ esetén ebben a pontban Hopf-bifurkáció lehet [5,9]. A H halmaz megadható a parametrikus reprezentáció módszerével az I^2 negatív valós paramétert használva, ahol I és $-I$ a polinom két tiszta képzetes gyökét jelöli:

$$\begin{aligned} u_0(I) &= -[g(I) + g(-I)]/2 \\ u_1(I) &= -[g(I) - g(-I)]/2I \end{aligned} \tag{2}$$

Könnyen látható, hogy mindkét kifejezésben az I paraméter csak páros hatványon szerepel, ezért I^2 valóban tekinthető paraméternek. Mivel a H halmaz az I^2 negatív valós paraméterrel paraméterezett görbe, azért a továbbiakban H-görbének nevezzük. A (2) paraméterezés az I^2 paraméter pozitív értékeire is értelmezhető, és kiterjeszthető az $I = 0$ értékre is. Ez a görbe az $I^2 > 0$ paraméterek esetén azon u együttható-vektorokat határozza meg, amelyekhez tartozó p_u polinomnak az I és $-I$ valós szám gyöke. Az $I = 0$ paraméterhez pedig az a polinom tartozik, amelynek a nulla legalább kétszeres gyöke, ugyanis $u_1(0) = u_0(0) = 0$. Megjegyezzük, hogy a paramétersík origója egyúttal a D_R-görbén is rajta van.

Jelölje H^* azt a görbét, amelyet a (2) paraméterezés meghatároz, amint az I^2 paraméter befutja a valós számok halmazát. Tehát a H-görbe a H^*-görbe azon része, amelyben az I^2 paraméter értéke negatív valós szám.

A H-görbe vizsgálata azért is jelentős, mert felrajzolásával megadható a paramétersíkon a stabilitási tartomány (S). Ugyanis a H-görbe segítségével megadható az N síkbeli megfelelője:

$$N = H \cup \{u \in \mathbb{R}^2 : u_0 = 0\}.$$

Ebben a halmazban találhatók azok a paraméterek, amelyekhez tartozó polinomnak van nulla valós részű gyöke, és a 2. Tétel szerint az N halmaz olyan tartományokra bontja a paramétersíkot, amelyekben a megfelelő polinom pozitív és negatív valós részű gyökeinek száma állandó. Tehát az N halmaz határozza meg a stabilitási tartományt is.

A továbbiakban a fenti elméletet alkalmazzuk a harmadfokú és negyedfokú polinomok vizsgálatára.

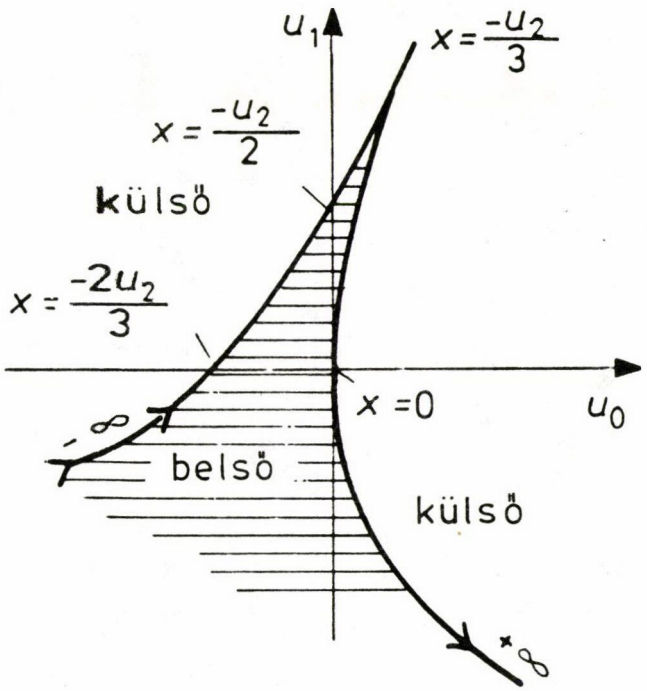
4. Harmadfokú polinomok

Legyen $p_u(x) = u_0 + u_1x + u_2x^2 + x^3$. Az (1) kifejezésből a D_R-görbe paraméteres egyenlete:

$$u_0(x) = x^2(2x + u_2)$$
$$u_1(x) = -x(3x + 2u_2)$$

Az 1. ábrán a D_R-görbét láthatjuk $u_2 > 0$ esetén. A görbe az $x = 0$ paraméternél érinti az u_1 tengelyt, az $x = -u_2/3$ értéknél pedig csúcsa van. A rajzon nyíl jelzi az x paraméter növekedésének irányát a görbe mentén. A D_R-görbe egy „belső" és egy „külső" részre osztja a paramétersíkot.

1. Ha u a „belső" részben van (2.a ábra), akkor abból három érintő húzható a D_R-görbéhez, így a p_u polinomnak három valós gyöke van, ezek értékét (x_1, x_2, x_3) az x paraméter érintési pontokban felvett értéke adja meg. Érdekes megjegyezni, hogy az x_i paraméterű pontba húzott érintő egy másik pontban (y_i) is metszi a D_R-görbét, és y_i értéke éppen a másik két gyök számtani közepével egyenlő. Ez abból



1. ábra

következik, hogy a három gyök összege a paramétersík bármely u pontjához tartozó p_u polinomnál ugyanaz $(-u_2)$ és az y_i paraméterű ponthoz tartozó polinomnak y_i kétszeres gyöke.

2. Ha u a „külső" részben van (2.b ábra), akkor az u pontból csak egy érintő húzható a görbéhez, az érintési pont paramétere (x_1) adja a p_u polinom egyetlen valós gyökének értékét. Az érintő másik metszéspontja a görbével az $x = R$ paraméternél van, az adja a másik két gyök valós részét, mivel a három gyök összege állandó $(-u_2)$. A gyökök képzetes részének (I) is szemléletes jelentése van a 2.b ábrán:

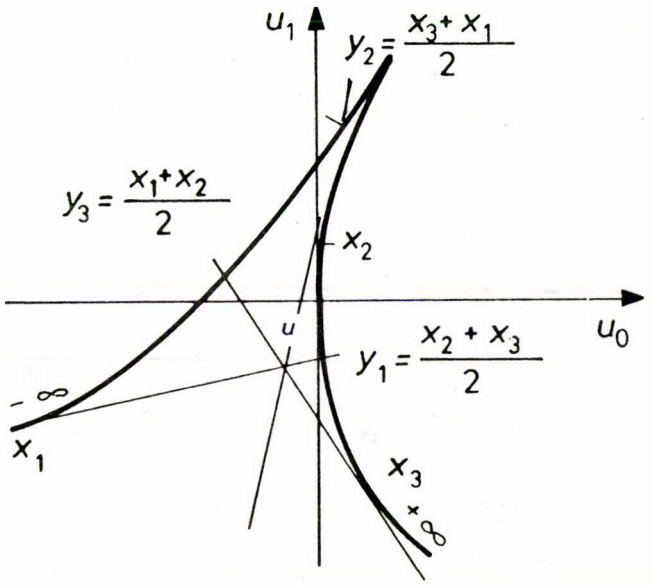
$$I = \pm\sqrt{u_1 - u_{10}}$$

itt u_1 az u pont, u_{10} pedig a metszéspont második koordinátája [3].

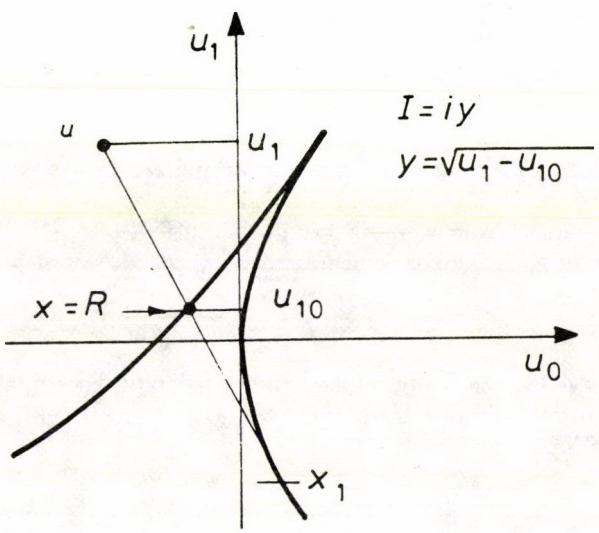
Ezután a stabilitási tartományt vizsgáljuk. A H^*-görbe egyenlete a (2) paraméteres felírás alapján:

$$u_0 = u_1 u_2$$

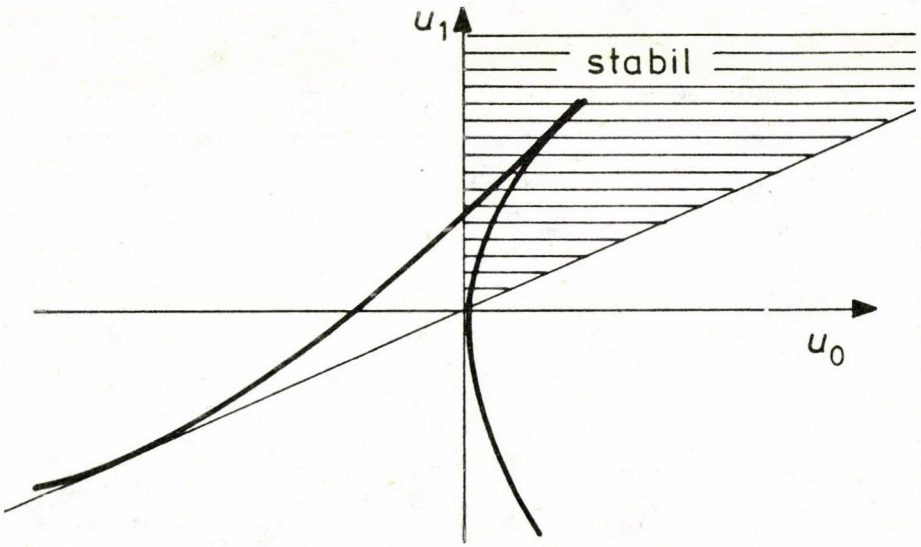
Tehát a H^*-görbe egy egyenes, ennek az $I^2 < 0$ paraméterekhez tartozó része, azaz a H-görbe egy félegyenes, amely az $\{(u_0, u_1) \in \mathbb{R}^2 : u_0 > 0, u_1 > 0\}$ pozitív síknegyedben van. Az S stabilitási tartományt a H-görbe és az u_1 tengely pozitív része határolja, amint az a 3. ábrán látható.



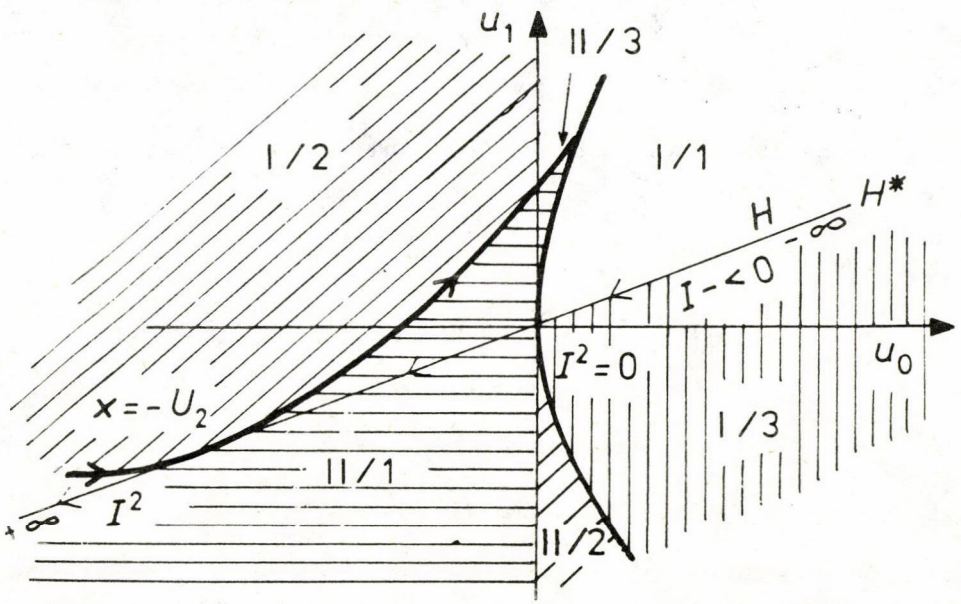
2.a ábra



2.b ábra


u_1
stabil
u_0

3. ábra


u_1
II/3
I/2
I/1
H
H^*
$-\infty$
I-<0
$I^2=0$
u_0
$x=-U_2$
II/1
I/3
I^2
$+\infty$
II/2

4. ábra

Összefoglalva: a paramétersík a gyökök valós részének előjele szerint a 4. ábrán látható hat részre osztható. Ezekben a gyökök (x_1, x_2, x_3) valós részének előjele az alábbi módon adható meg:

A „külső" részben	A „belső" részben
I/1. $x_1 < 0,\ \mathrm{Re}\, x_2 < 0,\ \mathrm{Re}\, x_3 < 0$	II/1. $x_1 < 0,\ x_2 < 0,\ x_3 > 0$
I/2. $x_1 > 0,\ \mathrm{Re}\, x_2 < 0,\ \mathrm{Re}\, x_3 < 0$	II/2. $x_1 < 0,\ x_2 > 0,\ x_3 > 0$
I/3. $x_1 < 0,\ \mathrm{Re}\, x_2 > 0,\ \mathrm{Re}\, x_3 > 0$	II/3. $x_1 < 0,\ x_2 < 0,\ x_3 < 0.$

5. Negyedfokú polinomok

Legyen

$$p_u(x) = u_0 + u_1 x + u_2 x^2 + x^3 + x^4. \tag{3}$$

Egy általános negyedfokú polinom a fenti alakra hozható a változó lineáris transzformációjával. Az (1) kifejezésből a D_R-görbe paraméteres egyenlete:

$$u_0(x) = x^2(u_2 + 2x + 3x^2)$$
$$u_1(x) = -x(4x^2 + 3x + 2u_2)$$

A D_R-görbe felrajzolásához a csúcspontok és az önmetszéspont helyét kell meghatározni.

A csúcsponthoz tartozó c_1, c_2 paraméterértékek a $p_u''(c) = 0$, azaz az

$$u_2 + 3c + 6c^2 = 0$$

egyenletből

$$c_{1,2} = \frac{-3 \pm \sqrt{3}d}{12}, \text{ ahol } d = \sqrt{3 - 8u_2}\ .$$

Látható, hogy az u_2 paraméter értékétől függően a D-görbének nulla, vagy két csúcspontja van.

Nézzük ezután az önmetszéspontokat. Ha a polinom

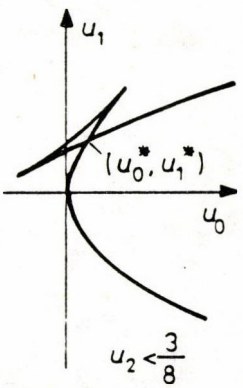
$$p_u(x) = (x - x_1)^2(x - x_2)^2 \tag{4}$$

alakú, és $x_1 \neq x_2$ valós számok, akkor a D_R-görbének az u pontban önmetszéspontja van. Ha $x_1 = x_2$ valós számok, akkor a p_u polinomnak négyszeres gyöke van. Ha pedig $x_1 = \bar{x}_2$ nem valós számok, akkor a p_u polinomnak egy kétszeres komplex gyöke van, ekkor az u pontot a D_R-görbe komplex kiegészítő pontjának nevezzük. Ekkor tehát az u paraméter a D halmazhoz tartozik, de nincsen rajta a D_R-görbén.

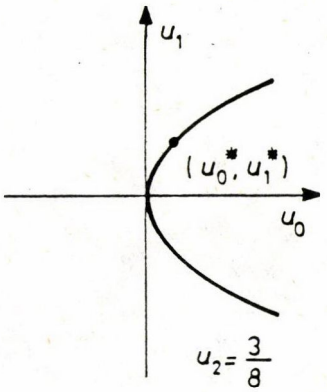
Ebben a három esetben az u pontot TDR-pontnak (Two Double Roots) nevezzük. A TDR-pont paramétereit a (3) és (4) összefüggésből kapjuk: $x_{1,2} = \frac{-1 \pm d}{4}$. Ebből a TDR-pont koordinátái az (u_0, u_1) síkon:

$$u_0^* = \left(\frac{4u_2 - 1}{8}\right)^2, \quad u_1^* = \frac{4u_2 - 1}{8}.$$

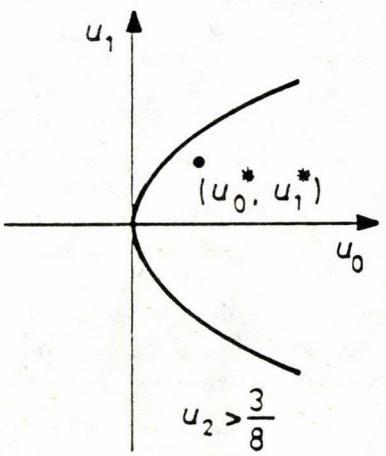
Tehát, ha $3 > 8u_2$, akkor (u_0^*, u_1^*) egy önmetszéspont;
ha $3 = 8u_2$, akkor (u_0^*, u_1^*) egy négyszeres gyökhöz tartozik;
ha $3 < 8u_2$, akkor (u_0^*, u_1^*) a komplex kiegészítő pont.
Ebben a három esetben a D_R-görbe az 5. ábrán látható.



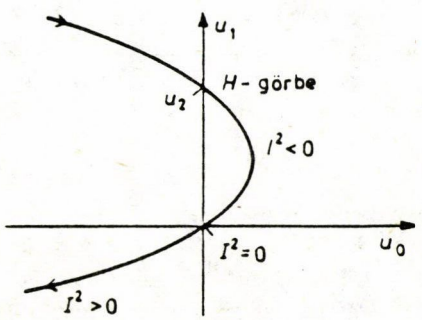
5.a ábra



5.b ábra



5.c ábra



6. ábra

A [3] munkában megvizsgáltuk, hogy az u_2 paramétert változtatva hogyan mozog az (u_0, u_1) síkon a TDR-pont és a csúcspontok.

Tekintsük végül a H^*-görbét. (2) alapján a H^*-görbe egyenlete:

$$u_0 = u_1(u_2 - u_1).$$

A 6. ábrán látható a H^*-görbe, ennek a felső félsíkban levő része a H-görbe, amely a stabilitási tartományt határolja.

FÜGGELÉK

Az alábbiakban bebizonyítjuk az 1., 2. és 3. Tételt. Az első két tétel bizonyításához szükség van egy segédtételre, amelyhez vezessük be a következő jelölést. Legyen $x \in \mathbb{R}^n$, $r > 0$. Jelölje

$$S(x, r) := \{y \in \mathbb{R}^n : |x - y| < r\}$$

az x pont körüli r sugarú nyílt gömböt. Ezt a jelölést a komplex síkon egy nyílt kör jelölésére is használni fogjuk.

1. LEMMA. *Legyen $x \in \mathbb{C}$ a p_a polinom k multiplicitású gyöke, ahol $a \in \mathbb{R}^n$ egy tetszőleges vektor. Ekkor minden $\delta > 0$ számhoz található olyan $\varepsilon > 0$ szám, hogy bármely $b \in S(a, \varepsilon)$ vektor esetén a p_b polinomnak az $S(x, \delta)$ körben multiplicitással számolva k gyöke van.*

Bizonyítás. A lemma más szavakkal azt fejezi ki, hogy egy polinom gyökei (többszörös gyökei is) az együtthatóktól folytonosan függenek. Ezért megmutatjuk, hogy az alább definiálandó $G_i : \mathbb{R}^n \to \mathbb{C}$ $(i = 1, 2, \ldots, n)$ függvények, amelyek egy adott együttható-vektorhoz az általa meghatározott polinom i-edik gyökét rendelik hozzá, folytonosak. Ahhoz, hogy ezeket a függvényeket egyértelműen megadhassuk, először meg kell határozni, hogy mit értünk a polinom i-edik gyökén. Egy adott polinom gyökeit a továbbiakban a következőképpen fogjuk sorszámozni: legyen az első gyök az, amelynek legkisebb a valós része, ha több ilyen van akkor azok közül az, amelyeknek legkisebb a képzetes része. Ha ez a gyök többszörös, akkor annyi sorszámot kap egymás után, amennyi a multiplicitása. A következő sorszámot az a gyök kapja, amelyik az első lenne, ha az előtte levőt elhagynánk. A polinom többi gyökét ugyanezen az elven sorszámozzuk.

Jelölje $b \in \mathbb{R}^n$ esetén $G_i(b) \in \mathbb{C}$ a p_b polinom i-edik gyökét $(i = 1, 2, \ldots, n)$. Jelölje $G : \mathbb{R}^n \to \mathbb{C}^n$ azt a függvényt, amelynek koordináta függvényei a G_i függvények. A G függvény folytonossága az inverzének folytonosságából következik. Legyen ugyanis $A_j : \mathbb{C}^n \to \mathbb{C}$ az a leképezés, amelyre $A_j(x_1, x_2, \ldots, x_n)$ a $\prod_{k=1}^{n}(x - x_k)$ polinom j-ed fokú tagjának együtthatója. Az A_j függvény maga is polinom (n változós), ezért folytonos. Jelölje $A : \mathbb{C}^n \to \mathbb{C}^n$ azt a függvényt, amelynek koordináta

függvényei az A_j függvények. Az A függvény inverze G (a G függvényt a komplex együtthatós polinomokra is értelmezve). Mivel egy kompakt halmazon értelmezett folytonos bijekció inverze is folytonos, azért csak egy alkalmas K kompakt halmazt kell megadni, hogy az A függvényt a K halmazra leszűkítve a G függvény $A(K)$ halmazra vett leszűkítésének folytonosságára következtethessünk. A K halmazt úgy kell megadni, hogy az $A(K)$ halmaz tartalmazza az állításban rögzített $a \in \mathbb{R}^n$ együttható-vektor valamely környezetét. Könnyen látható, hogy létezik olyan $R > 0$ szám, hogy minden $b \in S(a,1)$ esetén a p_b polinom gyökeinek abszolútértéke R-nél kisebb. Legyen $K := \overline{S(0,R)} \subset \mathbb{C}^n$ az R sugarú origó közepű gömb lezárása a $\mathbb{C}^n$ térben. Ekkor $A(K) \supset S(a,1)$, tehát a G függvény $S(a,1)$ gömbre vett megszorítása folytonos, amiből a Lemma állítása azonnal következik.

Az *1. Tétel bizonyítása:* (1) Először igazoljuk, hogy $B_1 \subset D_R$. Tegyük fel, hogy $a \in \mathbb{R}^n$ és $a \neq D_R$. Bebizonyítjuk, hogy ekkor $a \notin B_1$, azaz létezik olyan ε pozitív szám, hogy minden $b \in S(a,\varepsilon)$ esetén $a \sim b$. Válasszuk a $\delta > 0$ számot olyan kicsire, hogy a p_a polinom gyökei köré felvett δ sugarú nyílt körök diszjunktak legyenek, továbbá a nemvalós gyököket körülölelő körök a valós tengelytől is diszjunktak lesznek. Az 1. Lemma szerint ehhez a δ számhoz létezik egy olyan ε pozitív szám, hogy tetszőleges $b \in S(a,\varepsilon)$ esetén mindegyik δ sugarú körben a p_a és a p_b polinomnak ugyanannyi gyöke van multiplicitással számolva. Mivel $a \notin D_R$, azaz p_a valós gyökei egyszeresek és a felvett körök diszjunktak, a p_a polinom valós gyökei körüli körökben a p_b polinomnak is egy gyöke van, így az is valós, hiszen a nem valós gyökök párosával szerepelnek. Tehát p_a és p_b valós gyökeinek száma megegyezik, így $a \sim b$.

(2) Most bebizonyítjuk, hogy $D_R \subset B_1$. Legyen $a \in D_R$, igazolni kell, hogy minden ε_1 pozitív számhoz van olyan $b \in S(a,\varepsilon_1)$ vektor, amely nem ekvivalens az a vektorral. Jelölje y a p_a polinom egyik $k \geq 2$ multiplicitású valós gyökét (ilyen van, mert $a \in D_R$). Ekkor a p_a polinom

$$p_a(x) = (x-y)^k q(x)$$

alakba írható, ahol a q olyan polinom, amelynek y nem gyöke. Válasszunk az a vektorhoz δ és ε pozitív számokat most is ugyanúgy, mint a bizonyítás első részében, azzal a kiegészítéssel, hogy a q polinomnak az $S(y,\delta)$ körben ne legyen gyöke. Legyen c egy valós szám és $b \in \mathbb{R}^n$ az a vektor, melyre

$$p_b(x) = (x-y)^k q(x) + c(x-y)^{k-2} q(x) = (x-y)^{k-2} q(x)((x-y)^2 + c).$$

Legyen $\varepsilon_1 < \varepsilon$ tetszőleges, és c olyan negatív szám, hogy $b \in S(a,\varepsilon_1)$ (ilyen mindig található). Az $S(y,\delta)$ körben a p_a polinomnak csak egy valós gyöke van, y. A p_b polinomnak viszont ebben a körben legalább két valós gyöke van, $y \pm \sqrt{-c}$, ha c elég kis abszolútértékű. A p_b polinom többi gyöke viszont megegyezik a p_a polinom gyökeivel. Tehát a p_b polinomnak több valós gyöke van, mint a p_a polinomnak, azaz a b vektor nem ekvivalens az a vektorral. Ezzel a tételt igazoltuk.

A *2. Tétel bizonyítása:* (1) Először igazoljuk, hogy $B_2 \subset N$. Tegyük fel, hogy $a \in \mathbb{R}^n$ és $a \notin N$, azaz a p_a polinomnak nincs nulla valós részű gyöke. Bebizonyítjuk, hogy ekkor $a \notin B_2$, azaz létezik olyan ε pozitív szám, hogy minden $b \in S(a, \varepsilon)$ esetén $a \approx b$. Válasszunk az a vektorhoz δ és ε pozitív számokat most is ugyanúgy, mint az előző bizonyítás első részében. Legyen $b \in S(a, \varepsilon)$ tetszőleges, ekkor az 1. Lemma szerint a p_a és a p_b polinomnak ugyanannyi negatív valós részű, és ugyanannyi pozitív valós részű gyöke van, tehát $a \approx b$.

(2) Most bebizonyítjuk, hogy $N \subset B_2$. Legyen $a \in N$, igazolni kell, hogy minden ε_1 pozitív számhoz van olyan $b \in S(a, \varepsilon_1)$ vektor, amely nem ekvivalens az a vektorral. Két esetet különböztetünk meg: 1. A p_a polinomnak a nulla gyöke, 2. a p_a polinomnak tiszta képzetes gyökei vannak.

1. Legyen a p_b polinomnak a nulla k-szoros gyöke. Ekkor

$$p_a(x) = x^k q(x),$$

ahol $q(0) \neq 0$. Legyen $b \in \mathbb{R}^n$ az a vektor, melyre

$$p_b(x) = x^k q(x) + cx^{k-1} q(x).$$

Könnyen látható, hogy bármely ε_1 pozitív számhoz megadható úgy a $c \in \mathbb{R}$ szám, hogy $b \in S(a, \varepsilon_1)$ fennálljon. Ezután az előző tétel bizonyításának gondolatmenetét követve kapjuk, hogy a b vektor nem ekvivalens az a vektorral, ugyanis a p_b polinomnak eggyel kevesebb nulla valós részű gyöke van, mint a p_a polinomnak.

2. Legyen a p_a polinomnak az $i\omega$ szám ($\omega \in \mathbb{R}$) k-szoros gyöke. Ekkor

$$p_a(x) = (x^2 + \omega^2)^k q(x),$$

ahol $q(i\omega) \neq 0$. Legyen $b \in \mathbb{R}^n$ az a vektor, melyre

$$p_b(x) = (x^2 + \omega^2)^k q(x) + cx(x^2 + \omega^2)^{k-1} q(x).$$

Ezután a bizonyítás az előző esethez hasonló, ugyanis az $x^2 + \omega^2 + cx$ polinomnak $c \neq 0$ esetén nincs nulla valós részű gyöke, tehát a p_b polinomnak kettővel kevesebb nulla valós részű gyöke van, mint a p_a polinomnak. Ezzel a tételt igazoltuk.

A 3. Tétel bizonyításához szükségünk lesz a rezultáns fogalmára [6,8]. Legyenek p és q polinomok:

$$p(x) := \sum_{i=0}^{n} a_i x^i \qquad q(x) := \sum_{j=0}^{m} b_j x^j$$

E két polinom rezultánsa az alábbi $(n+m)\times(n+m)$-es determináns:

$$R(p,q)=\begin{pmatrix} a_n & \ldots\ldots & a_0 & 0 & \ldots\ldots & 0 \\ 0 & a_n & \ldots\ldots & a_0 & 0 \ \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & a_n & \ldots\ldots & \ldots\ldots & a_0 \\ b_m & b_0 & 0 & \ldots\ldots & \ldots\ldots & 0 \\ 0 & b_m & b_0 & 0 & \ldots\ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & b_m & \ldots\ldots & \ldots\ldots & b_0 \end{pmatrix}$$

A rezultáns jelentőségét mutatja az alábbi lemma.

2. LEMMA. *Két polinomnak pontosan akkor van közös gyöke, ha rezultánsuk nulla.* [6,8]

A *3. Tétel bizonyítása* Először megmutatjuk, hogy $a \in H^{**}$ pontosan akkor teljesül, ha az alábbi két egyenletnek van közös megoldása:

$$a_0 + a_2x^2 + a_4x^4 + \cdots = 0 \tag{5}$$

$$a_1 + a_3x^2 + a_5x^4 + \cdots = 0. \tag{6}$$

Legyen $a \in H^{**}$, ekkor két eset lehetséges:

A) A 0 szám többszörös gyöke a p_a polinomnak. Ekkor $a_0 = a_1 = 0$, ezért $x = 0$ az (5) és (6) közös megoldása.

B) Van olyan $x \neq 0$, melyre $p_a(x) = p_a(-x) = 0$. A $p_a(x) = 0$ és $p_a(-x) = 0$ egyenleteket összeadva, illetve kivonva kapjuk az (5) és (6) egyenleteket, melyeknek tehát x közös megoldása.

Amennyiben az (5) és (6) egyenleteknek van közös megoldása, akkor az iménti úton visszafelé haladva kapjuk, hogy $a \in H^{**}$.

Most megmutatjuk, hogy az (5) és (6) egyenleteknek pontosan akkor van közös megoldása, ha $\Delta_{n-1}(a) = 0$. Vegyük észre, hogy az (5) és (6) baloldalán lévő polinomok rezultánsa a sorok megfelelő átrendezésével éppen a $\Delta_{n-1}(a)$ determinánst adja. Ezért a 2. Lemma felhasználásával tételünk bizonyítását befejeztük.

IRODALOM

[1] ARNOLD V.I., *A differenciálegyenletek elméletének geometriai fejezetei* (Műszaki Könyvkiadó, Budapest, 1988).

[2] FARKAS, H., GYÖKÉR, S., WITTMANN, M., „Globális egyensúlyi bifurkációk vizsgálata a paraméteres reprezentáció módszerével", *Alk. Mat. Lapok* **14** (1989), 335–364.

[3] FARKAS, H., SIMON, P.L., „Use of the parametric representation method in revealing the root structure and Hopf bifurcation", *J. Math. Chem.* **9** (1992), 323–339.

[4] GILMORE, R., *Catastrophe Theory for Scientists and Engineers* (Wiley, New York, 1981).

[5] GUCKENHEIMER, J. and HOLMES, P.J., *Nonlinear Oscillations, Dynamical Systems and Bifurcations of Vector Fields* (Springer–Verlag, New York, Heidelberg, Berlin, 1983).

[6] KORN, G.A., KORN, T.M., *Matematikai kézikönyv műszakiaknak* (Műszaki Könyvkiadó, Budapest, 1975).
[7] PERKO, L., *Differential Equations and Dynamical Systems* (Springer–Verlag, New York, Heidelberg, Berlin, 1991).
[8] RÉDEI, L., *Algebra I.* (Akadémiai Kiadó, Budapest, 1954).
[9] WIGGINS, S., *Introduction to Applied Nonlinear Dynamical Systems and Chaos* (Spriger–Verlag, New York, Heidelberg, Berlin, 1990).

(Beérkezett: 1993. szeptember 20.)

SIMON L. PÉTER ÉS FARKAS HENRIK
BUDAPESTI MŰSZAKI EGYETEM, FIZIKAI INTÉZET
H–1521
E-MAIL: H2310SIM@ELLA.HU
H2306FAR@ELLA.HU

THE INVESTIGATION OF THE ROOT STRUCTURE OF POLYNOMIALS WITH THE PARAMETRIC REPRESENTATION METHOD

P. L. SIMON and H. FARKAS

In the course of the investigation of dynamical systems we often encounter the following problems: how many real roots has a polynomial, or how many roots has a polynomial with positive, negative and zero real part. In this paper we study the change of these qualitative properties, especially the loss of stability when the coefficients of the polynomial are varied. We establish a relation between the Routh-Hurwitz criterion and the Hopf bifurcation. We use the parametric representation method to reveal the root structure and determine its dependence on the coefficients of cubic and quartic polynomials.

Alkalmazott Matematikai Lapok **17** (1993) 57–69

NEMLINEÁRIS ÚTKÖVETŐ MÓDSZER TARTÓSZERKEZETEK STABILITÁSVIZSGÁLATÁRA

I. REGULÁRIS PONTOK

CSÉBFALVI A.

PÉCS

A dolgozatban egy olyan egységes, nemlineáris útkövető módszert ismertetünk, amely egyaránt alkalmas a reguláris és szinguláris pontok, az elsődleges és másodlagos állapotváltozási görbék meghatározására. A módszer a stabilitáselmélet perturbációs technikájának és a klasszikus lineáris homotópia módszernek egy ötvözete, amely a direkt módszerek információit is szolgálja.

1. Bevezetés

A módszer alapgondolatát az a felismerés adta, hogy a KOITER (1945), THOMPSON–HUNT (1973), illetve RIKS (1984) nevéhez fűződő stabilitáselmélet és a homotópia elvén alapuló útkövető módszer kiindulási alapja azonos, mindkettő lokális sorfejtésen alapszik. Az elsőrendű implicit differenciálegyenletek elméletén alapuló klasszikus homotópia módszer (RHEINBOLDT (1981), (1986); KUBICEK (1976); ABBOT (1978); KAMAT–WATSON–VENKAYYA (1983)) a vizsgált pont környezetét lineáris sorfejtéssel írja le. A stabilitásvizsgálat viszont megkívánja véges számú magasabbrendű tag figyelembevételét is.

A módszer kidolgozása során feltételezzük, hogy a szerkezetet konzervatív erőrendszer terheli, ezáltal a teljes potenciális energia függvény felírható a csomóponti eltolódások, illetve a teherparaméter függvényeként:

$$V(u_i, \lambda), \tag{1.1}$$

ahol u_i $(i = 1, 2, \ldots, n)$ a csomóponti eltolódásokat, λ pedig a teherintenzitási paramétert jelöli. Feltételezzük, hogy a $V(u_i, \lambda)$ teljes potenciális energia függvény az $n + 1$ dimenziós térben egy egyértelműen megadható sima függvény.

Az egyensúlyi egyenletek a potenciális energia függvény stacionaritási elve alapján adódnak:

$$V_{,i}(u_i, \lambda) = 0 \tag{1.2}$$

ahol $(\,)_{,i}$ az u_i csomóponti eltolódások szerinti parciális deriváltakat jelöli.

Feltételezzük, hogy tehermentes, deformálatlan $(u_i, \lambda) = (0,0)$ állapotban a szerkezet stabil egyensúlyi állapotban van, illetve, hogy a λ teherintenzitási paraméter növelésével kezdetben stabil egyensúlyi úton halad. Bizonyítható (ABBOTT (1978)), hogy stabil egyensúlyi állapotban a $V_{,ij}$ $(i = j = 1, 2, \ldots, n)$ Jacobi-mátrix pozitív definit, instabil egyensúlyi pontokban indefinit, kritikus pontokban a $V_{,ij}$ Jacobi-mátrix szinguláris lesz.

Jelölje y_k, $k = 1, 2, \ldots, n+1$ a csomóponti eltolódások és a teherintenzitási paraméter összekapcsolásával adódó vektort, ahol $y_i = u_i$, $i = 1, 2, \ldots, n$; és $y_{n+1} = \lambda$.

Jelölje $\tilde{V}_{,ik}$ $(i = 1, 2, \ldots, n;\ k = 1, 2, \ldots, n+1)$ ún. kibővített Jacobi-mátrixot, amely az y_k változók szerinti parciális deriváltakat tartalmazza. A kibővített Jacobi-mátrix első n oszlopa az eredeti Jacobi-mátrixnak felel meg, az $n+1$-edik oszlop pedig a λ szerinti deriváltakat jelöli.

2. Az útkövető módszer lényege

Induljunk ki a rendszer egy ismert y_k^a stabil egyensúlyi pontjából, például az $y_k^a = (0,0)$ pontból. Mivel y_k^a egyensúlyi pont, ezért kielégíti a $V_{,i}\,|^a = 0$ egyensúlyi egyenletet. Ebben az esetben $V_{,ij}$ $(i = j = 1, 2, \ldots, n)$ Jacobi-mátrix pozitív definit, invertálható, így y_k^a pont környezetében a megoldás egyértelmű. A szinguláris pontok környezetének vizsgálatát, valamint a szinguláris pontok típusának meghatározását a CSÉBFALVI (1993a) cikk tárgyalja.

Írjuk fel az egyensúlyi utat az y_k^a pont környezetében egy alkalmasan megválasztott s paraméter függvényében. Feltevésünknek megfelelően y_k^a elegendően kicsiny környezetében az egyensúlyi út egy folytonos görbe, amely a következő alakban állítható elő:

$$y_k(s) = y_k^a + y_k^{a(1)} s + \frac{1}{2!} y_k^{a(2)} s^2 + \frac{1}{3!} y_k^{a(3)} s^3 + \ldots . \tag{2.1}$$

ahol $y_k^{a(1)}, y_k^{a(2)}, \ldots$, pedig az $y_k(s)$ függvény s szerinti deriváltjait jelöli az y_k^a pontban.

Lokális paraméternek az ívhosszat választottuk, amelynek egyértelmű előnye, hogy lehetőséget nyújt több pontra támaszkodó módszerek alkalmazására.

Az egyensúlyi feltétel a következő alakban írható:

$$V_{,i}(y_k(s)) = 0\,, \tag{2.2}$$

amiből, kihasználva V sima voltát, egymásutáni differenciálással a következő egyenletek adódnak:

$$\tilde{V}_{,ij} y_j^{a(1)} = 0\,, \tag{2.3}$$

$$\tilde{V}_{,ij} y_j^{a(2)} = -\tilde{V}_{,ijk} y_j^{a(1)} y_k^{a(1)}\,, \tag{2.4}$$

$$\tilde{V}_{,ij} y_j^{a(3)} = -3\tilde{V}_{,ijk} y_j^{a(1)} y_k^{a(2)} - \tilde{V}_{,ijk\ell} y_j^{a(1)} y_k^{a(1)} y_\ell^{a(1)}\,, \tag{2.5}$$

ahol $i = 1, 2, \ldots, n$; $j, k, \ell = 1, 2, \ldots, n+1$.

A sorozat hasonló módon folytatható, az egyes egyenletek az előzőekre épülnek, az egyenletek baloldala szerkezetileg azonos. A (2.2) egyenlet a klasszikus homotópia módszer alapegyenlete. A (2.3)–(2.5) egyenletek egy implicit differenciálegyenlet-rendszer kiindulási adatait adják meg. A (2.1) felírási mód az egypontos explicit Taylor módszernek (HOUWEN (1977)) felel meg.

A (2.3)–(2.5) egyenletek baloldalán álló $\tilde{V}_{ij}$ kibővített Jacobi-mátrix $n \times (n+1)$-es, amelynek első $n \times n$-es blokkja az egyensúlyi egyenletek Jacobi-mátrixa, a mátrix $n+1$-edik oszlopa pedig a teherintenzitási paraméter szerinti parciális deriváltakat tartalmazza. Az egyensúlyi egyenletek Jacobi-mátrixa szimmetrikus, rendszerint sáv struktúrájú. A kibővített Jacobi-mátrix utolsó oszlopának kitöltöttsége a terhelés függvényében más és más lehet.

Az egyensúlyi út paraméteres alakja egy prediktor-korrektor típusú útkövető módszer kiindulási alapja, amelyben a

- prediktor fázis egy közönséges implicit differenciálegyenlet-rendszer megoldását jelenti, amelyet
- korrektor fázisként egy nemlineáris egyenletrendszer megoldása követ.

Induljunk ki egy pontos megoldásnak tekinthető y_k^a pontból, amelyből — az $y_k^{a(1)}, y_k^{a(2)}, \ldots$ deriváltak meghatározása után egy alkalmasan megválasztott differenciálegyenlet megoldó módszer segítségével — a következő y_k^b pont $\hat{y}_k^b$ első közelítését kapjuk, amely a következő alakú:

$$\hat{y}_k^b = y_k^a + y_k^{a(1)} s^b + \frac{1}{2!} y_k^{a(2)} \left(s^b\right)^2 + \cdots + \frac{1}{r!} y_k^{a(r)} \left(s^b\right)^r , \tag{2.6}$$

ahol r a közelítés rendjét, s^b az y_k^a ponthoz tartozó optimális lépésközt jelöli. Optimális lépésközön első megközelítésként kizárólag azt értve, hogy az $\hat{y}_k^b$ közelítés az y_k^b „pontos" értéktől legfeljebb ε_k^b hibával tér el.

A hagyományos útkövető módszerek az egyensúlyi utat kizárólag egyensúlyi pontok sorozatával határozzák meg, vagyis nem szolgáltatnak információt két egyensúlyi pontot összekötő útvonal tényleges alakjáról. A magasabb rendű előrejelzés megteremti annak a lehetőségét, hogy ne csupán egyensúlyi pontokat, hanem az egyensúlyi pontokat összekötő egyensúlyi íveket határozzuk meg.

3. A differenciálegyenlet-rendszer kiindulási adatai

Vizsgáljuk meg, hogyan határozhatók meg a differenciálegyenlet kiindulási adatait képező $y_k^{a(1)}, y_k^{a(2)}, \ldots$ deriváltak. Az $y_k^{a(1)}$ deriváltakra vonatkozó (2.3) egyen-

letek részletes alakja:

$$(3.1)\qquad \left[\begin{array}{cccc|c} V_{,11} & V_{,12} & \dots & V_{,1n} & V_{,1n+1} \\ V_{,21} & V_{,22} & \dots & V_{,2n} & V_{,2n+1} \\ \vdots & \vdots & \dots & \vdots & \vdots \\ V_{,n1} & V_{,n2} & \dots & V_{,nn} & V_{,nn+1} \end{array}\right] \left\{\begin{array}{c} y_1^{a(1)} \\ y_2^{a(1)} \\ \vdots \\ y_n^{a(1)} \\ y_{n+1}^{a(1)} \end{array}\right\} = \left\{\begin{array}{c} 0 \\ 0 \\ \vdots \\ 0 \end{array}\right\}$$

Az egyenletrendszer a $\tilde{V}_{,ij}$ kibővített Jacobi-mátrix nullterét határozza meg. Vizsgáljuk meg, hogy a (3.1) egyenletnek mikor van egyértelmű megoldása. A normált megoldás irányát, a nulltér folytonosságára támaszkodva, szögkorlátozási feltétellel (KAMAT–WATSON–VENKAYYA (1983)), különbségkorlátozási feltételekkel (LI–YORKE (1980)), illetve a Jacobi determináns előjelén alapuló módszerrel (RHEINBOLDT (1981)) állíthatjuk be.

Ha az állapotváltozási egyenlet Jacobi-mátrixa invertálható, vagyis rangja n, a feladat egyértelműen megoldható, mivel ekkor a nulltér egy egydimenziós altér.

A magasabbrendű deriváltakra vonatkozó egyenletrendszerek inhomogének, amelyek együttható mátrixa $n \times (n+1)$-es, az ismeretlenek száma pedig $n+1$. Így az egyenletrendszernek végtelen sok megoldása van.

A probléma kézenfekvő, egyértelmű megoldását akkor kapjuk, ha kihasználva a V teljes potenciális energia függvény sima voltát, az inhomogén egyenletrendszer megoldásaként az ún. általánosított megoldást választjuk. Az általánosított megoldást (lásd: NOBLE (1969), POPPER és CSIZMÁS (1993)) a kibővített Jacobi-mátrix $\left(\tilde{V}_{,ij}\right)^{+}_{k\ell}$, $(k = 1, 2, \dots, n+1;\ \ell = 1, 2, \dots, n)$ Moore–Penrose pszeudóinverze szolgáltatja.

Hangsúlyoznunk kell, hogy ez az inverz adja az összes magasabb rendű deriváltakra vonatkozó egyenletek megoldását is és a javító fázisban egy Newton-típusú iteráció alapjául szolgálhat. A numerikus megoldás lehetőségeit WATSON (1986), illetve DESA és tsai (1992) vizsgálták. A numerikus kezelés figyelemre méltó sajátossága, hogy a nulltér meghatározása, illetve a pszeudóinverz előállítása ugyanabban a fázisban történhet.

A lineáris közelítésen alapuló klasszikus homotópia módszerrel kapcsolatos eddigi vizsgálatok, például WATSON–KAMAT–REASER (1985), a módszer alkalmazhatóságának korlátját az egyszerű elágazások megjelenésében látták. Megítélésünk szerint a magasabbrendű deriváltak figyelembevételével ez a korlát feloldható. Az elágazási pontok környezetének vizsgálatát részletesen a témakörhöz kapcsolódó (CSÉBFALVI (1993a)) cikk tárgyalja.

4. A prediktor fázis

Ha az előzőekben körvonalazott módszerhez kapcsolódó problémák körét tekintjük, akkor nyilvánvaló, hogy a differenciálegyenlet-rendszert megoldó prediktor

fázis kidolgozása módszertanilag nem jelent különösebb nehézséget, mivel a megoldandó feladat a differenciálegyenletek elméletének egyik, teljes mértékben kidolgozott esetét jelenti.

A megoldási lehetőségek vizsgálatakor ilymódon a szakirodalomra, többek között HALL–WATT (1978), SHAMPE–GORDON (1975) munkáira hagyatkoztunk. A lineáris, egypontos útkövető módszerek esetében RHEINBOLDT (1986), lineáris, többpontos módszerek esetében ABBOTT (1980) és KUBICEK (1976) munkáit emelném ki. A megoldandó feladat szerkezetileg az explicit egypontos Taylor módszernek felel meg, ezért kézenfekvő a megoldást a Taylor módszerrel előállítani.

A módszer alkalmazása mellett szól, hogy az optimális, maximum ε_k^b hibájú megoldáshoz tartozó lépésköz a Taylor-sor maradéktagjának becsült értékéből egyszerűen meghatározható.

A módszer lényegét a következőkben foglalhatjuk össze. Jelölje:

$$y_k^b = y_k^a + y_k^{a(1)} s_k^b + \frac{1}{2!} y_k^{a(2)} \left(s_k^b\right)^2 + \cdots + \frac{1}{r!} y_k^{a(r)} \left(s_k^b\right)^r + R_{r+1}\left(s_k^b\right) \tag{4.1}$$

$$R_{r+1}\left(s_k^b\right) = \frac{1}{(r+1)!} y_k^{\xi(r+1)} \left(s_k^b\right)^{r+1} , \tag{4.2}$$

ahol y_k^b az s_k^b lépésközhöz tartozó pontos megoldást, a maradéktagban szereplő ξ a $[0, s_k^b]$ intervallum egy pontját jelöli, r a (2.3) egyenletrendszerrel kezdődő sorozat számított elemeinek száma.

Az r konkrét értékének meghatározása stabilitáselméleti megfontolások (GÁSPÁR és DOMOKOS (1991), DOMOKOS és GÁSPÁR (1992)) alapján, a vizsgált szerkezet ismeretében történik, tehát a jósló fázis szintjén adottságként jelenik meg.

Az r értékével kapcsolatos feltevésből következik, hogy az $r+1$–edik deriválttól kezdődően csak becsült — az előző lépések alapján extrapolált — $\hat{y}_k^{a(r+1)}, \hat{y}_k^{a(r+2)}, \ldots$ értékekre támaszkodhatunk.

A maradéktag becsült értéke másképpen is felírható:

$$\hat{R}_{r+1}\left(s_k^b\right) = \frac{1}{(r+1)!} \hat{y}_k^{a(r+1)} \left(s_k^b\right)^{(r+1)} + \\ + \frac{1}{(r+2)!} \hat{y}_k^{a(r+2)} \left(s_k^b\right)^{(r+2)} + \hat{R}_{r+3}\left(s_k^b\right) , \tag{4.3}$$

amelyből az $\hat{R}_{r+3}\left(s_k^b\right)$ tag elhanyagolásával az $\hat{R}_{r+1}\left(s_k^b\right)$ maradéktagban (4.3) szereplő $\hat{y}_k^{\xi(r+1)}$ derivált becsült értékére az alábbi lineáris összefüggést kapjuk:

$$\hat{y}_k^{\xi(r+1)} = \hat{y}_k^{a(r+1)} + \frac{1}{r+2} \hat{y}_k^{a(r+2)} \left(s_k^b\right) . \tag{4.4}$$

A (4.4) felírásmód annak a feltételezésnek felel meg, hogy a $[0, s_k^b]$ intervallumban az $\hat{y}_k^{\xi(r+1)}$ derivált lineáris függvény.

Az ε_k^b hibahatár ismeretében az egyes változókhoz tartozó s_k^b lépésköz egyszerű algebrai eszközökkel meghatározható:

$$\hat{R}_{r+1}\left(s_k^b\right) = \operatorname{sign}\left(\hat{R}_{r+1}\left(s_k^b\right)\right)\varepsilon_k\,. \tag{4.5}$$

Az s_k^b, $k = 1, 2, \ldots, n+1$ értékek alapján

$$s^b = \operatorname{Min}\{s_1, s_2, \ldots, s_m\} \tag{4.6}$$

adódik az y_k^a ponthoz tartozó s^b optimális lépésközre. Ha a halmozódó hibák hatásától eltekinthetünk (a Newton—típusú javító fázis miatt y_k^a pontos), akkor $\hat{y}_k^b$ az y_k^b ε_k^b sugarú környezetében helyezkedik el, ami a javító fázisban implicit feltételként jelenik meg. Az ε_k^b értékek meghatározásával kapcsolatos kérdéseket az 5. pontban tárgyaljuk, mivel ezek alapvetően a javító fázis részét képezik.

A Taylor módszer alternatíváját, a numerikusan stabil többpontos, magasabb rendű Adams formulák jelentenék, de ezek tárgyalásától eltekintünk. Ennek indokaként, a szakirodalmi eredményekre hivatkozhatunk (ABBOTT (1980), KUBICEK (1976)), amelyek azt bizonyítják, hogy az útkövető módszer pontosságát alapvetően a javító fázis pontossága határozza meg, vagyis a Taylor, illetve az Adams módszerek között számottevő különbségek nincsenek. A választást megerősíti az a tény, hogy a korszerű, elosztott differenciákon alapuló, változó rendű és lépésközű formulák esetében racionalitási okokból a lépésközváltás felezésre, kétszerezésre korlátozódik.

5. A korrektor fázis

A lineáris differenciálegyenletek elméletén alapuló klasszikus homotópia módszer első alkalmazásaiban (KAMAT–WATSON–VENKAYYA (1983), KUBICEK (1976)) a korrektor fázis beépítésének szükségessége lényegében még nem merült fel. A szerzők a kielégítő pontosságú útkövetést a lépésköz és a pontszám (egypontos–többpontos) megválasztásában, illetve később implicit differenciálegyenlet bázisú korrektorok alkalmazásában látták.

Kezdetben a szerzők nem tulajdonítottak különösebb jelentőséget annak a ténynek, hogy egy hosszabb útvonalon a számított görbe jelentős mértékben eltérhet az elméleti görbétől, ami egyrészt a lokális, másrészt a globális hibák halmozódásából fakad. Nyilvánvaló, hogy a homotópia módszer stabilásvizsgálatokra történő alkalmazásakor ezt a nagyvonalúságot nem engedhetjük meg. A halmozódó hibák miatt kialakult görbe a stabilitásvesztési sajátosságok tekintetében is eltérhet a valódi görbétől.

A differenciálegyenletek elméletére támaszkodó implicit prediktorok alkalmazása a problémát igazából nem oldotta meg. Ezért viszonylag hamar felmerült az a gondolat (ABBOTT (1980), RHEINBOLDT (1986)), hogy a differenciálegyenleteken

alapuló jósló fázist, bizonyos időközönként a jól ismert Newton-típusú korrektor módszerrel célszerű kombinálni. A Newton típusú korrektor alkalmazása kézenfekvő, hiszen ez egy olyan módszer, amellyel az „elcsúszási" hajlandóságot mutató görbe az eredeti pályára visszakényszeríthető.

A homotópia módszer és egy lokálisan konvergens Newton-típusú korrektor összekapcsolása nem tekinthető teljes mértékben megoldott prblémának, hiszen a konkrét megvalósítás számtalan tényezőnek, például a lokális parametrizálás módszerének függvénye.

Megjegyezzük, hogy paraméterként a lokálisan legjobb változót használó lineáris módszerek megközelítési logikája mögött lényegében az áll, hogy a Jacobi-mátrix rangcsökkenése maximum egy lehet, vagyis a szingularitásból adódó problémák egy alkalmas sor hozzávételével a prediktor, illetve korrektor fázisban egyaránt megoldhatók. Lokálisan legjobb változónak azt a változót nevezzük, amelynek koordináta iránya a legkisebb szöget zárja be a lineárisan extrapolált útvonallal.

Mielőtt a javító fázis részleteit ismertetnénk, foglalkoznunk kell az előzőekben nyitvahagyott kérdéssel, nevezetesen az ε_k^b tűréshatár megválasztásával. Mivel a jósló fázishoz egy lokálisan konvergens javító fázis kapcsolódik, az ε_k^b értékeket úgy kellene megválasztanunk, hogy biztosak lehessünk, hogy az $\hat{y}_k^b$ induló megoldás az y_k^b pontos megoldáshoz konvergál. Természetesen ez igaz minden elegendően kicsiny s^b lépésközre, hiszen y_k^a — feltevésünk szerint — az előző javító fázis, egy konvergens iteráció eredménye.

A prediktor–korrektor fázisok összekapcsolásának problémakörét elméletileg RHEINBOLDT (1981) vizsgálta.

Legyen az y_k^b a pontos megoldás, amelyre teljesül $V_{,i}\left(y_k^b\right) = 0$ és tételezzük fel, hogy rendelkezésünkre áll egy lokálisan konvergens iterációs eljárás. Definíció szerint az y_k^b ponthoz hozzárendelhető egy pozitív ϱ_k^b érték, úgy, hogy ha $\hat{y}_k^b$ induló megoldás y_k^b ϱ_k^b sugarú környezetében helyezkedik el, akkor az eljárás az y_k^b ponthoz konvergál.

A ϱ_k^b konvergencia sugár becslésére számos módszer található a szakirodalomban (ABBOTT (1978), RHEINBOLDT (1981)). Ezek a módszerek azonban igen érzékenyek a megoldandó probléma sajátosságaira és a konvergencia sugár becsléseként meglehetősen konzervatív eredményeket szolgáltatnak. Mindegyik módszer azon a feltételezésen alapul, hogy az előző lépések minőségi jellemzői (pl. konvergencia sugár) elegendő információt tartalmaznak a következő lépés konvergencia sugarának meghatározásához. Bizonyítható (RHEINBOLDT (1981)), hogy az előző lépések minőségi jellemzői nem nyújtanak elegendő információt a konvergencia sugár alsó, illetve felső korlátjának meghatározásához, de az előző lépés (lépések) alapján megadható egy olyan ε_k^b tűréshatár, amely annak ellenére, hogy nem tekinthető a ϱ_k^b konvergencia sugár becslésének, a gyakorlati szempontoknak mégis teljes mértékben megfelel.

Jelölje ε_k^a az előző korrektor fázishoz tartozó

$$\varepsilon_k^a = y_k^a - \hat{y}_k^a \tag{5.1}$$

eltérést, ahol $\hat{y}_k^a$ az iteráció induló megoldása, y_k^a az iteráció eredménye. Az eljárás azon a természetes felismerésen alapszik, hogy az egymásután következő lépésekben a megoldandó probléma minőségi jellemzői viszonylag lassan változnak, így az előző lépés alapján kapott ε_k^a eltérés az aktuális ε_k^b tűréshatárnak elfogadható első közelítése. Ettől csak abban az esetben térünk el, ha ez a választás numerikus instabilitást (pl. egyre csökkenő lépéshosszt) eredményezne, vagy az így adódó s^b lépéshossz a lépéshossz relatív és abszolút változását szabályozó feltételek valamelyikét sértené.

Ezek után a megoldandó feladat a következőképpen fogalmazható meg. Keressük a

$$V_{,i}\left(y_k^b\right) = 0 \tag{5.2}$$

egyenletrendszer megoldását az $\hat{y}_k^b$ becslésből kiindulva.

A Newton-típusú javító fázis kidolgozásakor, az előzőekben vázolt szakirodalmi előzményekre támaszkodva, arra törekedtünk, hogy lehetőleg elkerüljük a kiindulási egyenlet kibővítését, vagyis a korrektor fázis feltételes minimalizálási feladatként történő kezelését.

Módszerünkben, az ε_k^b tűréshatár alkalmas megválasztása miatt, a feltételkezelés lényegében csak ellenőrzést, az iteráció eredményeképpen kapott y_k^b pont minősítését jelenti, magába az iterációs eljárásba nem épül be.

Ha a feladatot feltételes minimalizálási problémaként oldanánk meg, akkor ez a Jacobi-mátrix módosítását eredményezné, hiszen a büntető függvényben szereplő egyenlőségi feltétel deriváltjai megjelennének a Jacobi-mátrixban, ezáltal a prediktor, illetve korrektor fázis alapmátrixa különbözővé válna. A módosítás — a többletmunkán túlmenően — nem eredményezné a megoldási folyamat biztonságának növelését, mivel esetünkben csak igen extrém, r rögzített értékéhez képest erősen görbülő útvonalon várható, hogy az explicit módon nem kezelt feltétel aktivvá váljon. Ilyen esetekben viszont elegendő információt szolgáltat a rendszer az s^b lépéshossz alkalmas csökkentéséhez.

Írjuk fel a Newton módszer első iterációs lépését az alábbi formában:

$$\overset{(1)^b}{y_k} = \hat{y}_k^b + \hat{\alpha}^b \Delta \hat{y}_k^b\,, \tag{5.3}$$

ahol $\overset{(1)^b}{y_k}$ a prediktor fázis utáni első korrektor lépés eredményét jelöli, $\hat{\alpha}^b$ skalár az aktuális lépéshossz, $\Delta \hat{y}_k^b$ pedig az alábbi egyenletből határozható meg:

$$V_{,ik} \Delta \hat{y}_k^b = -V_{,i}\,. \tag{5.4}$$

Az $\hat{\alpha}^b$ lépéshossz meghatározása a Powell elven alapul (lásd: POPPER és CSIZMÁS (1993)), amely azt a feltételt jelenti, hogy az iteráció eredménye nem lehet rosszabb közelítés, mint amilyen az iteráció kiindulási alapja.

A fentieknek megfelelően, a korrektor fázis alapját egy olyan n egyenletet tartalmazó lineáris egyenletrendszer megoldása képezi, amelyben a változók száma

$n+1$, az együttható mátrixról pedig csupán annyit tudunk, hogy rangja, általános esetben legfeljebb n. Az együttható mátrix rangjára vonatkozó megjegyzés azt jelenti, hogy a kritikus pontok megjelenése a mátrix rangját csökkentheti.

Ha az egyenletrendszer megoldását az összes szóbajöhető eset figyelembevételével vizsgáljuk, akkor újra az általánosított megoldáshoz, vagyis a kibővített Jacobi-mátrix $\left(\tilde{V}_{,ij}\right)^{+}_{k\ell}$, $(k = 1,2,\ldots,n+1;\ \ell = 1,2,\ldots,n)$ Moore-Penrose–féle pszeudóinverzhez jutunk.

A megoldandó problémával kapcsolatban itt csupán azt emelnénk ki, hogy a legkisebb négyzetek módszeréhez hasonló alapelvekkel jellemezhető, vagyis az egyenletrendszer egyértelműen meghatározott, minimális normájú megoldása a $V_{,i}V_{,i}$ függvényt minimalizálja.

Hangsúlyoznunk kell, hogy az általánosított inverz alkalmazásával a korrektor fázis matematikai megoldó módszerének magja ugyanaz, mint amit a prediktor fázisban használtunk. Az álatánosított inverz és a nulltér ugyanannak a lineáris algebrai eljárásnak az eredménye.

Tudjuk, hogy a prediktor-korrektor fázisban a módszer az előrejelzés r rendjét adottságként kezeli. Ez természetesen nem jelenti azt, hogy az adott s^b lépésben minden változóra, minden tag r-ig bezárólag valóban hordoz információt, mivel az egyes változók görbülete az adott intervallumban jelentős mértékben eltérhet, egyrészt egymáshoz, másrészt az előző intervallumhoz viszonyítva.

Módszerünkben a korrektor fázis az eredményül adódó y_k^b pont minősítésével végződik. Ennek célja kettős, egyrészt ellenőrízni kell, hogy az adott pont valóban az $\hat{y}_k^b$ közelítéshez tartozó pontos megoldás, másrészt meg kell határozni azokat az információkat, amelyek az adott $[0, s^b]$ intervallumhoz tartozó egyensúlyi ív kielégítően pontos meghatározásához szükségesek. A feladatnak ez a második része egyben az egyes változókra vonatkozóan a közelítés tényleges rendjének meghatározását is jelenti.

A megoldandó feladat első része nem jelent különösebb problémát, mivel az adott y_k^b pont akkor tekinthető $\hat{y}_k^b$ pontos megoldásának, ha

$$y_k^b - \hat{y}_k^b \leq \varepsilon_k^b \,. \tag{5.5}$$

A feladat második része lényegében az általánosított Rolle-tétel (pl. SZÉP (1972)) alkalmazási feltételeinek ellenőrzését jelenti.

Ennek ismertetése előtt azonban meg kell jegyeznünk, hogy az y_k^b pontos megoldás (a ponthoz tartozó ívhossz szerinti deriváltak) ismeretében az adott $[0, s^b]$ intervallum belsejére vonatkozóan minden derivált becslése pontosítható, hiszen — extrapolált értékek helyett — megbízhatóbb interpolált értékekre támaszkodhatunk. Ez a lehetőség — az y_k^a, illetve az y_k^b végpontokat is beleértve — különösen a már eleve becsült értékű $r+1$, $r+2$ rendű deriváltak esetében eredményez javulást a becslési pontosságban. Ennek megfelelően feltehetjük, hogy az első becsült értékű ($r+1$ rendű) ívhossz szerinti $\hat{y}_k^{a\,(r+1)}$, $\hat{y}_k^{b\,(r+1)}$ deriváltak pontosak, hiszen értékük egy legalább három pontos interpoláció eredményeképpen adódik. Ennek megfelelően ezekre a deriváltakra a továbbiakban $\hat{y}_k^{a\,(r+1)}$, $\hat{y}_k^{b\,(r+1)}$ jelöléssel hivatkozunk.

Képezzük a következő sorozatot:

$$d_k^1 = (y_k^b - y_k^a)/s^b\,, \tag{5.6}$$

$$d_k^i = i!\left(y_k^b - y_k^a - \sum_{j=1}^{i-1} y_k^{a(1)}(s^b)^j/j!\right) \Big/ (s^b)^i, \quad i = 2,3,\ldots,r,r+1\,. \tag{5.7}$$

Jelölje I_k^b azon i, $i = 1,2,\ldots,r+1$ indexek halmazát, amelyekre fennáll az alábbi feltétel:

$$I_k^b = \left\{ i \mid y_k^{a(1)} \le d_k^i \le y_k^{b(1)} \right\}. \tag{5.8}$$

Legyen r^k az I_k^b indexhalmaz maximális indexű eleme. A „visszaszámolási eljárásnak" megfelelően r^k az y_k^b változóra a közelítő függvény tényleges rendjét, $d_k^{r^k}$ pedig y_k^b r^k rendű maradéktagjában szereplő derivált értékét adja meg:

$$R_{r^k}(s^b) = d_k^{r^k}(s^b)^{r^k}/\,r^k! \tag{5.9}$$

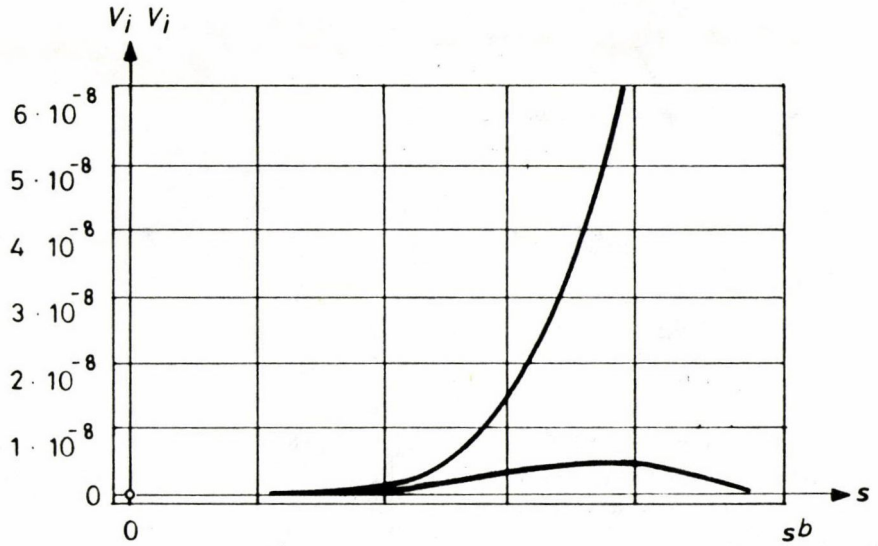
A maradéktagban szereplő deriváltat a $[0,s^b]$ intervallumban lineáris függvénnyel közelítve a következő összefüggés adódik:

$$R_{r^k}(s) = \left(y_k^{a(r^k)} + s\frac{d_k^{r^k} - y_k^{a(r^k)}}{s^b}\right)(s)^{r^k}/\,r^k!, \qquad s \in [0,s^b]\,, \tag{5.10}$$

Az y_k^b pontos megoldásából kiindulva, a maradéktag (5.10) becslésével a $[0,s^b]$ intervallum minden pontjának megbízhatósága javítható, vagyis a monoton növekedő $V_{,i}V_{,i}$ hibafüggvény alakulása visszafordítható. A maradéktagban szereplő deriváltak lineáris közelítése a prediktor fázis feltételezéseivel (a lépéshossz meghatározásával) összhangban van. Nemlineáris közelítés szükségessé tenné $[0,s^b]$ intervallum egy vagy több belső pontjában a pontos megoldás ismeretét.

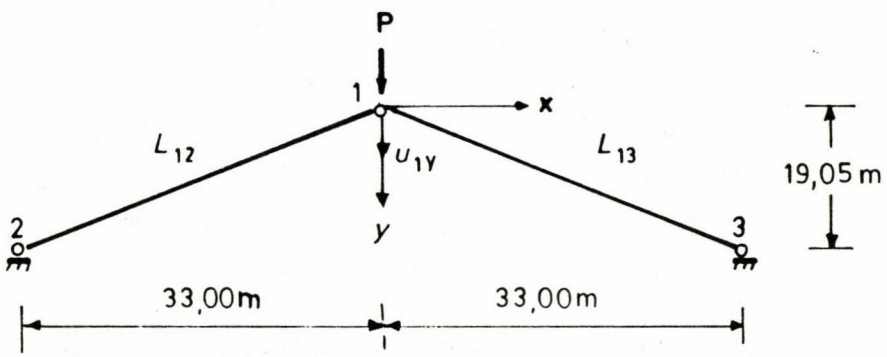
Az 5.1 ábra a lineárisan becsült maradéktag jellegzetes hatását szemlélteti a $V_{,i}V_{,i}$ hiba alakulására. A maradéktag hatására a meredeken ívelő eredeti hibagörbe megfordul, az intervallum második felében a közelítés pontossága, az eredetihez képest, legalább egy nagyságrenddel javul. Ugyanakkor az ábra jól érzékelteti a

nemlineáris becslésben rejlő további lehetőségeket.

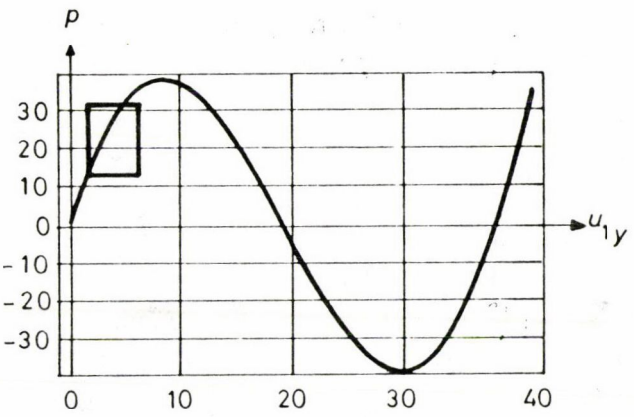


5.1 ábra: A hibanorma alakulása becsült maradéktaggal

A módszer alkalmazási lehetőségeit teszt feladatokban vizsgáltuk. A hibanorma 5.1 ábra szerinti alakulását az 5.2 ábra szerinti egyszerű síkbeli rácsos tartó vizsgálat során kaptuk.



5.2 ábra



5.3 ábra

Az 5.3 ábra a szerkezet állapotváltozási görbéjét szemlélteti. Egy négyzettel jelöltük azt a görbeszakaszt, amelyre az 5.1 ábrán található hibanorma vonatkozik. Mivel az adott görbeszakasz görbülete kicsi, ezért ezen a szakaszon az eljárás nagy lépésekkel halad.

IRODALOM

[1] ABBOTT, J. P., „An efficient algorithm for the determination of certain bifurcation points", *J. Comput. Appl. Math.* **4** (1978), 19–27.

[2] ABBOTT, J. P., „Computing solution arcs of nonlinear equation with a parameter", *The Computer Journal* **23** (1980), 85–89.

[3] CSÉBFALVI, A. and CSÉPFALVI, GY., „Geometrically Non-Linear Analysis of Space Trusses Using Sparse Quasi-Newton Algorithms", *Proceedings, 11th International Conference on Mathematical Programming*, Mátrafüred, Hungary, March 21–26, 1992,, 2–3.

[4] CSÉBFALVI, A. and VÁSÁRHELYI, A., „Some computational aspects of nonlinear space trusses", *Acta Tech. Acad. Sci. Hung.* (1993), (megjelenés alatt).

[5] CSÉBFALVI, A. and CSÉBFALVI, GY., „Post-buckling analysis of frames by a hybrid path-following method", *Generalized Convexity* (S. Komlósi, T. Rapcsák, S. Schaible, eds.) (Springer Verlag, 1993), 311–321.

[6] CSÉBFALVI, A., „Nemlineáris útkövető módszer tartószerkezetek stabilitásvizsgálatára II. Elágazási és határpontok", *Alkalmazott Matematikai Lapok* (1993a), (megjelenés alatt).

[7] CSÉBFALVI, A., „Szakaszonként folytonosan differenciálható energia függvénnyel jellemezhető tartószerkezetek stabilitásvizsgálata", *Alkalmazott Matematikai Lapok* (1993b), (megjelenés alatt).

[8] CRANDALL, M. G. and RABINOWITZ, P. H., „Bifurcation from simple eigenvalues", *J. Functional Analysis* **8** (1971), 321–340.

[9] DESA, C., IRANI, K. M., WATSON, L. T. and WALKER, H. M., „Preconditioned iterative methods for homotopy curve tracking", *SIAM J. Sci. Stat. Comput.* **13** (1992), 30–46.

[10] DOMOKOS, G. és GÁSPÁR, ZS., *Posztkritikus állapot vizsgálata diszkrét és folytonos modellel*, OTKA Kutatási részjelentés (1992), 97–123.

[11] GÁSPÁR, ZS. und DOMOKOS, G., „Über das post-kritische Verhalten von Systemen mit C_n-Symmetrie", *ZAMM* **72** (1992), 4, T 110–T111.

[12] HALL, G. and WATT, J. M., *Modern Numerical Methods for Ordinary Differential Equations* (Oxford, 1978).
[13] HOUWEN, P. J., *Construction of integration formulas for initial value problems* (North-Holland Publishing Company, Amsterdam, 1977).
[14] KAMAT, M. P., WATSON, L. T. and VENKAYYA, V. B., „A quasi-Newton versus a homotopy method for structural analysis", *Comput. & Struct.* **17** (1983), 579–585.
[15] KOITER, W. T., *Over de Stabiliteit van het elastisch Evenwicht*, Dissertation (Delft Technical University, H. J. Paris, Amsterdam, Holland, 1945).
[16] KUBICEK, M., „Dependence of solution of nonlinear systems on a parameter, Algorithm 502", *CACM-Toms* **2** (1976), 98–107.
[17] LI, T. Y. and YORK, J. A., „A Simple Reliable Numerical Algorithm for Following Homotopy Paths", in Robinson, S. M.: Analysis and Computation of Fixed Point (Academic Press, New York, 1980).
[18] NOBLE, B., *Applied linear algebra* (Prentice-Hall, New Jersey, 1969).
[19] POPPER GY. és CSIZMÁS F., *Numerikus módszerek mérnököknek* (Akadémiai Kiadó, Budapest, 1993).
[20] RIKSE, E., „Some Computational aspects of the stability analysis of non-linear structures", *Comp. Meth. Appl. Mech. Eng.* **47** (1984), 219–259.
[21] RHEINBOLDT, W. C., „Numerical analysis of continuation methods for non-linear structural problems", *Comput. & Struct.* **13** (1981), 103–113.
[22] RHEINBOLDT, W. C., *Numerical Analysis of Parametrized Nonlinear Equations*, vol. 7 (J. Wiley & Sons, New York, 1986).
[23] SHAMPE, L. F. and GORDON, M. K., *Computer Solution of Ordinary Differential Equations, The Initial Value Problem* (Freeman & Company, San Fracisco, 1975).
[24] SZÉP J., *Analízis* (Közgazdasági és Jogi Könyvkiadó, Budapest, 1972).
[25] THOMPSON, J. M. T. and HUNT, G. W., *A general Theory of Elastic Stability* (Wiley, New York, 1973).
[26] WATSON, L. T., KAMAT, M. P and REASER, M. H., „A robust hybrid algorithm for computing multiple equilibrium solutions", *Eng. Comput.* **2** (March. 1985), 30–34.
[27] WATSON, L. T., „Numerical linear algebra aspects of globally convergent homotopy methods", *SIAM Rev.* **28** (1986), 529–545.
[28] WRIGGERS, P., WAGNER, W. and MIEHE, C., „A quadratically convergent procedure for the calculation of stability points in finite element analysis", *Computer Meth. Appl. Mech. and Eng.* **70** (1988), 329–347.

(Beérkezett: 1993. október 28.)

CSÉBFALVI ANIKÓ
POLLACK MIHÁLY MŰSZAKI FŐISKOLA
7625 PÉCS, BOSZORKÁNY ÚT 2.

NONLINEAR PATH-FOLLOWING METHOD FOR STABILITY OF STRUCTURES I. REGULAR POINTS

A. CSÉBFALVI

In this paper we present a nonlinear path-following method, which able to detect regular and singular points, and basic and secondary paths of the equilibrium equations. This method based on the perturbation technique and the linear homotopy method and able to compute the information of direct methods.

Alkalmazott Matematikai Lapok **17** (1993) 71–86

NEMLINEÁRIS ÚTKÖVETŐ MÓDSZER TARTÓSZERKEZETEK STABILITÁSVIZSGÁLATÁRA

II. ELÁGAZÁSI ÉS HATÁRPONTOK

CSÉBFALVI A.

Pécs

A dolgozatban az állapotváltozási görbék szinguláris pontjainak (az elágazási és határpontok) meghatározására egy útkövető módszert mutatunk be. Az útkövető módszer az állapotváltozási görbe egyes pontjai helyett annak egyes szakaszait határozza meg, ami a szinguláris pontok meghatározásakor kihasználható.

1. Bevezetés

A kritikus pontok felderítésére szolgáló módszerek egyöntetűen azon a tényen alapulnak, hogy egy kritikus pont környezetében a szerkezet minőségi jellemzői megváltoznak.

Az egyes módszerek csak abban különböznek egymástól, hogy a kritikus pont jelzésére milyen minőségi jellemzőt használnak. Mivel számos rendszerjellemző hordoz egyenértékű információt a kritikus pontokkal kapcsolatban, ezért nem véletlen, hogy a témakör szakirodalma igen széleskörű. Az útkövető módszerekhez kapcsolódó kritikus pont felderítő eljárások témakörében WRIGGERS–WAGNER–MIEHE (1988), WRIGGERS–SIMO (1990) és ERIKSSON (1988), (1989), (1991) munkáit emelnénk ki.

A szakirodalomban ismertetett eljárások közös vonása, hogy az adott útvonal szakasz kezdő- és végpontjához tartozó minőségi jellemzőket figyelik, és ezek változásából következtetnek a kritikus pont jelenlétére.

Megjegyezzük, hogy az irodalomban található felderítő módszerek alapjában véve csak a kritikus pont jelenlétének tényét jelzik. A kritikus pont tényleges meghatározása egy további lépésben, az ún. kibővített egyenletrendszeren alapuló direkt, vagy valamely indirekt módszerrel, például a felezéses eljárással, történik. Mindkét megközelítési irány számos további problémát vet fel. A rendszerint a Newton módszeren alapuló, kvadratikus konvergenciával kecsegtető direkt eljárások érzékenyek az induló megoldás megválasztására. Az indirekt eljárások lassú konvergenciája pedig közismert tény.

Külön problémát jelent az az eset, amikor az adott szakaszon több izolált kritikus pont fordul elő, hiszen ezek a kezdő- és végponthoz kötött vizsgálat miatt rejtve maradhatnak, illetve a vizsgálat a valóságosnál egyszerűbb esetet jelezhet. Teljesen

természetes, hogy ez utóbbi esetben a direkt és indirekt módszerek eredménye egyaránt megbízhatatlan. A vázolt problémákra a kiindulási ponthoz legközelebb álló kritikus pont jelzésére alkalmas módszerek jelenthetnek megoldást. Ezek a módszerek ma még csupán felvetés szintjén jelennek meg a szakirodalomban. ERIKSSON (1989) a direkt — egy kritikus pontot adó —, illetve az indirekt módszerek összehasonlításakor az utóbbiak egyik fontos előnyeként azt emelte ki, hogy a fenti probléma kezelésére alkalmazhatók.

Első ránézésre úgy tűnhet, hogy a kritikus pont jelenlétére utaló jelekből a pont típusára vonatkozó információk is automatikusan adódnak. Ez azonban csak akkor igaz, ha a kritikus pont jelzésére szolgáló módszer nem tartalmaz valamilyen prekoncepciót a kritikus pontok típusával kapcsolatosan. A teherkomponens lokális paraméter szerinti első deriváltjának előjelváltása jelzi a határpont jelenlétét, de nyilvánvalóan nem ad információt a határpont multiplicitásával kapcsolatban.

Hasonló helyzet állhat elő többszörös elágazási pontok esetében, ha a minőségi változás jelzése az eredeti Jacobi-mátrix legkissebb sajátértékének előjelváltozásán alapul. Ez a módszer elágazási pontot jelez, feltételezve, hogy a teherkomponens lokális paraméter szerinti első deriváltja nem vált előjelet. Ebből a jelzésből azonban még nem tudhatjuk, hogy többszörös elágazási pontról van-e szó, hiszen például egy kétszeres elágazási pont jelenlétére az eredeti Jacobi-mátrix első két legkisebb sajátértékének együttes előjelváltásából következtethetünk.

Feltételezzük, hogy a szerkezetet konzervatív erőrendszer terheli, így a teljes potenciális energia függvény felírható a csomóponti eltolódások, illetve a teherparaméter függvényeként:

$$V(u_i, \lambda), \tag{1.1}$$

ahol u_i $(i = 1, 2, \ldots, n)$ a csomóponti eltolódásokat, λ pedig a teherintenzitási paramétert jelöli. Feltételezzük továbbá, hogy a $V(u_i, \lambda)$ teljes potenciális energia függvény az $n + 1$ dimenziós térben egy egyértelműen megadható sima függvény.

Az egyensúlyi egyenletek a potenciális energia függvény stacionaritási elve alapján adódnak:

$$V_{,i}(u_i, \lambda) = 0 \tag{1.2}$$

ahol $(\ )_{,i}$ az u_i csomóponti eltolódások szerinti parciális deriváltakat jelöli.

Feltételezzük, hogy tehermentes, deformálatlan $(u_i, \lambda) = (0, 0)$ állapotban a szerkezet stabil egyensúlyi állapotban van, illetve, hogy a λ teherintenzitási paraméter növelésével kezdetben stabil egyensúlyi úton halad. Stabil egyensúlyi állapotban a $V_{,ij}$ $(i = j = 1, 2, \ldots, n)$ Jacobi-mátrix pozitív definit, instabil egyensúlyi pontokban indefinit, kritikus pontokban a $V_{,ij}$ Jacobi-mátrix szinguláris lesz.

Jelölje y_k, $k = 1, 2, \ldots, n+1$ a csomóponti eltolódások és a teherintenzitási paraméter összekapcsolásával adódó vektort, ahol $y_i = u_i$, $i = 1, 2, \ldots, n$; és $y_{n+1} = \lambda$.

Jelölje $\tilde{V}_{,ik}$ $(i = 1, 2, \ldots, n;\ k = 1, 2, \ldots, n + 1)$ ún. kibővített Jacobi-mátrixot, amely az y_k változók szerinti parciális deriváltakat tartalmazza. A kibővített Jacobi-

mátrix első n oszlopa az eredeti Jacobi-mátrixnak felel meg, az $n+1$–edik oszlop pedig a λ szerinti deriváltakat jelöli.

A kritikus pontok legegyszerűbb típusát a határpontok alkotják. Az y_k^h pontot határpontnak nevezzük, ha y_k^h pont környezetében $V_{,ij}$ nem szinguláris, az y_k^h pontban eredeti $V_{,ij}$ Jacobi-mátrix rangja $n-1$, de a kibővített $\tilde{V}_{,ik}$ Jacobi-mátrix rangja n marad:

$$\operatorname{rang}(V_{,ij}|^h) = n-1 \tag{1.3}$$

$$\operatorname{rang}(\tilde{V}_{,ik}|^h) = n\,. \tag{1.4}$$

A fenti összefüggésben szereplő rang(.) szimbólum a mátrix rangját, vagyis az adott mátrix oszlopvektoraiból (sorvektoraiból) alkotott vektorrendszer lineárisan független elemeinek számát (a vektortér dimenzióját) jelöli.

Az y_k^e kritikus pontot egyszerű elágazási pontnak nevezzük, ha y_k^e környezetében $V_{,ij}$ nem szinguláris, az y_k^e pontban eredeti $V_{,ij}$ Jacobi-mátrix rangja $n-1$, a kibővített $\tilde{V}_{,ik}$ Jacobi-mátrix rangja szintén $n-1$:

$$\operatorname{rang}(V_{,ij}|^e) = n-1 \tag{1.5}$$

$$\operatorname{rang}(\tilde{V}_{,ik}|^e) = n-1\,. \tag{1.6}$$

Az y_k^e kritikus pontot többszörös elágazási pontnak nevezzük, ha y_k^e környezetében $V_{,ij}$ nem szinguláris, az y_k^e pontban eredeti $V_{,ij}$ Jacobi-mátrix rangja $n-d$, a kibővített $\tilde{V}_{,ik}$ Jacobi-mátrix rangja szintén $n-d$:

$$\operatorname{rang}(V_{,ij}|^e) = n-d, \tag{1.7}$$

$$\operatorname{rang}(\tilde{V}_{,ik}|^e) = n-d, \tag{1.8}$$

ahol:

$$2 \le d \le n\,. \tag{1.9}$$

Induljunk ki a rendszer egy ismert y_k^a stabil egyensúlyi pontjából, például az $y_k^a = (0,0)$ ponból. Mivel y_k^a egyensúlyi pont, ezért kielégíti a $V_{,i}|^a = 0$ egyensúlyi egyenletet. Ebben az esetben $V_{,ij}$ $(i = j = 1,2,\ldots,n)$ Jacobi-mátrix pozitív definit, invertálható, így y_k^a pont környezetében a megoldás egyértelmű.

Írjuk fel az egyensúlyi utat az y_k^a pont környezetében egy alkalmasan megválasztott s paraméter függvényében. Feltevésünknek megfelelően y_k^a elegendően kicsiny környezetében az egyensúlyi út egy folytonos görbe, amely a következő alakban állítható elő:

$$y_k(s) = y_k^a + y_k^{a(1)} s + \frac{1}{2!} y_k^{a(2)} s^2 + \frac{1}{3!} y_k^{a(3)} s^3 + \ldots\,. \tag{1.10}$$

ahol $y_k^{a^{(1)}}, y_k^{a^{(2)}}, \ldots$, pedig az $y_k(s)$ függvény s szerinti deriváltjait jelöli az y_k^a pontban.

Lokális paraméternek az ívhosszat választottuk, amelynek egyértelmű előnye, hogy lehetőséget nyújt több pontra támaszkodó módszerek alkalmazására.

Az egyensúlyi feltétel a következő alakban írható:

$$V_{,i}(y_k(s)) = 0\,, \tag{1.11}$$

amiből, kihasználva V sima voltát, egymásutáni differenciálással a következő egyenletek adódnak:

$$\tilde{V}_{,ij} y_j^{a^{(1)}} = 0\,, \tag{1.12}$$

$$\tilde{V}_{,ij} y_j^{a^{(2)}} = -\tilde{V}_{,ijk} y_j^{a^{(1)}} y_k^{a^{(1)}}\,, \tag{1.13}$$

$$\tilde{V}_{,ij} y_j^{a^{(3)}} = -3\tilde{V}_{,ijk} y_j^{a^{(1)}} y_k^{a^{(2)}} - \tilde{V}_{,ijk\ell} y_j^{a^{(1)}} y_k^{a^{(1)}} y_\ell^{a^{(1)}}\,, \tag{1.14}$$

ahol $i = 1, 2, \ldots, n$; $j, k, \ell = 1, 2, \ldots, n+1$.

Az útkövető módszer részletes tárgyalása a témához kapcsolódó CSÉBFALVI (1993a) dolgozatban található.

A kritikus pont jelzésére szolgáló eljárás kidolgozásánál az alábbi szempontokat tartottuk figyelemre méltónak:

- Lehetőleg ne okozzon jelentős többletmunkát az alapeljáráshoz képest, vagyis optimális esetben azokon az információkon alapuljon, melyeket az útkövető módszer eleve szolgáltat, vagy amelyek viszonylag csekély többletráfordítással az útkövető módszerből adódó információk alapján meghatározhatók;
- A kritikus pont jelzése a szakasz kezdőpontjához legközelebb álló kritikus pont jelzését jelentse;
- Lehetőleg ne csak a kritikus pont jelenlétét, hanem annak típusát is jelezze; és
- Elegendő információval szolgáljon a kritikus pont pontos meghatározásához.

2. A szinguláris pontok meghatározása

A korábban megfogalmazott szempontokkal összhangban, úgy véljük, hogy egy prekoncepció mentes eljárás

- a teherkomponens ívhossz szerinti első deriváltján ($y_{n+1}^{a^{(1)}}$, illetve $y_{n+1}^{b^{(1)}}$), valamint
- az eredeti Jacobi-mátrix ($V_{,ij}$, $i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, n$) negatív sajátértékeinek számán alapulhat.

Tudott, hogy a fenti két jellemző együttes változása határpontot jelez. Elágazási pontok esetében a változás csupán a negatív sajátértékek számában következik be, a teherparaméter szerinti első derivált előjele változatlan marad.

Az útkövető módszer, a prediktor-korrektor fázis y_k^a pontból kiindulva nem csak az y_k^b pontot, hanem a két pontot összekötő egyensúlyi utat is megadja.

Vizsgáljuk meg közelebbről, hogy ez milyen előnyökkel jár a kritikus pontok felderítésére szolgáló eljárás esetében.

Az első fontos eredmény a határpontok jelzéséhez kapcsolódik és azon alapul, hogy az adott útvonal szakaszon a módszer az egyensúlyi utat paraméteres alakban adja meg:

$$y_k = y_k(s), \quad 0 \le s \le s^b, \quad y_k(0) = y_k^a, \quad y_k(s^b) = y_k^b . \tag{2.1}$$

A határpont megkeresése $y_{n+1}(s)$ extremumának meghatározását jelenti, ami a paraméteres alak következtében egyszerű algebrai eszközökkel megoldható feladat. Az eredményül adódó s^h paraméter értéke alapján a határponthoz tartozó többi koordináta az $y_k^h = y_k(s^h)$, $k = 1, 2, \ldots, n$ összefüggés alapján meghatározható.

A megközelítési módból következik, hogy az így adódó y_k^h pont a pontos megoldásnak tekinthető. Az útkövetőmódszer a határpontok környezetében, az útvonal erőteljes görbülete miatt, a lépésközt automatikusan lecsökkenti, így a határpontot tartalmazó intervallumban magának az útvonalnak, illetve a teherparaméter ívhossz szerinti első deriváltjának becslése kiemelkedően jó. Így a határpont becslése gyakorlatilag egyenértékű a határpont pontos meghatározásával.

A módszer természetesen alkalmas bármely más változóra vonatkozóan az ún. fordulópontok (turning point) pontos meghatározására. Fordulópont alatt egy olyan egyensúlyi pontot értve, amelynek környezetében az adott változó ívhossz szerinti első deriváltja előjelet vált.

Az elágazási pontok kezelése első ránézésre nehezebb feladatnak tűnik, hiszen ez annak a pontnak a meghatározását jelentené, ahol az eredeti $V_{,ij}$ Jacobi-mátrix determinánsa nullává válik, vagyis legalább egy sajátértéke előjelet vált.

A módszerben viszont nincs olyan elem, ami az elágazási pontok jelzésének problémáját a határpontokhoz hasonló egyszerűséggel megoldaná. Ha olyan megoldást keresünk, amely teljes mértékben kihasználja, hogy a kiindulási pont környezetében ismerjük az egyensúlyi út paraméteres alakját, akkor a következő egyszerű, a szakirodalomban ezideig nem ismertetett megoldás adódik.

Az eljárás kiindulási adatait a jósló-javító fázis eredményeképpen adódó pontok $V_{,ij}$ Jacobi-mátrixának sajátértékei alkotják. Ezek ismeretében, a második lépéstől kezdődően, egyre növekvő rendben, a legkisebb sajátérték ívhossz szerinti deriváltjai becsülhetővé válnak. Ez azt jelenti, hogy nincs akadálya annak, hogy a legkisebb sajátértéket, mint az s paraméter függvényét, az egyensúlyi útnak megfelelő pontossággal, az egyensúlyi út meghatározásakor követett eljárással előrejelezzük. Jelölje:

$$\eta(s) = \eta^a + \eta^{a^{(1)}} s + \frac{1}{2!}\eta^{a^{(2)}} s^2 + \cdots + \frac{1}{r!}\eta^{a^{(r)}} s^r + R_{r+1}(s) \tag{2.2}$$

a $[0, s^b]$ intervallumban a legkisebb sajátérték alakulását megadó függvényt. Ennek ismeretében az elágazási pont $\hat{y}_k^e$ közelítése egyszerűen megadható, hiszen ehhez

csupán az $\eta(s)$ függvény $\hat{s}^e$ zérushelyét kell meghatároznunk, amely, mivel egyváltozós függvényről van szó, egyszerű algebrai eszközökkel megoldható feladat. Az $\hat{s}^e$ paraméter ismeretében az $\hat{y}_k^e = y_k(\hat{s}^e)$ értékek már egyszerűen adódnak.

Megjegyezzük, hogy a fentiekben ismertetett, a legkisebb sajátértéken alapuló megoldás nem prekoncepció mentes, mivel a szóbajöhető kritikus pontok körét az egyszerű elágazások szintjén rögzíti, vagyis feltételezi, hogy az adott $[0, s^b]$ intervallumban csak egy sajátérték vált előjelet. A probléma megnyugtató megoldását úgy kapjuk, ha a legkisebb sajátérték vizsgálatát egy meghatározott d számú legkisebb sajátérték vizsgálatával helyettesítjük.

A módszerben az alágazási pontok meghatározása egy egyváltozós (közvetve $n+1$ változós) szélsőérték feladat megoldásával javítható.

Tudjuk, hogy a $[0, s^b]$ intervallumban

$$V_{,i}(y_k(s)) = V_{,i}(s) \approx 0, \quad s \in [0, s^b]. \tag{2.3}$$

Az $\hat{s}^e$ közelítő értékből kiindulva, a $[0, s^b]$ intervallumban, keressük a

$$V_{,ij}(s)\varphi_j(s) = 0 \tag{2.4}$$

egyenletrendszernek az alábbi

$$\|\varphi_j(s)\| = 1 \tag{2.5}$$

normalizálási feltételnek megfelelő s^e megoldását.

A (2.4) egyenletrendszer az eredeti $V_{,ij}$ Jacobi-mátrix φ_j nullvektorának meghatározását jelenti. Az egyensúlyi út paraméteres alakja miatt $V_{,ij}(s)$ a $[0, s^b]$ intervallum minden pontjában ismert. A (2.5) normalizálási feltétel a $\varphi_j(s) = 0$ triviális megoldását zárja ki. A $\varphi_j(s)$ nullvektor a $V_{,ij}$ Jacobi-mátrix zéró sajátértékéhez tartozó sajátvektor. Az egyszerű elágazás feltételezése miatt, a zéró sajátérték egyszeres sajátérték.

Tudjuk (pl. NOBLE (1969)), hogy mivel a $V_{,ij}$ Jacobi-mátrix szimmetrikus, ezért a $[0, s^b]$ intervallum minden s pontjában a Rayleigh-hányadosra vonatkozó feltételes minimalizálási feladat:

$$\varphi_i(s)V_{,ij}(s)\varphi_j(s) \implies \text{Minimum!}, \tag{2.6}$$

$$\varphi_i(s)\varphi_i(s) = 1 \tag{2.7}$$

$\varphi_i(s)$ megoldásához tartozó $f(s) = \varphi_i(s)V_{,ij}(s)\varphi_j(s)$ célfüggvényérték az adott s helyen a $V_{,ij}(s)$ Jacobi-mátrix minimális sajátértékét adja meg.

Ennek megfelelően az eredeti (2.6)–(2.7) feladat a következő alakban adható meg:

$$f(s) = 0, \quad s \in [0, s^b], \tag{2.8}$$

amely az s paraméterre egy egyváltozós, implicite egy $n+1$ változós programozási feladat, amelynek s^e megoldása az $y_k^e = y_k(s^e)$ elágazási ponthoz tartozó ívhossz paramétert adja meg.

Meg kell jegyeznem, hogy a (2.8) feladat esetében $f(s)$ ívhossz szerinti $\dot{f}(s)$ deriváltja az alábbi kifejezés alapján számítható:

$$\dot{f}(s) = \varphi_i(s)\dot{V}_{,ij}(s)\varphi_j(s)\,, \tag{2.9}$$

ahol $\dot{V}_{,ij}(s)$ a $V_{,ij}$ Jacobi-mátrix ívhossz szerinti deriváltját jelöli. Az egyváltozós explicit feladat Newton módszerrel, a Rayleigh-hányadosra vonatkozó $n+1$ változós implicit feladat például a konjugált gradiens módszerrel oldható meg (PAPADRAKAKIS (1984)).

3. Az elágazási pontokra vonatkozó tételek

Az útkövető módszer alapját képező (1.12), (1.13), (1.14) egyenletek megoldása nem jelent problémát abban az esetben, amikor a Jacobi-mátrix invertálható, hiszen ekkor az irányítástól eltekintve, egységnyi normájú megoldást választva, az (1.12) egyenlet megoldása egyértelmű. A magasabb rendű deriváltak meghatározására szolgáló (1.13), (1.14) egyenletek az általánosított megoldást választva ugyancsak egyértelművé tehetők.

Az előzőekhez hasonlóan, ugyancsak problémamentes az az eset, amikor az eredeti Jacobi-mátrix rangja $n-1$, de $\tilde{V}_{,in+1}$, azaz a teherparaméter szerinti deriváltaknak megfelelő oszlop nincs benne az eredti $V_{,ij}$ Jacobi-mátrix oszlopvektorainak terében, hiszen ekkor a kibővített $\tilde{V}_{,ij}$ Jacobi-mátrix rangja változatlanul n. Így nulltere változatlanul egydimenziós altér. Ezt az esetet vizsgáltuk a határpontok meghatározásakor.

A módszer alapgondolatából következik, hogy ha lokális paraméterként a differenciálegyenlet megoldásakor használt „semleges" ívhossz paramétert használjuk, akkor a határpontok kezelésének kérdése kizárólag a határpontok helyének pontos meghatározására redukálódik. Magában az útkövetés folyamatában a határpontok megjelenése módszertanilag nem jelent problémát. A semleges paraméter választás miatt a módszerben nem jelentkeznek azok a nehézségek, amelyek a lokálisan legjobb változó szerinti paraméterezés esetében adódnának. A hagymányos útkövető módszerekbe ugyanis egy külön eljárást kell beépíteni, amely „a második legjobb" változót választja paraméterként, ha az adott intervallumban, az előző intervallum alapján becsült legjobb változó iránya extremális tulajdonságokat mutat (az irány előjele megváltozik).

Elágazási pontok esetén az (1.12) egyenlet elveszti egyértelműségét, mivel a nulltér dimenziója megnövekszik. Így a lehetséges egyensúlyi irányok meghatározásának kérdése sokkal nehezebb feladattá válik.

Megjegyezzük, hogy a határpontok kezelésének problémamentessége természetesen csak az egyszerű határpontok esetében áll fenn. A többszörös határpontok

esete viszont, módszertanilag nem különbözik a lehetséges elágazási irányok vizsgálatától, így ezt részleteiben külön nem vizsgáljuk.

Tudjuk, hogy az elágazási pontokban az eredeti Jacobi-mátrix szinguláris válik, valamint a $\tilde{V}_{,in+1}$ oszlopvektor „belép" az eredeti Jacobi-mátrix oszlopvektorainak terébe.

Tegyük fel, hogy ismerjük az egyensúlyi út egy y_k^e elágazási pontját, amelyre a következők igazak:

$$V_{,i}|^e = 0\,, \tag{3.1}$$

azaz kielégíti az egyensúlyi egyenletet, a $V_{,ij}|^e$ mátrix nullterének dimenziója:

$$\begin{aligned} &\dim\left(\mathrm{null}\left(V_{,ij}|^e\right)\right) = d \\ &1 \leq d \leq n\,, \end{aligned} \tag{3.2}$$

a $V_{,ij}|^e$ mátrix nulltere a $\varphi_k^1, \varphi_k^2, \ldots, \varphi_k^d$ bázisvektorok által kifeszített altér:

$$\mathrm{null}\left(V_{,ij}|^e\right) = \mathrm{altér}(\varphi_k^1, \varphi_k^2, \ldots, \varphi_k^d)\,, \tag{3.3}$$

a $V_{,ij}|^e$ mátrix képtere:

$$\mathrm{kép}\left(V_{,ij}|^e\right) = \{\psi_k \mid \psi_k \varphi_k^m = 0,\ m = 1, 2, \ldots, d\} \tag{3.4}$$

és a teherparaméter szerinti deriváltak oszlopvektorára fennáll, hogy

$$\tilde{V}_{,in+1} \in \mathrm{kép}\left(V_{,ij}|^e\right)\,. \tag{3.5}$$

A fenti összefüggésekben szereplő dim(..) szimbólum az argumentumban szereplő vektortér dimenzióját, az altér(..) szimbólum az argumentumban szereplő bázisvektorok által kifeszített alteret jelöli.

Mátrixok baloldali, jobboldali nullvektorterére a továbbiakban bnull(..), jnull(..) szimbólumokkal hivatkozom. Egy mátrix baloldali (jobboldali) nullvektorterét azok a vektorok alkotják, amelyekkel a mátrixot balról (jobbról) szorozva a zéróvektort kapjuk. Ha a mátrix baloldali nullvektortere megegyezik a jobboldali nullvektortérrel, akkor a null(..) jelölést használom. A fenti (3.3) összefüggésben a $V_{,ij}|^e$ mátrix szimmetrikus. Tudjuk, hogy szimmetrikus mátrixok baloldali és jobboldali nullvektortere azonos, így a null(..) szimbólum használható.

A (3.4), (3.5) összefüggésben szereplő kép(..) szimbólum az argumentumban szereplő mátrix képterét, vagyis oszlopvektorainak lineáris kombinációit jelöli.

1. Megjegyzés. A (3.5) feltevés zárja ki a határpontok esetét.

1. LEMMA. *A fenti feltételek mellett*

$$\dim\left(\text{jnull}\left(\tilde{V}_{,ij}|^e\right)\right) = d+1\,, \tag{3.6}$$

$$\text{jnull}\left(\tilde{V}_{,ij}|^e\right) = \text{altér}(\Phi_k^1, \Phi_k^2, \ldots, \Phi_k^d, \Phi_k^{d+1})\,, \tag{3.7}$$

ahol:

$$\begin{gathered}\Phi_k^m = \varphi_k^m\,,\\ \Phi_{n+1}^m = 0\,,\\ k = 1,2,\ldots,n;\quad m = 1,2,\ldots,d;\end{gathered} \tag{3.8}$$

valamint:

$$\begin{gathered}\Phi_k^{d+1} = \eta_k\,,\\ \Phi_{n+1}^{d+1} = \gamma\,,\\ k = 1,2,\ldots,n;\end{gathered} \tag{3.9}$$

ahol:

$$\begin{gathered}\eta_k \in \text{kép}\,(V_{,ij}|^e)\\ \gamma \neq 0\\ \tilde{V}_{,ij}|^e \Phi_j^{d+1} = 0\,.\end{gathered} \tag{3.10}$$

Bizonyítás. Könnyen belátható, hogy

$$\tilde{V}_{,ij}|^e \Phi_j^m = 0, \quad m = 1,2,\ldots,d; \tag{3.11}$$

valamint

$$\Phi_j^m \Phi_j^{d+1} = 0, \quad m = 1,2,\ldots,d; \tag{3.12}$$

mivel $\Phi_k^m = \varphi_k^m$, $\Phi_{n+1}^m = 0$, $k = 1,2,\ldots,n$; $m = 1,2,\ldots,d$, és $V_{,ik}\varphi_k^m = 0$, $\eta_k \in \text{kép}\,(V_{,ij}|^e)$.

Az az állítás pedig, hogy

$$\Phi_j^{d+1} \in \text{jnull}\left(\tilde{V}_{,ij}|^e\right) \tag{3.13}$$

Φ_j^{d+1} definíciójából azonnal következik.

2. LEMMA. *A fenti feltételek mellett a* $\{\varphi_k^1, \varphi_k^2, \ldots, \varphi_k^d\}$ *vektorok a* $\tilde{V}_{,ij}|^e$ *kibővített Jacobi-mátrix baloldali nullvektorterének bázisát alkotják, azaz*

$$\text{bnull}\,(V_{,ij}|^e) = \text{altér}(\varphi_k^1, \varphi_k^2, \ldots, \varphi_k^d)\,. \tag{3.14}$$

Bizonyítás. Az állítás a (3.5) feltételből egyenesen következik.

2. Megjegyzés. A $d = 1$ esetben egyszerű elágazásról, a $d > 1$ esetben pedig többszörös elágazásról beszélünk.

3. Megjegyzés. Az eredeti és a kibővített Jacobi-mátrix nullterét kifeszítő vektorok közötti kapcsolat teszi lehetővé, hogy az elágazási pontok jelenlétének vizsgálatát az eredeti Jacobi-mátrix legkisebb sajátértékeinek alakulásához kössük. Ennek előnyei alapvetően abban jelentkeznek, hogy mivel szimmetrikus mátrixról van szó, a legkisebb sajátértékek meghatározására felhasználható módszerek köre kiszélesedik.

4. Megjegyzés. Az *1. Lemma* megfordítása is igaz, vagyis az eredeti és a kibővített Jacobi-mátrix közötti „közlekedés" lehetősége mindkét irányban fennáll.

Most nézzük meg részletesebben a $d = 1$ esetet. Tudjuk, hogy ekkor y_k^e elegendően kicsiny környezetében az egyensúlyi utak két folytonos görbét írnak le, amelyek csak az adott kritikus pontban metszik egymást. Egyszerű elágazások esetén a nulltér dimenziója

$$\dim\left(\text{jnull}\left(\tilde{V}_{,ij}|^e\right)\right) = 2\,, \tag{3.15}$$

a $\{\Phi_k^1, \Phi_k^2\}$ nullvektorok egy síkot feszítenek ki, azaz a két egyensúlyi irány meghatározásához az (1.12) egyenlet nem elegendő. A probléma megoldása a második deriváltakra vonatkozó (1.13) egyenletből adódik, figyelembe véve azt a tényt, hogy $\varphi_k^1 \in \text{bnull}\left(\tilde{V}_{,ij}|^e\right)$.

Az (1.13) egyenlet mindkét oldalát φ_i^1-gyel megszorozva, az alábbi egyenlet adódik:

$$\varphi_i^1 \tilde{V}_{,ijk}|^e y_j^{e(1)} y_k^{e(1)} = 0\,, \tag{3.16}$$

mint kompatibilitási feltétel.

Keressük az egyenlet megoldását az alábbi alakban

$$y_k^{e(1)} = \xi^1 \Phi_k^1 + \xi^2 \Phi_k^2\,, \tag{3.17}$$

olymódon, hogy a ξ^1, ξ^2 valós együtthatókra teljesüljön a

$$(\xi^1)^2 + (\xi^2)^2 = 1 \tag{3.18}$$

összefüggés. Az együtthatókra vonatkozó (3.18) feltétel miatt az egyenlet megoldásai normált megoldások lesznek. A (3.17) egyenlet a (3.18) feltétel figyelembevétele mellett az alábbi formában írható:

$$\varphi_i^1 \tilde{V}_{,ijk}|^e(\xi^1\Phi_j^1 + \xi^2\Phi_j^2)(\xi^1\Phi_k^1 + \xi^2\Phi_k^2) = 0\,. \tag{3.19}$$

Legyen

$$A = \varphi_i^1 \tilde{V}_{,ijk}|^e \Phi_j^1 \Phi_k^1 \tag{3.20}$$

$$B = \varphi_i^1 \tilde{V}_{,ijk}|^e \Phi_j^1 \Phi_k^2 + \varphi_i^1 \tilde{V}_{,ijk}|^e \Phi_j^2 \Phi_k^1 \tag{3.21}$$

$$C = \varphi_i^1 \tilde{V}_{,ijk}\,|^e\, \Phi_j^2 \Phi_k^2\,, \tag{3.22}$$

akkor ξ^1, ξ^2 valós együtthatókra a következő egyenletrendszert kapjuk:

$$\begin{aligned} A(\xi^1)^2 + B\xi^1\xi^2 + C(\xi^2)^2 &= 0 \\ (\xi^1)^2 + (\xi^2)^2 &= 1\,, \end{aligned} \tag{3.23}$$

amiből $\{\xi^1, \xi^2\}$ együtthatókra két megoldás adódik. Az egyenletrendszer elemi algebrai eszközökkel megoldható. Általános esetben ($A \neq 0$; $B \neq 0$; $C \neq 0$) egy egységsugarú kör és egy másodfokú görbe metszéspontjainak meghatározását jelenti.

A vizsgálat jelenlegi szakaszában az elágazási pontok típusára vonatkozóan csak azt tudjuk eldönteni, hogy az adott elágazás szimmetrikus, vagy aszimmetrikus elágazásnak tekinthető. Szimmetrikus elágazás alatt azt értjük, amikor a két egymást metsző görbe irányát meghatározó vektorok közül az egyik teherkomponense zérus lesz. A (3.23) egyenletrendszer szerkezetéből következik, hogy az elágazási pont szimmetrikus–aszimmetrikus volta az A együttható függvénye. Az $A = 0$ esetben az adott egyszerű elágazási pont szimmetrikus.

5. Megjegyzés. Figyelembe véve, hogy az eredeti $V_{,ij}|^e$ Jacobi-mátrix szingularitásából fakadó Φ_k^1 nullvektorban $\Phi_{n+1}^1 = 0$, az A együttható

$$V_{,ijk}|^e \varphi_i^1 \varphi_j^1 \varphi_k^1$$

alakban is felírható. Az így adódó alak megegyezik FLORES–GODOY (1992) klasszifikációs ismérvével. Különbségek, az eltérő megközelítési módon túlmenően abból fakadnak, hogy a Flores–Godoy féle összefüggések alapvetően a másodlagos útra vonatkoznak, így bizonyos esetekben nem alkalmasak az elsődleges útvonal meghatározására.

Az elágazási ponton átmenő két egyensúlyi úthoz tartozó magasabb rendű deriváltakat az (1.13), (1.14), illetve a sorozat folytatásából adódó egyenletrendszerek megoldása szolgáltatja, hangsúlyozva ismételten, hogy a módszerben az egyenletrendszerek baloldala minden derivált rendszám esetében azonos, így a megoldást

mindig ugyanaz a mátrix, a kibővített Jacobi-mátrix $(\tilde{V}_{,ij})^{\dagger}_{k\ell}$, $(k = 1, 2, \ldots, n+1;$ $m = 1, 2, \ldots, n)$ pszeudóinverze szolgáltatja. Természetesen, megoldás alatt általánosított megoldást értünk.

A megközelítés másik lehetséges módját az jelentené, amikor az egyre magasabb rendű deriváltak irányába haladva, az egyes lépésekben hiányzó egyenletek bevonásával pótolnánk. Ezt a megoldási módot vitathatónak érezzük abból a szempontból, hogy így minden egyes lépésben más és más együtthatómátrixszal rendelkező egyenletrendszert kellene megoldani.

Ezek után röviden foglalkoznunk kell a $d > 1$ esettel, vagyis a többszörös elágazások kérdésével. Az előzőekhez hasonlóan, a megoldást

$$y_k^{e(1)} = \xi^1 \Phi_k^1 + \xi^2 \Phi_k^2 + \cdots + \xi^d \Phi_k^d + \xi^{d+1} \Phi_k^{d+1} \tag{3.24}$$

alakban keressük, ahol $\Phi_k^1, \Phi_k^2, \ldots, \Phi_k^d$ az eredeti $V_{,ij}|^e$ Jacobi-mátrix szingularitásából adódó nullvektorokat jelöli, Φ_k^{d+1} a Jacobi-mátrix kibővítéséből, illetve a teherkomponenshez tartozó $\tilde{V}_{,in+1}|^e$ deriváltakra vonatkozó (3.5) megkötésből adódó nullvektor.

A megoldás, a $d = 1$ esethez hasonlóan, a másodrendű deriváltakra vonatkozó (1.13) egyenlet kompatibilitási feltétellé történő átírásából, azaz a másodrendű deriváltak kiejtéséből adódik:

$$\varphi_i^m \tilde{V}_{,ijk}|^e \left(\sum_{p=1}^{d+1} \xi^p \Phi_j^p \right) \left(\sum_{p=1}^{d+1} \xi^p \Phi_k^p \right) = 0, \quad m = 1, 2, \ldots, d. \tag{3.25}$$

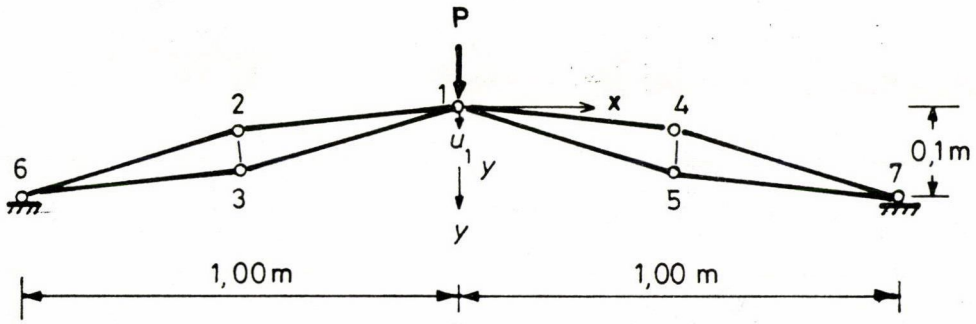
A d egyenletből álló nemlineáris egyenletrendszer $d+1$ változót tartalmaz, amelyet a

$$\sum_{p=1}^{d+1} (\xi^p)^2 = 1 \tag{3.26}$$

normálási feltétellel egészíthetünk ki. A felírási módból látszik, hogy a $d > 1$ esetekben az elágazási pontból kiinduló egyensúlyi irányok meghatározása d növelésével egyre nehezebbé válik. Figyelembe véve, hogy nemlineáris egyenletrendszerről van szó, a nullától különböző megoldások száma a $d+1$ értéket meghaladja. A különböző megoldások számára vonatkozóan BEZOUT (lásd NAAS és SCHMID (1967)) adott felső korlátot.

Megjegyezzük, hogy az elágazási pontok meghatározásával kapcsolatos nehézségek jelentős mértékben csökkenthetők, ha az elágazási útvonal „örökli" az elsődleges útvonal szimmetrikus sajátosságait, mivel ekkor a (3.26) nemlineáris egyenletrendszer megoldása kikerülhető. Így nem véletlen, hogy a témakör szakirodalma rendkívül széleskörű és sokrétű (pl. WERNER–SPENCE (1984), WEINITSCHKE (1985)).

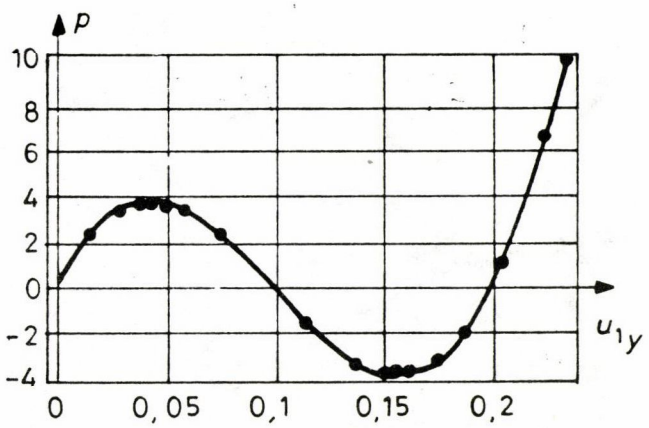
Módszerünkben a vizsgált szerkezet szimmetriájából fakadó egyszerűsítési lehetőségek egy az egyben fennállnak, így részletes ismertetésüktől eltekinthetünk.



3.1 ábra

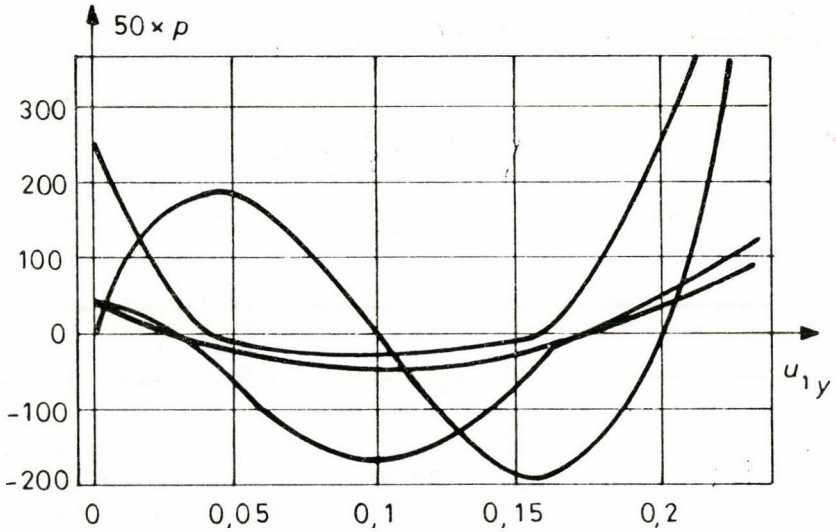
Az ismertetett útkövető módszer szinguláris pontok meghatározására történő alkalmazását egy egyszerű példán (lásd 3.1 ábra) keresztül szemléltetjük.

A 3.2 ábra az állapotváltozási görbét szemlélteti, amely a becsült ívszakaszok sorozata. Az ábra az ívszakaszokat, illetve a becsült pontokat tartalmazza, így jól látszik, hogy a határpontok (forduló pontok) környezetében a pontok besűrűsödnek, a becsült ívszakaszok megrövidülnek, amely egyben azt is jelenti, hogy az interpoláción alapuló határpont meghatározás gyakorlatilag egzakt eredményeket szolgáltat.

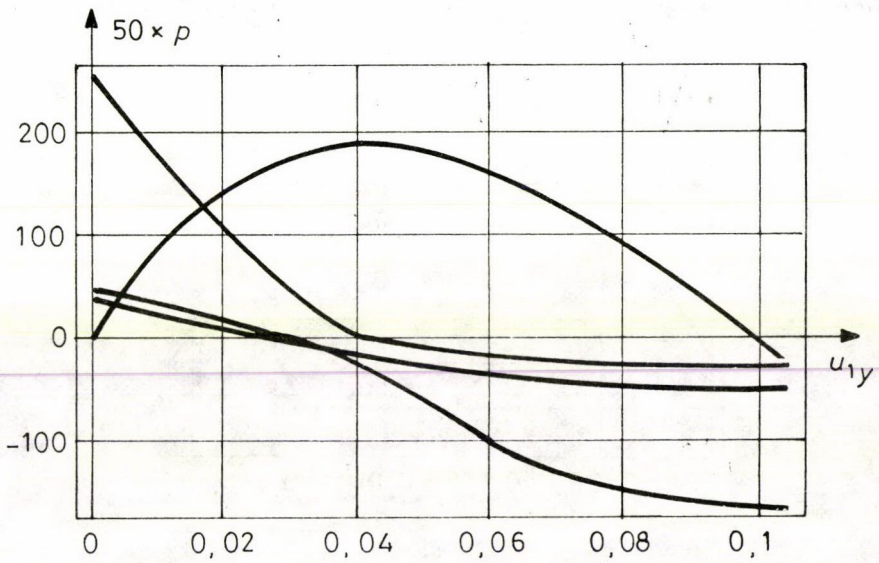


3.2 ábra

A 3.3 ábra illetve 3.4 ábra az állapotváltozási görbét, valamint az eredeti Jacobi-mátrix három legkisebb sajátértékének alakulását szamlélteti. A 3.3 ábra magyarázatot ad WRIGGERS (1988) azon észlelésével kapcsolatban, hogy a feladatban az elágazási pontok környezetében nagyfokú numerikus instabilitás tapasztalható, ami a felszínen a direkt módszer lassú konvergenciájaként jelentkezik. Még a kinagyított



3.3 ábra



3.4 ábra

3.4 ábrán sem különülnek el a két sajátérték függvény zérushelyének megfelelő pontok. A numerikus nehézségeket fokozza, hogy az elágazási pontok környezetében határpont is található, ami a Jacobi-mátrixot tovább „színezi". A feladat érdekes-

ségét az adja, hogy nem lehet egyértelműen eldönteni, hogy két egymáshoz igen közelálló egyszeres, vagy egy többszörös elágazási pontról van-e szó.

IRODALOM

[1] ABBOTT, J. P., „An efficient algorithm for the determination of certain bifurcation points", *J. Comput. Appl. Math.* **4** (1978), 19–27.

[2] ABBOTT, J. P., „Computing solution arcs of nonlinear equation with a parameter", *The Computer Journal* **23** (1980), 85–89.

[3] CSÉBFALVI, A. and CSÉBFALVI, GY., „Geometrically Non-Linear Analysis of Space Trusses Using Sparse Quasi-Newton Algorithms", *Proceedings, 11th international Conference on Mathematical Programming*, Mátrafüred, Hungary, March 21–26 , 1992, (1992), 2–3.

[4] CSÉBFALVI, A. and VÁSÁRHELYI, A., „Some computational aspects of nonlinear space trusses", *Acta Techn. Acad. Sci. Hung.* (1993), (megjelenés alatt).

[5] CSÉBFALVI, A. and CSÉBFALVI, GY., „Post-buckling analysis of frames by a hybrid path-following method", in Generalized Convexity (S. Komlósi, T. Rapcsák, S. Schaible, eds.) (Springer Verlag, 1993), 311–321.

[6] CSÉBFALVI, A., „Nemlineáris útkövető módszer tartószerkezetek stabilitásvizsgálatára I. Reguláris pontok", *Alkalmazott Matematikai lapok* (1993a), (megjelenés alatt).

[7] CSÉBFALVI, A., „Szakaszonként folytonosan differenciálható energia függvénnyel jellemezhető tartószerkezetek stabilitásvizsgálata", *Alkalmazott Matematikai lapok* (1993b), (megjelenés alatt).

[8] CRANDALL, M. G. and RABINOWITZ, P. H., „Bifurcation from simple eigenvalues", *J. Functional Analysis* **8** (1971), 321–340.

[9] ERIKSSON, A., „On some path-related measures for nonlinear structural F. E. problems", *I. J. Num. Meth. Eng.* **26** (1988), 1791–1803.

[10] ERIKSOON, A., „On linear constraints for Newton-Raphson corrections and critical point searches in structural F. E. problems", *I. J. Num. Meth. Eng.* **28** (1989), 1317–1334.

[11] ERIKSOON, A., „Derivatives of tangential stiffnes matrices for equilibrium path descriptions", *I. J. Num. Meth. Eng.* **32** (1991), 1093–1113.

[12] FLORES, F. G. and GODOY, L. A., „Elastic postbuckling analysis via finite element and perturbation techniques", Part 1: Formulation, *I. J. Num. Meth. Eng.* **33** (1992), 1775–1794.

[13] HALL, G. and WATT, J. M., *Modern Numerical Methods for Ordinary Differential Equations* (Oxford, 1978).

[14] HOUWEN P. J., *Construction of integration formulas for initial value problems* (North-Holland Publisching Company, Amsterdam, 1977).

[15] KAMAT, M. P., WATSON, L. T. and VENKAYYA, V. B., „A quasi-Newton versus a homotopy method for stuctural analysis", *Comput. & Struct.* **17** (1983), 579–585.

[16] KOITER, W. T., *Over de Stabiliteit van het elastisch Evenwicht*, Dissertation (Delft Technical University, H. J. Paris, Amsterdam, Holland, 1945).

[17] KUBICEK, M., „Dependence of solution of nonlinear systems on a parameter, Algorithm 502", *CACM-Toms* **2** (1976), 98–107.

[18] LI, T. Y. and YORK, J. A., „A Simple Reliable Numerical Algorithm for Following Homotopy Paths", in Robinson, S. M.: Analysis and Computation of Fixed point (Academic Press, New York, 1980).

[19] NAAS, J. and SCHMID, H. L., *Mathematisches Wörterbuch*, Band 1, 3 Auflage (Akademie Verlag, Berlin, Stuttgart, 1967), 201.

[20] NOBLE, B., *Applied linear algebra* (Prentice-Hall, New Jersey, 1969).

[21] PAPADRAKAKIS, M., „Solution of the partial eigenproblem by iterative methods", *I. J. Num. Meth. Eng.* **20** (1984), 2283–2301.

[22] SHAMPE, L. F. and GORDON, M. K., *Computer Solution of Ordinary Differential Equations, The Initial Value Problem* (Freeman & Company, San Francisco, 1975).

[23] SZÉP J., *Analízis* (Közgazdasági és Jogi Könyvkiadó, Budapest, 1972).
[24] THOMPSON, J. M. T. and HUNT, G. W., *A general Theory of Elastic Stability* (Wiley, New York, 1973).
[25] WATSON, L. T., KAMAT, M. P. and REASER, M. H., „A robust hybrid algorithm for computing multiple equilibrium solutions", *Eng. Comput.* **2** (1985, March), 30–34.
[26] WATSON, L. T., „Numerical linear algebra aspects of globally convergent homotopy methods", *SIAM Rev.* **28** (1986), 529–545.
[27] WEINITSCHKE, H. J., „On the calculation of limit and bifurcation points in stability problems of elastic shells", *Int. J. Solids Structures* **21** (1985), 79–95.
[28] WERNER, B. and SPENCE, A., „The computation of symmetry-breaking bifurcation points", *SIAM J. Num. Anal.* **21** (1984), 388–399.
[29] WRIGGERS, P., WAGNER, W. and MIEHE, C., „A quadratically convergent procedure for the calculation of stability points in finite element analysis", *Computer Meth. Appl. Mech. and Eng.* **70** (1988), 329–347.
[30] WRIGGERS, P. and SIMO, J. C., „A general procedure for the direct computation of turning and bifurcation points", *I. J. Num. Meth. Eng.* **30** (1990), 155–176.

(Beérkezett: 1993. október 28.)

CSÉBFALVI ANIKÓ
POLLACK MIHÁLY MŰSZAKI FŐISKOLA
7625 PÉCS, BOSZORKÁNY ÚT 2.

NONLINEAR PATH-FOLLOWING METHOD FOR STABILITY OF STRUCTURES II. BIFURCATION AND LIMIT POINTS

A. CSÉBFALVI

In this paper we present a path-following method for detect singular points (bifurcation and limit points) of the equilibrium curve. This path-following method is capable of computing segments of the equilibrium path. This fact is the base of investigation of the singular points.

Alkalmazott Matematikai Lapok **17** (1993) 87–99

NÉHÁNY TÉRCSOPORT OPTIMÁLIS GÖMBKITÖLTÉSE*

SZIRMAI JENŐ

Budapest

A dolgozat négy nevezetes kristálycsoport szerinti egyszeresen tranzitív gömbkitöltéseket vizsgál, vagyis olyan gömbrendszereket, amelyeknek a szimmetriacsoportja az adott tércsoport és a gömbkitöltés bármely két gömbjét az előbbi csoportnak pontosan egy eleme viszi egymásba.

Az általam vizsgált tércsoportok ($P\bar{4}3m$; $Fd\bar{3}m$; $Pn\bar{3}m$; $I\bar{4}3m$) az $F\bar{4}3m$ jelű tükrözéscsoport bővítésével származtathatók. A dolgozat megadja az adott tércsoporthoz tartozó optimális gömbkitöltést és annak sűrűségét

A feladat minden esetben visszavezethető a szfenoid nevezetű egybevágó lapokkal rendelkező tetraéder 2 illetve 4 gömbbel történő legsűrűbb kitöltésére. A szfenoid lapsíkjaira vonatkozó síktükrözések generálják az $F\bar{4}3m$ tércsoportot, amelyeknek a szfenoid alaptartománya.

Az eredmények közül kiemelem az $Fd\bar{3}m$ jelű tércsoporhoz tartozó optimális elrendezést, amely igen ritka 0,199 optimális sűrűséget szolgáltat.

1. Bevezetés

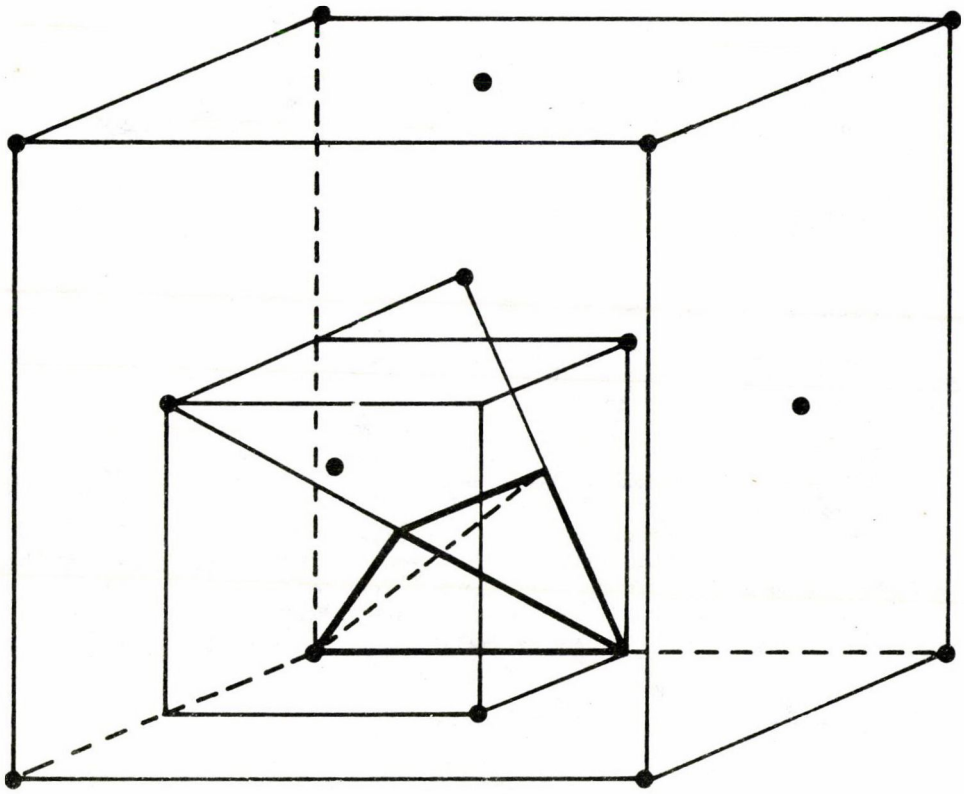
A századfordulón a fizika kristályvizsgálataival párhuzamosan előtérbe került a kristályok geometriájának vizsgálata is. Érdekes lehet az egyes kristályok felépítésével kapcsolatban az adott tércsoporthoz tartozó optimális gömbkitöltések megadása és ezek sűrűségének meghatározása is.

Ezt a problémát U. SINOGOWITZ az 1940-es évbekben írt [5] cikkében kezdeményezte a legsűrűbb rácsszerű gömbkitöltés analógiájára. Mint ismeretes a legsűrűbb rácsszerű gömbkitöltés az $Fm\bar{3}m$ jelű tércsoporthoz tartozik, amely a lapcentrált kockarács szimmetriacsoportjának felel meg. Ugyanilyen sűrűséggel kaphatunk még szabályos és nem szabályos gömbrendszereket is.

A dolgozat adott tércsoporthoz tartozó egyszeresen tranzitív gömbkitöltéseket vizsgál, vagyis olyan gömbrendszereket, amelyeknek a szimmetriacsoportja az adott tércsoport és a gömbkitöltés bármely két gömbjét az előbbi csoportnak pontosan egy eleme viszi egymásba.

Az általunk vizsgált tércsoportok az $F\bar{4}3m$ jelű tükrözéscsoport bővítésével származtathatók. Az $F\bar{4}3m$ tércsoport lapcentrált kockarácsát és speciális tetraéder (szfenoid) alaptartományát az 1. ábra mutatja. Ennek a tetraédernek két szemközti

* Készült az OTKA 1615 (1991) támogatásával.



1. ábra

élénél 90 fokos lapszög, a többi élnél 60 fokos lapszög van. A lapokra vonatkozó tükrözések generálják az $F\bar{4}3m$ csoportot. Dolgozatunk megadja a bővítéssel származtatható néhány tércsoport optimális gömbkitöltését, ennek sűrűségét.

a. A $2 \circ F\bar{4}3m = P\bar{4}3m$ jelű tércsoport 180 fokos tengely körüli forgatással, röviden 2-forgatással, való bővítéssel származtatható. A 2-forgás tengelye a 90 fokos lapszögű szemközti élek felezőpontjait köti össze (2. ábra).

Az optimális gömbkitöltés sűrűsége $\cong 0,224$.

b. A $2 \circ F\bar{4}3m = Fd\bar{3}m$ tércsoport a 3. ábrán látható módon származtatható. A 2-forgás tengelye 60 fokos lapszögű éleket köt össze.

Az optimális gömbkitöltés sűrűsége $\cong 0,199$.

c. A $222 \circ F\bar{4}3m = Pn\bar{3}m$ jelű técsoport a 4. ábrán látható módon származtatható a 3 darab 2-forgás tengelyével.

Az optimális gömbkitöltés sűrűsége $\cong 0,398$.

d. A $\bar{4} \circ F\bar{4}3m = I\bar{4}3m$ jelű tércsoport az 5. ábrán látható módon származtatható. 90 fokos forgatás és középpontos tükrözés kompozíciója a bővítő $\bar{4}$ jelű

forgástükrözés.

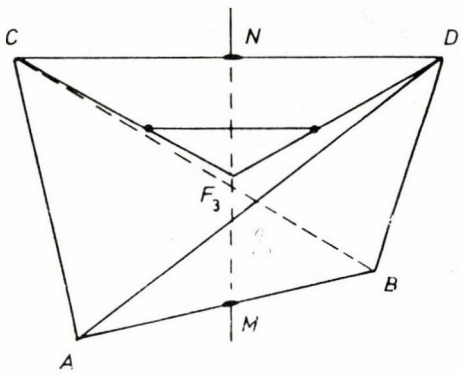
Az optimális gömbkitöltés sűrűsége $\cong 0,398$.

Az utóbbi két optimális elrendezés megegyezik egymással.

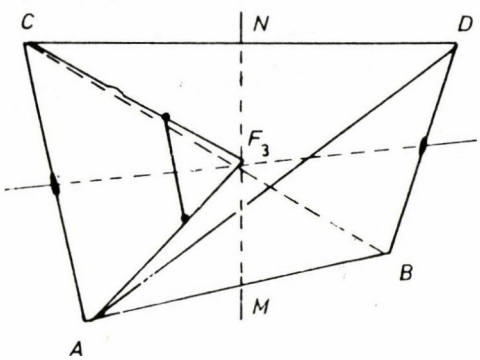
A teljesség kedvéért említjük, hogy az eredeti $F\bar{4}3m$ csoport szerinti gömbrendszerek közül a maximális sűrűségű kitöltést az 1. ábra tetraéderébe beírt gömb szolgáltatja. Az optimális sűrűség $\cong 0,433$.

Az optimális gömbkitöltések meghatározása lehetséges az analízis és a lineáris algebra apparátusának felhasználásával, azonban ez adott tércsoport esetén igen nehézkessé válik. A probléma megoldására további lehetőség az optimális sűrűségek számítógépes programmal történő megkeresése. Ennek hátránya, hogy csak közelítő megoldásokat eredményez ([2]).

Az általunk alkalmazott eljárás a szélsőértékszámítás elemi geometriai eszközeit használja, a megsejtett legkedvezőbb gömbkitöltésekről látjuk be azok optimális voltát. Hasonló kérdésfeltevés más tércsoportokra is aktuális és az alkalmazások számára is érdekes feladat. Módszerünk további tércsoportok optimális gömbkitöltésének meghatározására is alkalmazható (lásd még [6]).



2. ábra



3. ábra

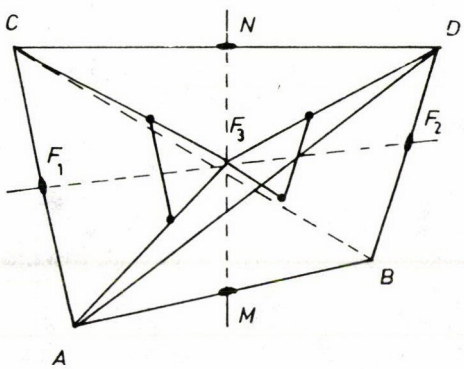
2. Alapfogalmak

Az E^3 euklideszi tér egybevágóságainak a csoportját jelölje Iso E^3.

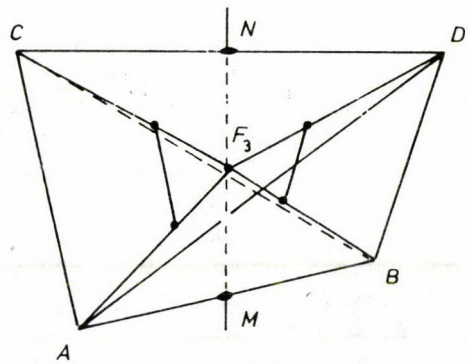
2.1. Definíció. A G transzformáció-csoportot az E^3 tér diszkrét csoportjának nevezzük, ha teljesülnek az alábbi feltételek:

a. $G \subset \text{Iso } E^3$

b. Tetszőleges $X \in E^3$ esetén az X pont pályája (orbitja):
$X^G := \{X^\alpha \in E^3 : \alpha \in G\}$ diszkrét ponthalmaz az E^3 térben (nincs torlódási pontja).



4. ábra



5. ábra

2.2 Definíció. Az F_G zárt ponthalmazt a G diszkrét csoport alaptartományának (fundamentális tartományának) nevezzük, ha teljesülnek a következő feltételek:

a. Minden $P \in E^3$ esetén létezik olyan $A \in F_G$ pont, hogy $P \in A^G$.
b. Tetszőleges A, $B \in \text{int}F_G$ belső pontokra igaz, hogy ha $B \in A^G$ akkor $A = B$.
c. $\text{int}F_G$ egyszeresen összefüggő E^3-ban.

2.3. Definíció. A G diszkrét csoportot kristálycsoportnak mondjuk, ha létezik korlátos alaptartománya.

2.4. Definíció. Egy $A \in E^3$ pontnak a G diszkrét csoporthoz tartozó G_A stabilizátorcsoportja az A-t helybenhagyó G-beli transzformációkból áll:

$$G_A := \{\alpha \in G : A^\alpha = A\}.$$

A továbbiakban, rögzített G diszkrét csoport esetén, olyan $P \in E^3$ pontokkal foglalkozunk, amelyeknek a stabilizátorcsoportja az identitásból áll, azaz $G_P = 1$. Szemléletesen fogalmazva a P pont szabadsági foka három. (Ha $G_P \neq 1$, akkor a P pont szabadsági foka értelemszerűen csökken.)

Legyen ezek után G kristálycsoport, X, $Y \in E^3$ és $\rho(X,Y)$ az E^3-térbeli távolságfüggvény.

2.5. Definíció. Az X^G pályához tartozó, X magpontú Dirichlet–Voronoj cella, röviden D–V cella:

$$D(X^G) := \{y \in E^3 : \rho(X,Y) \leq \rho(Y,X^g) \text{ bármely } g \in G \text{ esetén}\}.$$

Ha $G_x = 1$, akkor $D(X^G)$ az E^3 térben G kristálycsoporthoz tartozó alaptartomány.

2.6. Definíció. A $D(X^G)$ cellába írható X középpontú maximális gömb sugara:

$$r(X) := \min_{g \in G \setminus \{I\}} \left\{ \frac{1}{2} \rho(X, X^g) \right\}.$$

2.7. Definíció. Az X^G orbithoz tartozó gömbkitöltés sűrűsége:

$$\delta(X^G) := \frac{\frac{4}{3} r^3(x) \pi}{\mathrm{Vol}(D(X^G))}.$$

Legyen $Y, Z \in X^G$ és $h \in G$-re teljesüljön, hogy $Y^h = Z$, ekkor $(D(Y^G))^h = D(Z^G)$. Ez nyilván teljesül az egyes cellákhoz tartozó maximális sugarú gömbökre is. Az X^G pontrendszer és a kialakuló gömbrendszer $\mathrm{Sym}(X^G)$ szimmetriacsoportja mindenképpen tartalmazza G szimmetriacsoportját, de lehet gazdagabb is nála: $G \leq \mathrm{Sym}(X^G)$.

2.8. Definíció. Ha $G = \mathrm{Sym}(X^G)$, akkor az X^G pályát karakterisztikusnak mondjuk. Egyébként a pálya (orbit) nem karakterisztikus.

3. A probléma általános fölvetése

Adott G csoport esetén keressük azt az X^G orbitot, ahol az orbithoz tartozó gömbkitöltés sűrűsége a maximális.

A G csoporthoz tartozó optimális sűrűség:

$$\delta(G) := \max_{X, p(G)} (\delta(X^G))$$

(ahol $p(G)$ a tércsoport esetleg fellépő szabad (affin) paramétereit jelöli).

4. Általános észrevételek

a. Az egyes orbitokat ekvivalencia-osztályokba soroljuk $X^G \sim Y^G$, ha létezik $h \in \mathrm{Iso}\, E^3$, amelyre $(X^G)^h = Y^G$. Azok a $h \in \mathrm{Iso}\, E^3$ egybevágóságok, amelyekre minden $x \in E^3$ esetén $(X^G)^h = (X^h)^G$ a G kristálycsoport metrikus normalizátor csoportját alkotják:

$$N(G) := \{h : h^{-1} G h = G, \quad h \in \mathrm{Iso}\, E^3\}.$$

Legyen az $N(G)$ metrikus csoport alaptartománya $F(N(G))$.

b. Amikor az optimális sűrűségű gömbkitöltéshez tartozó orbitot keressük, az orbit egy elemét elég az $F(N(G)) \subseteq F_G$ alaptartományból keresni.

$$\delta(G) := \max_{\substack{X \in F(N(G)) \\ p(G)}} (\delta(X^G))$$

(A metrikus normalizátor csoport alaptartománya is függhet a G kristálycsoport paramétereitől.)

c. A b pontban leírt vizsgálat során sok esetben érdemes a normalizátor alaptartományát is alkalmasan választott résztartományokra bontani.

d. Ha valamely esetben $G_x \neq 1$, akkor az optimális gömbkitöltéshez tartozó X^G orbit egy elemét az F_G és így $F(N(G))$ alaptartomány határán kereshetjük, nulla, egy vagy kétdimenziós tartományban.

e. Az optimális gömbkitöltéshez tartozó orbit legyen X^G opt. A szimmetriacsoportja legyen $\mathrm{Sym}(X^G \text{ opt.}) \geq G$. Különösen érdekesek azok az esetek, amikor $\mathrm{Sym}(X^G \text{ opt.}) = G$, ha tehát optimális orbit karakterisztikus. Különben az optimum egy gazdagabb szimmetriacsoporthoz tartozó gömbelhelyezést eredményez. Látni fogjuk, hogy a bevezetésben szereplő 4 csoport közül a 3. ábra $Fd\bar{3}m$ csoportjánál az optimális orbit karakterisztikus, a többi csoport esetében ez nem lesz igaz.

5. Néhány tércsoport optimális gömbkitöltésének meghatározása

A most vizsgálandó tércsoportok mind egy nevezetes tetraéderhez, a szfenoid nevű tetraéderhez kapcsolódnak. A szfenoid a 6. ábrán látható módon keletkezik, csupa egybevágó lapokból áll és szemközti két egység hosszú éleinél 90°-os, a többi $\sqrt{3}$ hosszú éleinél 60°-os lapszögek vannak. Ezt a tetraédert a lapsíkjaira tükrözve és ezt ismételve egy térkitöltést kapunk, amelyhez az $F\bar{4}3m$ jelű tércsoport tartozik. További tércsoportokat kapunk, ha az előbbi tetraédert önmagába vivő másodrendű forgatásokkal bővítjük az előbbi síktükrözések által generált $F\bar{4}3m$ tércsoportot. Olyan gömbrendszereket vizsgálunk, amelyekben a gömbkitöltés bármely két elemét egy előbbi csoportnak pontosan egy eleme viszi egymásba. Az ilyen gömbkitöltést egyszeresen tranzitívnak nevezzük.

Tekintsük tehát a szfenoidot és használjuk a 6. ábra jelöléseit.

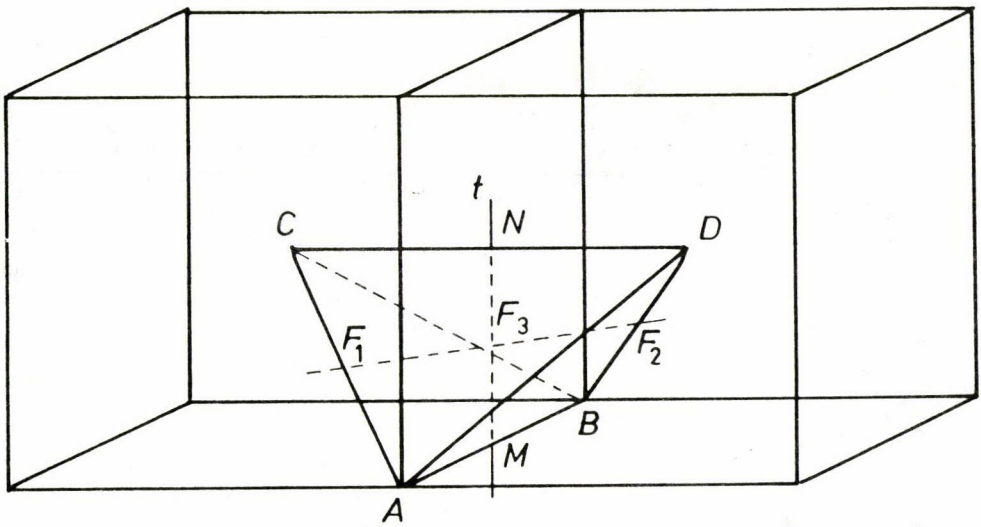
6. A $P\bar{4}3m$ tércsoport optimális gömbkitöltése

Bővítsük az $F\bar{4}3m$ tércsoportot a $t = MN$ tengely körüli másodrendű forgatással, ami a szfenoidot önmagára képezi le. Ekkor kapjuk a $2 \circ F\bar{4}3m = P\bar{4}3m$ tércsoportot (2. ábra). E tércsoport optimális gömbkitöltésének meghatározása a következő probléma megoldását igényli:

Keressük az $ABCD$ tetraéderben elhelyezkedő azon két gömböt, amelyekre teljesül:

a. nem nyúlnak egymásba
b. egymásból t körüli 180°-os forgatással származtathatók
c. sugaruk maximális.

Jelöljük g-vel az $ABCD$ tetraéder által tartalmazott tetszőleges gömböt, g'-vel ennek a t körüli 180 fokos elforgatottját.



6. ábra

Észrevételek:

i. A g' gömböt is tartalmazza az $ABCD$ tetraéder.

ii. A g gömb nem metszhet bele t-be.

iii. Az $ABCD$ tetraéder szimmetriáira hivatkozva és utalva a 4.b és 4.c pontokra, elegendő a maximális sugarú g gömb középpontját a metrikus normalizátor alaptartományában, például az AF_3CM tetraéder pontjai között keresni, leszámítva az AMC háromszöglemezt. Hiszen az ABN síkra a CDM síkra való tükrözések, továbbá a szemközti 60°-os lapszögű élek felezőpontjait összekötő tengelyek körüli 2-forgások az $ABCD$ tetraédert és az NM 2-forgás tengelyét is önmagába viszik. $N(P\bar{4}3m) = Im3m$, melynek egy alaptartománya az AF_3CM tetraéder ([4]).

A maximális sugarú g gömb középpontjának és sugarának meghatározása

Megsejtjük, hogy melyik lesz a maximális sugarú g gömb, majd egy tetszőlegesen választott g gömböt a sugarának nem csökkentésével eljuttatjuk a sejtett gömbbe.

Sejtés:

Az optimális g gömb az, amelynek középpontja az MNC háromszöglapon van és érinti az ANC és a CAM síkokat továbbá t tengelyt is.

Válasszunk egy tetszőleges g gömböt, középpontja legyen O. Keressünk olyan mozgatásokat, amelyek a g gömböt sugarának nem csökkentésével eljuttatják a sejtett optimális gömbbe!

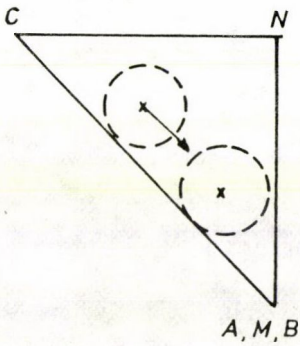
Mozgatások

Legyen a g gömb középpontja, O, az MNC háromszöglemezen.

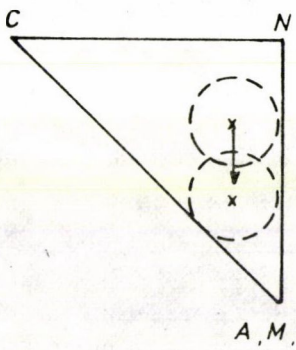
Növeljük a g gömb sugarát folytonosan az O pont körül. Ha ezen növelés során először a CM szakaszt vagyis az ABC lapot érinti, akkor g-t mintegy gurítsuk le az ABC lapon úgy, hogy a CM szakaszon maradjon az érintési pont. Addig guruljon, amíg NM-et nem érinti. Ezt a lépést megtehetjük, hiszen O-nak a CNA illetve a CDB lapoktól való távolsága növekedett, miközben az ABC laptól való távolsága változatlan maradt. Tehát eljutottunk egy olyen g_1 gömbbe, amely érinti t-t és az ABC síkot. Ezekután g_1-et az M pontból középpontosan kinagyítjuk, amíg nem érinti az ANC, illetve a CBD lapokat (7. ábra).

Ha a g gömb sugarának folytonos növelése során először t-t érinti, akkor a következő eljárást vegezzük: Mozgassuk g-t MN-nel párhuzamosan úgy, hogy az érintési pont az MN szakaszon maradva M felé közelítsen. Ezt megtehetjük, hiszen ezen mozgatásnál az O pont távolsága az ANC illetve a CBD lapoktól nő — eljutunk egy t-t és az ABC lapot érintő gömbhöz. Ezt M-ből kinagyítjuk addig, amíg az ANC és a CBD lapokat nem érinti (8. ábra).

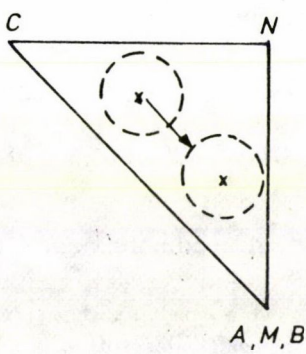
Ha a g gömb növelésekor először az ANC illetve a CBD lapokat érinti, akkor a g gömböt a CM-mel párhuzamosan mozgatjuk — O-nak az ABC síktól való távolsága nem változik az ANC illetve a CBD lapoktól való távolsága pedig nő, mert nő a g-nek a CA egyenestől való távolsága. A g gömböt addig gurítjuk le, amíg nem érinti a t tengelyt. Ezután az előző bekezdés módszerét alkalmazva eljutunk a „sejtett" gömbbe (9. ábra).



7. ábra



8. ábra



9. ábra

Így ha az O pontot a CMN háromszöglapon választottuk, akkor sejtésünk helyesnek bizonyult.

Most válasszuk az O pontot, (figyelembe véve a 6.iii. észrevételét) az AMF_3C tetraéder pontjai közül, leszámítva az AMC háromszöglapot.

Ha g a sugarának a növelése során először az ABC lapot érinti, akkor mozgassuk g-t AB-vel párhuzamosan, úgy hogy középpontja, O, rákerüljön az AF_3C

szögfelezősíkra. Ezt megtehetjük, hiszen O-nak az ABC laptól való távolsága állandó, t-től való távolsága nő és pontosan addig tart a mozgás, amíg az ANC lapot nem érinti.

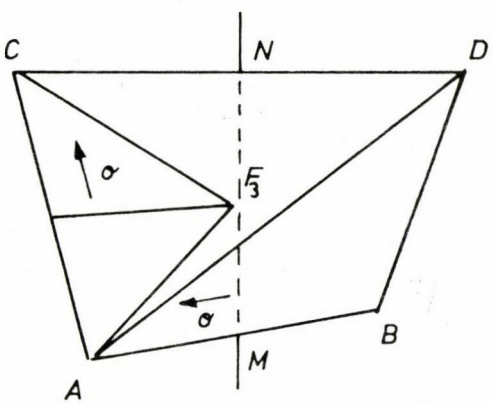
Tegyük fel, hogy g sugarának növelése során először t-t érinti. Alkalmazzunk olyan mozgatást, amely $B \to A$ irányú és addig végezzük, amíg az O pont az AF_3C szögfelezősíkra nem kerül.

Tehát tetszőleges g-hez tudunk olyan mozgatást alkalmazni, amely g sugarának nem csökkentésével a középpontját az AF_3C háromszöglemezre vitte. (Ha az O pont az AF_3C háromszöglapon van, akkor a következő pontot alkalmazzuk. 10. ábra)

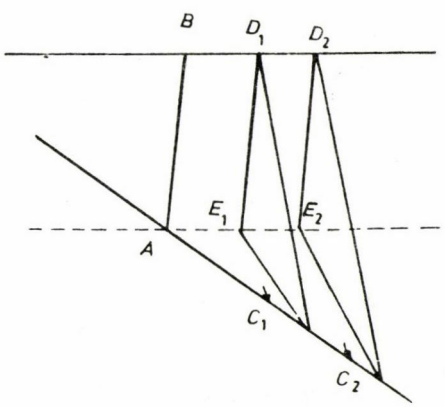
6.1. LEMMA. *Legyenek a, b kitérő egyenesek, a normáltranszverzális szakaszuk végpontjai rendre A és B. Ha egy C pont az a egyenesen az A ponttól (adott irányba) távolodik, akkor C-nek a b egyenestől való tábvolsága monoton nő (11. ábra).*

Bizonyítás. Legyen AC_2 nagyobb mint AC_1 — legyen a C_i-ből b-re állított merőleges talppontja D_i $(i = 1, 2)$. C_i-ből (b, A) síkjára bocsátott merőleges talppontja E_i $(i = 1, 2)$. $D_1C_1E_1$ és $D_2C_2E_2$ derékszögű háromszögekben $D_1E_1 = D_2E_2$ és $E_1C_1 < E_2C_2$ ezért $D_1C_1 < D_2C_2$.

Ezzel a lemmát beláttuk.



10. ábra



11. ábra

Bármelyik O gömbközéppontot bejuttattuk az AF_3C háromszöglemezre a g gömb sugarának csökkentése nélkül. Az $ABCD$ tetraéder szimmetriáiból adódóan nyilvánvaló, hogy az O pontot választhatjuk az F_1F_3C háromszöglemez pontjaiból (10. ábra).

Mozgassuk az O pontot az AC-vel párhuzamosan úgy, hogy a CMN háromszöglemezre kerüljön. A 6.1. lemma értelmében az O pont távolsága a t egyenestől

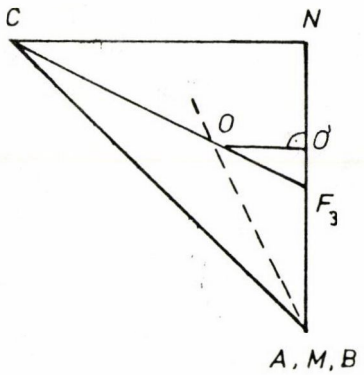
nő, az ANC és az AMC síkoktól való távolsága nem változik, ezért a mozgatás végrehajtható. Ezekután a 6. fejezet mozgatásaival eljutunk a „sejtett" gömbbe.

Tehát egy tetszőleges g gömböt sugarának csökkentése nélkül eljuttattuk a „sejtett" gömbbe, így az lesz a maximális sugarú a feltételeknek eleget tevő gömb.

A gömbkitöltés sűrűsége

A előzőek szerint az O pontnak az MNC háromszöglapon a CF_3 szakaszon és az NMC szög szögfelezőjén kell lennie. Legyen r a keresett sugár. Az O pontnak MN-re való merőleges vetülete O'.

Legyen $MN = 1$. Nyilván $O'F_3 = \frac{r}{2}$ és $F_3M = \frac{1}{2}$ (12. ábra).



12. ábra

MOO' háromszögből: $\dfrac{2r}{1+r} = \operatorname{tg}\dfrac{\pi}{8} = \dfrac{1}{1+\sqrt{2}}$ innen

$$r = \frac{1}{1+2\cdot\sqrt{2}} = \frac{2\sqrt{2}-1}{7} \cong 0,261.$$

A gömbkitöltés sűrűsége:

$$\delta = \frac{V_g}{\frac{1}{2}V_T} = \frac{\frac{4}{3}r^3\pi}{\frac{1}{3}} = 4r^3\pi \cong 0,224.$$

Így meghatároztuk a $2 \circ F\bar{4}3m = P\bar{4}3m$ tércsoport optimális gömbkitöltését, és megadtuk annak a sűrűségét.

7. $APn\bar{3}m$ tércsoport optimális gömbkitöltése

Használjuk a korábban alkalmazott jelöléseket (6. ábra). Bővítsük az $F\bar{4}3m$ tükrözéscsoportot a négy elemű diédercsoporttal. Ezt a 222 jelű csoportot az F_1F_3 és MN körüli másodrendű forgatások generálják. A tércsoport jele $222 \circ F\bar{4}3m = Pn\bar{3}m$.

E tércsoport oiptimális gömbkitöltésének a meghatározásánál most már csak azt emeljük ki, hogy az $ABCD$ tetraéderben elhelyezkedő négy egymásba nem nyúló gömböt az F_1F_3 és MN körüli 180 fokos forgatás páronként egymásba viszi.

Jelöljük g-vel az $ABCD$ tetraéderben elhelyezett gömböt, amely érinti F_1F_3 egyenesét és CMA illetve CNA lapokat, továbbá O középpontja a CF_3 szakaszon van. A g gömb sugara: $r = \dfrac{\sqrt{3}}{2(\sqrt{6}+1)} \cong 0,251$.

Forgassuk el a g gömböt F_1F_3 körül 180 fokkal úgy, hogy a középpontja az AF_3 szakaszra kerüljön, legyen ez a gömb g^1. Majd g-t és g^1-et forgassuk el MN körül 180 fokkal, ekkor g képe legyen g^2 és g^1 képe legyen g^3. Legyen a g^i gömb középpontja O^i $(i = 1, 2, 3)$. Nyilvánvaló, hogy ezen négy gömböt az $ABCD$ tetraéder tartalmazza, továbbá, hogy ezekre teljesülnek a 7. probléma feltételei.

Az optimalitás bizonyításában felhasználjuk a következő lemmát. (Az MN tengelyű $r = \dfrac{\sqrt{3}}{2(\sqrt{6}+1)}$ sugarú henger az $OO^1O^2O^3$ tetraéderből minden csúcsnál levág egy szögletet.)

7.1 LEMMA. *Az $OO^1O^2O^3$ tetraéderben az O és O^1 csúcsoknál az előbbi módon keletkezett szögletek tetszőleges két pontjának távolsága kisebb vagy egyenlő mint $OO^1 = 2r$.*

Az optimalitás bizonyítása a 6. pontban ismertetett módszerrel történik, ezért a bizonyítást nem részletezzük.

A $Pn\bar{3}m$ tércsoport optimális gömbkitöltésének sűrűsége:

$$\delta = \frac{V_g}{\frac{1}{4}V_T} = \frac{\frac{4}{3}r^3\pi}{\frac{1}{6}} = 8r^3\pi \cong 0,398.$$

Következmény. A 7. probléma megoldása a $\bar{4} \circ F\bar{4}3m = I\bar{4}3m$ tércsoport optimális gömbkitöltését is megadta. Az $I\bar{4}3m$ tércsoport az $F\bar{4}3m$ tükrözéscsoport négyelemű $\bar{4}$ csoporttal való bővítésével keletkezett. A $\bar{4}$ csoportot az MN körüli negyedrendű forgatás és ponttükrözés kompozíciója generálja, így az MN körüli másodrendű forgatás is hozzátartozik. Az $I\bar{4}3m$ tércsoport optimális gömbkitöltése és annak a sűrűsége megegyezik a $Pn\bar{3}m$ csoportéval.

8. Az $Fd\bar{3}m$ tércsoport optimális gömbkitöltése

Használjuk a korábbi jelöléseket és a 7. pont eredményeit. A $2 \circ F\bar{4}3m = Fd\bar{3}m$ tércsoport az $F\bar{4}3m$ tércsoport F_1F_3 körüli másodrendű forgatással való bővítésével keletkezett.

a. Ezen tércsoport optimális gömbkitöltésének meghatározásánál olyan $ABCD$ tetraéderben elhelyezkedő két gömböt tekintünk, amelyek nem nyúlnak egymásba, és egymásból F_1F_3 körüli 180 fokos forgatással származtathatók.

b. Nyílván az O és O' középpontú $r = \dfrac{\sqrt{3}}{2(\sqrt{6}+1)}$ sugarú gömbök eleget tesznek a fenti feltételnek. Látható, hogy ha lenne kedvezőbb két középpont azt az $OO^1O^2O^3$ tetraéderből kell választani, továbbá ezen két gömbközéppont távolsága F_1F_3 szakasztól nagyobb kell, hogy legyen mint r. Azonban az $OO^1O^2O^3$ tetraéder benne van egy F_1F_3 tengelyű r sugarú hengerben, vagyis az O és O^1 középpontú r sugarú gömböknél kedvezőbb elhelyezés nincs.

Az optimális gömbkitöltés sűrűsége: $\delta = \dfrac{V_g}{\frac{1}{2}V_T} \cong 0,199.$

Kiemeljük, hogy az optimális gömbkitöltésnek mint gömbrendszernek is $Fd\bar{3}m$ a szimmetriacsoportja.

Így egy meglepően ritka optimális sűrűségű gömbkitöltést kapunk.

IRODALOM

[1] COXETER, H. S. M., *A geometriák alapjai* (Műszaki Könyvkiadó, Budapest, 1987).

[2] MÁTÉ CSILLA, *Tércsoportok optimális gömbkitöltésének meghatározása számítógéppel*, Szakdolgozat (ELTE TTK, Budapest, 1994).

[3] *International tables for X-Ray Crystallography* (Kynoch, Birmingham, 1969, 1979).

[4] MOLNÁR E., „Konvexe Fundamentalpolyeder und D–V Zellen für 29 Raumgruppen, die Coxetersche Spigelungsuntergruppen enhalten", *Beitrage zur Algebra und Geometrie* **14** (1984), 33–75.

[5] SINOGOWITZ, U., „Herleitung aller homogenen nicht kubischen Kugelpackungen", *Z. Kristallographie* **105** (1943), 23–52.

[6] SZIRMAI J., „Optimale Kugelpackungen für die Raumangruppen $F23$, $P432$ und $F432$", *Periodica Polytechnica* **36** (1992), 317–331.

[7] *Atlas prostransvennyh groupp kubicheskoj sistemy* (Nauka, Moszkva, 1980).

[8] DELONE, B. N. and SANDAKOVA, N. N., „Teorija stereoedrov", *Trudy MIAN SSSR* **64** (1961), 28–51.

(Beérkezett: 1992. március 10.)

(Átdolgozva beérkezett: 1993. május 15.)

SZIRMAI JENŐ
BME GEOMETRIA TANSZÉK
BUDAPEST XI. EGRY JÓZSEF U. 1. H. 22

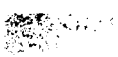
OPTIMALE KUGELPACKUNGEN UNTER EINIGEN RAUMGRUPPEN

J. SZIRMAI

In der Arbeit werden vier einfach transitive Kugelpackungen untersucht: d.h. solche Kugelsysteme, deren Symmetriegruppe eine gegebene kristallographische Raumgruppe ist, ferner, zu zwei beliebigen Kugeln der Packung gibt es genau ein Element der gegebenen Raumgruppe, das die erste Kugel in die zweite überfürt.

Diese vier Raumgruppen $P\bar{4}3m$, $Fd\bar{3}m$, $Pn\bar{3}m$, $I\bar{4}3m$ (siehe z.B. im [4]) kann man durch die Erweiterung der Spiegelungsgruppe $F\bar{4}3m$ (Abb. 1. mit einem einem sphenoidalen Fundamentalbereich) gewinnen.

In der Arbeit bestimmen wir die optimalen Kugelpackungen und ihre maximalen Dichten für die obenen vier Gruppen mit synthetischen Hilfsmitteln. Es scheint sich eine interessante Anordnung su sein, die zu der gruppe $Fd\bar{3}m$ gehört, denn diese Anordnung hat eine sehr kleine maximale Dichte 0.199.



Alkalmazott Matematikai Lapok **17** (1993) 101–121

A HŐVEZETÉSI EGYENLET ÉS NUMERIKUS MEGOLDÁSÁNAK KVALITATÍV TULAJDONSÁGAI*
I. AZ ELSŐFOKÚ KÖZELÍTÉSEK NEMNEGATIVITÁSA

FARAGÓ ISTVÁN, HAROTEN HARITON, KOMÁROMI NÁNDOR, PFEIL TAMÁS

Budapest

Egy folytonos feladat numerikus megoldási módszerének lényeges tulajdonsága, hogy a konvergencia mellett a megoldásfüggvény legfontosabb kvalitatív tulajdonságai átöröklődjenek az egyes numerikus megoldásokra, mivel ezek a tulajdonságok alapvető módon jellemzőek a folytonos feladatra. Cikkünk első részében a parabolikus típusú, másodrendű, lineáris parciális differenciálegyenlet fontosabb kvalitatív tulajdonságait fogalmazzuk meg, majd ezek közül a nemnegativitási tulajdonságnak a numerikus megoldás során történő megőrzésére adunk meg feltételeket. Ebben a részben a feladat térbeli diszkretizálására lineáris véges elemeket illetve az ismert véges differenciás sémákat alkalmazzuk. Megvizsgáljuk a kérdést a különböző peremfeltételekre, illetve a térben kétdimenziós feladatra is.

1. Bevezetés

A parabolikus típusú, másodrendű, lineáris parciális differenciálegyenlet az egyik leggyakrabban vizsgált és alkalmazott matematikai modell. Numerikus megoldásának alapvető követelménye a konvergencia, azaz, hogy a numerikus megoldások sorozata — valamilyen értelemben — konvergáljon a folytonos feladat megoldásához. A különböző módszerekre általában ezt a kérdéskört szokás tárgyalni. Ugyanakkor lényeges, hogy a folytonos feladat megoldásának legfontosabb kvalitatív tulajdonságai átöröklődjenek az egyes numerikus megoldásokra, hiszen ezek a tulajdonságok alapvető módon jellemzőek nemcsak a folytonos feladatra, hanem a modellezett fizikai jelenségre is.

Cikkünkben — az ismert eredményeket összefoglalva illetve újakat adva — az utóbbi kérdéskörrel foglalkozunk. Így mi nem foglalkozunk az alkalmazott numerikus sémák konvergenciájának kérdésével, ez több helyen is megtalálható, [4], [16], [21].

A cikk két részből áll.

Az első részben a folytonos feladat fontosabb kvalitatív tulajdonságaival, majd ezek közül a nemnegativitási tulajdonság numerikus megoldásra történő megmaradásával foglalkozunk.

Az első fejezetben összefoglaljuk az általános alakú parabolikus, lineáris, másodrendű feladatok legfontosabb kvalitatív tulajdonságait és megfogalmazzuk ezeket

*A dolgozat a T 4385 számú OTKA kutatási program keretében készült.

a speciális, hővezetési feladatokra is. A második fejezetben a térbeli diszkretizációra a lineáris véges elemeket, míg az időbelire a jól ismert egylépéses, egyparaméteres sémát alkalmazzuk. Megvizsgáljuk a nemnegativitás kérdését a különböző peremfeltételekre, illetve a térben kétdimenziós feladatokra is. A térbeli felosztás számának függvényében mondunk ki szükséges és elégséges feltételeket a véges differenciás illetve véges elemes térbeli diszkretizációkra.

1. A megoldás kvalitatív tulajdonságai

Ebben a fejezetben az egy- és többdimenziós homogén hővezetési egyenlet és a hozzá tartozó ún. vegyes feladat megoldásának egyes jól ismert kvalitatív tulajdonságait foglaljuk össze, majd az időbeli monotonitás kérdésével foglalkozunk.

Legyen $n \in \mathbb{N}^+$, $\Omega \subset \mathbb{R}^n$ korlátos, sima peremű tartomány, valamint $T \in \mathbb{R}^+$ vagy $T = +\infty$ esetén $Q_T := (0,T) \times \Omega$. Tekintsük a következő másodrendű parabolikus egyenletet:

$$\frac{\partial u}{\partial t} = \sum_{i=1}^{n} \frac{\partial^2 u}{\partial x_i^2}; \quad (t,x) \in Q_T, \tag{1.1}$$

ahol $u \in C^{1,2}(Q_T) \cap C(\bar{Q}_T)$. Az egyenlet egy megoldásának legnagyobb értékéről szól a jól ismert maximumelv (bizonyítása megtalálható a [3] vagy [19] könyvben), amely szerint ha u az (1.1) egyenlet megoldása, akkor u a maximális értékét felveszi a zárt alaplapon vagy a zárt palláston is, azaz a $(\{0\} \times \bar{\Omega}) \cup ([0,T] \times \partial\Omega)$ halmazon is. A maximumelvet u helyett $-u$-ra elmondva a megoldás legkisebb értékére vonatkozó hasonló állítás kapható.

A általánosabb alakú egyenletekre és tartományokra vonatkozó maximumelv megtalálható FRIEDMAN [7] és SMOLLER [19] könyvében. Ezen eredmények speciális esete az, hogy az (1.1) egyenlet jobb oldalából x-nek egy tetszőleges nemnegatív függvényét levonva a maximumelv állítása érvényben marad.

Jelölje $\Omega_0 := \{0\} \times \Omega$ a Q_T tartomány alaplapját, $\Gamma_T := [0,T) \times \partial\Omega$ pedig a palástját! Legyen u_0 az Ω tartományon, g pedig a Γ_T palást lezártján adott folytonos függvény. Az (1.1) egyenletet az

$$u(0,x) = u_0(x); \quad x \in \Omega \tag{1.2}$$

kezdeti feltétellel és az

$$u(t,x) = g(t,x); \quad t \in [0,T], \quad x \in \partial\Omega \tag{1.3}$$

peremfeltétellel első (Dirichlet-) peremfeltételű vegyes feladatnak hívjuk.

Az (1.1)–(1.3) vegyes feladat megoldásának egyértelműsége a maximumelv egyszerű következményeként adódik. A maximumelv kisebb átfogalmazása az alábbi állítás.

1.1. KÖVETKEZMÉNY. *Ha az* (1.1)–(1.3) *vegyes feladatnak van megoldása az* $u_0 \geq 0$ *és* $g \geq 0$ *feltételek mellett, akkor az* u *megoldás is nemnegatív.*

Legyen az $\Omega \subset \mathbb{R}^n$ tartomány pereme háromszor folytonosan differenciálható $(n-1)$-dimenziós felület. Tegyük fel, hogy az u_0 és g egyesítéseként kapható $\Omega_0 \cup \bar{\Gamma}_T$-n adott függvény előáll egy $f \in C^{2,3}(Q_T) \cap C(\bar{Q}_T)$ függvény leszűkítéseként.

1.1. TÉTEL. *A megadott feltételek mellett az* (1.1)–(1.3) *vegyes feladatnak van megoldása a* $C^{1,2}(Q_T) \cap C(\bar{Q}_T)$ *függvénytérben.*

Megjegyzés. Ha a kezdeti feltételt jelentő u_0 függvény nem folytonos, akkor a vizsgált vegyes feladatnak nyilvánvalóan nincs klasszikus megoldása. Ugyanakkor $u_0 \in L_2(\Omega)$ és $g \in L_2(\bar{\Gamma}_T)$ esetén létezik ún. gyenge megoldás (ld. pl. [18]), amire a homogén Dirichlet-peremfeltétel esetén szintén fennáll a nemnegativitási tulajdonság [10]. A vizsgált vegyes feladatnak akkor sincs minden esetben klasszikus megoldása, ha a mellékfeltételként szereplő u_0 és g függvényről megköveteljük a folytonosságot. A Hölder-folytonos függvény definícióját bevezetve az 1.1. tétel feltételeinél általánosabb esetben is igazolható a klasszikus megoldás létezése (ld. pl. FRIEDMAN [7]).

Legyen most $h \geq 0$ a $\partial\Omega$-n adott folytonos függvény, valamint $g \in C(\bar{\Gamma}_T)$. Az (1.1) egyenletet az (1.2) kezdeti és a

$$\partial_\nu u(t,x) + h(x)u(t,x) = g(t,x), \quad t \in [0,T], \quad x \in \partial\Omega \tag{1.4}$$

peremfeltétellel harmadik peremfeltételes vegyes feladatnak nevezzük (itt ν jelölte a tartományból kifelé mutató normális irányt).

Speciálisan, ha $h = 0$, akkor második (Neumann-) peremfeltételes vegyes feladatról van szó.

SMOLLER [19] 10.1. tételéből nyilvánvalóan adódik, hogy az (1.1), (1.2), (1.4) vegyes feladatra is érvényes az 1.1. következményhez hasonló nemnegativitási tulajdonság.

1.2. KÖVETKEZMÉNY. *Tegyük fel, hogy az* $\Omega \subset \mathbb{R}^n$ *tartomány pereme egyszer folytonosan differenciálható. Legyen* $u_0 \geq 0$, $g \geq 0$ *és* $h \geq 0$. *Ha az* (1.1), (1.2), (1.4) *vegyes feladatnak van megoldása, akkor az is nemnegatív.*

Legyen az $\Omega \subset \mathbb{R}^n$ tartomány pereme kétszer folytonosan differenciálható $(n-1)$-dimenziós felület. Legyen $h \in C(\partial\Omega)$, $g \in C(\bar{\Gamma}_T)$ és $u_0 \in C(\bar{\Omega})$.

1.2. TÉTEL. *Az előbbi feltételek mellett az* (1.1), (1.2), (1.4) *vegyes feladatnak van megoldása a* $C^{1,2}(Q_T) \cap C^{0,1}(\bar{Q}_T)$ *függvénytérben.*

Ennek bizonyítása szintén megtalálható FRIEDMAN [7] könyvében.

Tekintsük az (1.1) egyenletet $T = +\infty$ esetén az (1.2) kezdeti feltétellel és az

$$u(t,x) = 0, \quad t \geq 0, \quad x \in \partial\Omega \tag{1.6}$$

homogén első peremfeltétellel. FRIEDMAN [7] általánosabb parabolikus egyenletekre vonatkozó eredményének speciális eseteként kapjuk, hogy az (1.1), (1.2), (1.6) vegyes feladat megoldása $t \to +\infty$ esetén nullához tart x-ben egyenletesen az Ω halmazon. Ez más szavakkal azt jelenti, hogy a megoldás $t \to +\infty$ esetén $\bar{\Omega}$-beli maximumnormában tart a nullához, azaz érvényes a következő állítás.

1.3. KÖVETKEZMÉNY. *Ha az* (1.1), (1.2), (1.6) *vegyes feladat megoldása* u, *akkor*

$$\lim_{t \to +\infty} \max\{|u(t,x)| : x \in \bar{\Omega}\} = 0,$$

azaz a $t \mapsto u(t,\cdot)$ *függvény* $C(\bar{\Omega})$*-beli maximumnormában nullához tart, ha* $t \to +\infty$.

Igazolható (PFEIL [13]), hogy a

$$\lim_{t \to +\infty} u(t,x) = 0, \quad x \in \Omega$$

konvergencia minden rögzített $x \in \Omega$ esetén megfelelően nagy t-re monoton. Emellett, ha az u_0 függvénynek a homogén peremfeltétellel ellátott Laplace-operátor sajátfüggvényei szerinti Fourier-sorában az első együttható nullától különböző, akkor található olyan $T > 0$, hogy tetszőleges $x \in \Omega$ mellett a $t \mapsto u(t,x)$ függvény szigorúan monoton a $[T, +\infty)$ intervallumon [13], [14]. Amennyiben az u_0 függvény első Fourier-együtthatója nulla, úgy a T monotonitási küszöb nem adható meg x-től függetlenül egyenletesen [13].

Most vizsgáljuk az (1.1), (1.2), (1.6) homogén peremfeltételű vegyes feladatot a $(0, L)$ egydimenziós intervallumon. Ebben a speciális esetben a következő tétel nemcsak a fenti tulajdonságú T létezését mondja ki, hanem egyben becslést is ad rá.

Legyen $u_0 \in C^5[0, L]$, és szinuszos Fourier-sorának első együtthatója (ξ_1) különbözzön nullától. Legyen továbbá $K := \max\{|u_0^{(5)}(x)| : x \in [0, L]\}$.

1.3. TÉTEL. *Ha a kezdeti feltételt megadó* u_0 *függvényre teljesül* $u_0 \in C^5[0, L]$, *valamint* $u_0^{(i)}(0) = u_0^{(i)}(L) = 0$ $i = 0, 2, 4$ *esetén, akkor*

$$T := \frac{L^2}{3\pi^2} \log \frac{KL^5}{3|\xi_1|\pi^3}$$

mellett minden rögzített $x \in (0, L)$ *esetén a* $t \mapsto u(t,x)$ *függvény szigorúan monoton fogy a* $[T, +\infty)$ *intervallumon, ha* u_0 *első Fourier-együtthatója pozitív; ha pedig ez negatív, akkor a vizsgált függvény szigorúan monoton növő.*

Bizonyítás. Az (1.1), (1.2), (1.6) vegyes feladat megoldása

$$u(t,x) = \sum_{k=1}^{\infty} \xi_k e^{-\frac{k^2\pi^2}{L^2}t} \sin \frac{k\pi}{L} x, \quad (t,x) \in [0, +\infty) \times [0, L], \tag{1.7}$$

ahol

$$u_0(x) = \sum_{k=1}^{\infty} \xi_k \sin \frac{k\pi}{L} x, \quad x \in [0, L].$$

Az (1.7) Fourier-sor t szerint tagonként differenciálható a tagonkénti deriválással kapott függvénysor egyenletes konvergenciája miatt, így

$$\frac{\partial u}{\partial t}(t, x) = -\sum_{k=1}^{\infty} \xi_k \frac{k^2\pi^2}{L^2} e^{-\frac{k^2\pi^2}{L^2}t} \sin \frac{k\pi}{L} x, \quad (t, x) \in (0, +\infty) \times (0, L).$$

A sort tovább alakítva kapható

$$\frac{\partial u}{\partial t}(t, x) = -\frac{\pi^2}{L^2} e^{-\frac{\pi^2}{L^2}t} \sin \frac{\pi}{L} x \left(\xi_1 + e^{-3\frac{\pi^2}{L^2}t} \sum_{k=2}^{\infty} k^2 \xi_k e^{-(k^2-4)\frac{\pi^2}{L^2}t} \frac{\sin \frac{k\pi}{L} x}{\sin \frac{\pi}{L} x} \right), \tag{1.8}$$

$$(t, x) \in (0, +\infty) \times (0, L).$$

Legyen $f_k(x) := \dfrac{\sin \frac{k\pi}{L} x}{\sin \frac{\pi}{L} x}$, ahol $x \in (0, L)$, $k \in \mathbb{N}^+$. Teljes indukcióval igazolható, hogy $|f_k| \leq k$ a $(0, L)$ intervallumon, ezért tetszőleges $t > 0$ esetén az (1.8) formulában szereplő sor a következőképpen becsülhető:

$$\left| \sum_{k=2}^{\infty} k^2 \xi_k e^{-(k^2-4)\frac{\pi^2}{L^2}t} f_k(x) \right| \leq \sum_{k=2}^{\infty} k^3 |\xi_k|. \tag{1.9}$$

Mivel $u_0 \in C^5[0, L]$, valamint $u_0^{(i)}(0) = u_0^{(i)}(L) = 0 \quad i = 0, 2, 4$ esetén, így parciális integrálásokkal az u_0 függvényre az intervallum szélein kirótt feltételek miatt

$$\xi_k = \frac{2}{L} \int_0^L u_0(x) \sin \frac{k\pi}{L} x \, dx = \frac{2}{k\pi} \int_0^L u_0'(x) \cos \frac{k\pi}{L} x \, dx = \cdots =$$

$$= \frac{2L^4}{k^5\pi^5} \int_0^L u_0^{(5)}(x) \cos \frac{k\pi}{L} x \, dx,$$

így

$$|\xi_k| \leq \frac{2KL^5}{\pi^5} \cdot \frac{1}{k^5},$$

ahol $K := \max\{|u_0^{(5)}(x)| : x \in [0, L]\}$. Ezért az (1.9) egyenlőtlenségben szereplő mindkét sor konvergens, továbbá

$$\left| \sum_{k=2}^{\infty} k^2 \xi_k e^{-(k^2-4)\frac{\pi^2}{L^2}t} f_k(x) \right| \leq \frac{KL^5}{3\pi^3}.$$

Fennáll $\operatorname{sign}\left(\frac{\partial u}{\partial t}(t,x)\right) = -\operatorname{sign}(\xi_1)$, ha az (1.8) formulában a zárójelben álló kifejezés előjelét ξ_1 határozza meg. Ez teljesül, amennyiben érvényes

$$\left| e^{-3\frac{\pi^2}{L^2}t} \frac{KL^5}{3\pi^3} \right| \leq |\xi_1|,$$

azaz

$$t \geq \frac{L^2}{3\pi^2} \log \frac{KL^5}{3|\xi_1|\pi^3}.$$

Megjegyzés. A kezdeti feltételt megadó u_0 függvényről további simaságot feltéve más eredmények is kaphatók T-re.

Megjegyzés. Ha u_0 első szinuszos Fourier-együtthatója nulla, és ξ_p jelöli az első nullától különböző Fourier-együtthatóját, akkor bármely $\left(\frac{k-1}{p}L, \frac{k}{p}L\right)$, $k = 1, 2, \ldots, p$ intervallum kompakt részhalmazához adható meg a monotonitási küszöb x-től függetlenül [13].

Az előzőhöz hasonló tétel kapható a térben többdimenziós téglatesten megadott (1.1), (1.2), (1.6) homogén peremfeltételű vegyes feladat megoldására. Kétdimenzióban például a következőképpen szól az állítás, amely az előzőhöz hasonlóan igazolható.

Legyen $\Omega := (0, L) \times (0, M)$, ahol feltehető, hogy $L \leq M$. Az ezen a kétdimenziós téglalapon adott négyzetesen integrálható és a homogén első peremfeltételt teljesítő függvények terében a

$$\left\{ \sin \frac{j\pi}{L} x_1 \cdot \sin \frac{k\pi}{M} x_2 \ : \ j, k \in \mathbb{N}^+ \right\} \tag{1.10}$$

függvényrendszer teljes ortogonális rendszert alkot.

Tegyük fel, hogy $u_0 \in C^{10}([0, L] \times [0, M])$. E függvény (1.10) rendszer szerinti Fourier-sorának együtthatóit jelölje ξ_{jk}. Legyen továbbá

$$K^* := \max \left\{ \left| \frac{\partial^{10} u_0}{\partial^5 x_1 \partial^5 x_2}(x_1, x_2) \right| : (x_1, x_2) \in [0, L] \times [0, M] \right\}.$$

(A fenti simaságot a kétváltozós megoldásfüggvény Fourier-együtthatóinak az előzőhöz hasonló becslése teszi indokolttá.)

1.4. TÉTEL. *Ha a kezdeti feltételt megadó u_0 függvény, valamint a $\dfrac{\partial^{j+k} u_0}{\partial^j x_1 \partial^k x_2}$ függvények $(j, k) = (2, 0), (4, 0), (0, 5), (2, 5), (4, 5)$ esetén eltűnnek a $(0, L) \times (0, M)$ téglalap peremén, továbbá $\xi_{11} \neq 0$, akkor*

$$T^* := \frac{M^2}{3\pi^2} \log \frac{2L^3 M^5 K^*}{9\pi^6 \left(\frac{1}{L^2} + \frac{1}{M^2}\right) |\xi_{11}|}$$

mellett minden rögzített $(x_1, x_2) \in (0, L) \times (0, M)$ *esetén a* $t \mapsto u(t, x_1, x_2)$ *függvény szigorúan monoton fogy a* $[T^*, +\infty)$ *intervallumon, ha* u_0 *első Fouier-együtthatója pozitív, ha pedig ez negatív, akkor a vizsgált függvény szigorúan monoton növő lesz.*

Az 1.3. és 1.4. tétel következménye az, hogy ha a kezdeti feltételt megadó függvény első Fourier-együtthatója nullától különböző, akkor a megoldás oszcillációmentes, azaz rögzített $x \in (0, L)$, illetve $(x_1, x_2) \in (0, L) \times (0, M)$ esetén a $t \mapsto u(t, x)$, illetve $t \mapsto u(t, x_1, x_2)$ függvények zérushelyei halmazának csak akkor torlódási pontja $+\infty$, ha egy t érték fölött a vizsgált függvény azonosan nulla. Az alábbi eredmény ennek általánosítása, ami speciális esete [13] 2. tételének.

1.4. KÖVETKEZMÉNY. *Az* $\Omega \subset \mathbb{R}^n$ *sima peremű tartományon adott* (1.1), (1.2), (1.6) *vegyes feladatban minden* $x \in \Omega$ *esetén megadható olyan* $T > 0$ *szám, hogy az u megoldással képzett* $t \mapsto u(t, x)$ *függvény monoton a* $[T, +\infty)$ *intervallumon.*

Ebből következően az u megoldás oszcillációmentes.

A homogén peremfeltételű vegyes feladat megoldásával képzett $t \mapsto u(t, .)$ függvény nemcsak a $C(\bar{\Omega})$ tér normájában, hanem közismerten $L_2(\Omega)$-normában is monoton fogyó, ahogy ezt a következő állítás mutatja.

1.5. TÉTEL. *Az* $\Omega \subset \mathbb{R}^n$ *sima peremű tartományon vizsgált* (1.1), (1.2), (1.6) *homogén peremfeltételű vegyes feladat megoldása* $L_2(\Omega)$*-normában monoton fogyó.*

A megoldás $L_2(\Omega)$-normában való monotonitása a [13] dolgozatban definiált általánosabb homogén peremfeltételű parabolikus vegyes feladat esetén is teljesül.

2. A numerikus megoldás nemnegativitása

Ebben a részben megvizsgáljuk, hogy a folytonos feladat előzőekben ismertetett nemnegativitási tulajdonsága milyen feltételek mellett öröklődik át a numerikus megoldásra. (Az ilyen tulajdonságú sémákat monoton sémáknak is szokásos nevezni [17].)

Tekintsük a továbbiakban a

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}; \quad x \in (0, L), \ t > 0, \tag{2.1}$$

$$u(x, 0) = u_0(x), \quad x \in (0, L); \tag{2.2}$$

$$u(0, t) = u(L, t) = 0; \quad t \geq 0 \tag{2.3}$$

egydimenziós, lineáris, parabolikus típusú feladat numerikus megoldását.

Először állítsuk elő a (2.1)–(2.3) probléma numerikus megoldását a véges differenciák módszerével az

$$\Omega_{h,\tau}\{(x_i, t_j), \ x_i = ih, \ h = L/(n+1), \ i = 0, 1, \ldots, n+1, \\ t_j = j\tau, \ \tau > 0, \ j = 0, 1, 2, \ldots\} \tag{2.4}$$

ekvidisztáns rácshálón. A jól ismert approximációk alkalmazásával az $u(x_i, t_j)$ pontos megoldás y_i^j közelítését megkaphatjuk a következő lineáris algebrai egyenletrendszer megoldásával:

$$\frac{y_i^{j+1} - y_i^j}{\tau} = \gamma \frac{y_{i+1}^{j+1} - 2y_i^{j+1} + y_{i-1}^{j+1}}{h^2} + (1-\gamma)\frac{y_{i+1}^j - 2y_i^j + y_{i-1}^j}{h^2} \tag{2.5}$$

$$i = 1, 2, \ldots, n;\ j = 0, 1, \ldots;$$

$$y_i^0 = u_0(x_i);\ i = 1, 2, \ldots, n; \tag{2.6}$$

$$y_0^j = y_{n+1}^j = 0, \quad j = 0, 1, 2, \ldots, \tag{2.7}$$

ahol $\gamma \in [0, 1]$ valamilyen tetszőleges, rögzített paraméter [17]. Jelölje y^j a j-edik időréteghez tartozó ismeretlenek vektorát, azaz az $[y_1^j, y_2^j, \ldots, y_n^j]^T$ n-dimenziós oszlopvektort. Ekkor a (2.5)–(2.7) feladat felírható

$$\mathbf{X}_1 y^{j+1} = \mathbf{X}_2 y^j; \quad j = 0, 1, 2, \ldots; \tag{2.8}$$

$$y^0 \text{ adott} \tag{2.9}$$

alakban, ahol $\mathbf{X}_1$ és $\mathbf{X}_2$ az alábbi, $n \times n$ méretű, egyenletesen kontinuáns mátrixok:

$$\mathbf{X}_1 := \text{tridiag}[-\tau\gamma/h^2;\ 1 + 2\tau\gamma/h^2;\ -\tau\gamma/h^2];$$
$$\mathbf{X}_2 := \text{tridiag}[(1-\gamma)\tau/h^2;\ 1 - 2\tau(1-\gamma)/h^2;\ (1-\gamma)\tau/h^2].$$

(Az (1.7) feltételből közvetlenül adódó $y_0^j = y_{n+1}^j = 0$ értékeket nem vesszük be az egyenletbe.)

Feladatunk a következő: milyen, a τ-ra és h-ra vonatkozó feltételek mellett öröklődik át a (2.1)–(2.3) feladat megoldásának nemnegativitása a (2.8)–(2.9) feladat teljes diszkretizációval szolgáltatott megoldására?

A numerikus megoldás nemnegativitásának szükséges és elégséges feltétele, hogy a nyilvánvalóan invertálható $\mathbf{X}_1$ mátrix inverzének és az $\mathbf{X}_2$ mátrixnak a szorzata nemnegatív legyen, azaz az

$$\mathbf{X} := \mathbf{X}_1^{-1} \cdot \mathbf{X}_2 \tag{2.10}$$

mátrixra teljesüljön az

$$\mathbf{X} \geq 0 \tag{2.11}$$

feltétel. Ehhez egy elégséges feltétel, ha a szorzatban szereplő mindkét mátrix nemnegatív.

Jelölje

$$q = \tau/h^2. \tag{2.12}$$

Az $\mathbf{X}^{-1} \geq 0$ feltétel teljesüléséhez egy elégséges feltétel, ha az $\mathbf{X}_1$ mátrix M-mátrix [23], ami esetünkben γ és q tetszőleges megengedett értéke esetén automatikusan teljesül. Az $\mathbf{X}_2 \geq 0$ feltétel a

$$
\begin{aligned}
& q \leq \frac{1}{2(1-\gamma)}; \quad \text{ha } \gamma \in [0,1); \\
& q \text{ tetszőleges}, \quad \text{ha } \gamma = 1
\end{aligned}
\tag{2.13}
$$

feltételt jelenti. Így (2.13) egy elégséges feltétele a (2.8) séma nemnegativitásának [17]. A fent szereplőnél nagyobb felső korlát is megadható elégséges feltételként. Megengedve az $\mathbf{X}_2$ mátrix főátlójában lévő elemek nennegativitását, LORENZ, J. [11] eredményeit felhasználva STOYAN, G. [20] megmutatta, hogy $n \geq 2$ esetén az alábbi, (2.13)-nél élesebb elégséges feltétel is megadható a vizsgált séma nemnegativitására:

$$
\begin{aligned}
& q \leq \frac{-1 + 2\gamma + \sqrt{1-\gamma(1-\gamma)}}{3\gamma(1-\gamma)}; \quad \text{ha } \gamma \in (0,1); \\
& q \leq 1/2, \quad \text{ha } \gamma = 0; \\
& q \text{ tetszőleges}, \quad \text{ha } \gamma = 1.
\end{aligned}
\tag{2.14}
$$

Megmutatható, hogy n növelésével ezek a korlátok növelhetők [6], ugyanakkor a korlátok sorozata lineáris sebességgel monoton konvergál a

$$
2\gamma(1-\gamma)q^2 + (2-4\gamma)q - 1 + A \leq 0, \tag{2.15}
$$

$$
A = (2q(1-\gamma) - 1\sqrt{1+4\gamma q} \tag{2.16}
$$

q-ra vonatkozó egyenlőtlenség megoldását jelentő szükséges feltételhez. Mindemellett, tetszőleges, rögzített n esetén az $\mathbf{X}_1$ mátrix továbbra is M-mátrix marad.

Megjegyzés. A fenti eredmények az

$$
\begin{aligned}
& \frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + f(x,t); \quad x \in (0,L),\ t > 0, \\
& u(x,0) = u_0(x); \quad x \in [0,L]; \\
& u(0,t) = u(L,t) = 0; \quad t > 0,
\end{aligned}
$$

inhomogén típusú (2.1) egyenlet numerikus megoldásának nemnegativitására is közvetlenül alkalmazhatók. Erre a feladatra ugyanis (2.5) helyett a

$$
\begin{aligned}
& \frac{y_i^{j+1} - y_i^j}{\tau} = \gamma \frac{y_{i+1}^{j+1} - 2y_i^{j+1} + y_{i-1}^{j+1}}{h^2} + (1-\gamma)\frac{y_{i+1}^j - 2y_i^j + y_{i-1}^j}{h^2} + f_{i,\gamma}^{j+\frac{1}{2}} \\
& i = 1,2,\ldots,n; \quad j = 0,1\ldots
\end{aligned}
$$

lineáris algebrai egyenletet kapjuk, ahol az új tag az inhomogenitást jelentő f függvény (x_i, t_j) rácspontban értelmezett valamilyen approximációját jelenti. Ha az f függvény nemnegatív, akkor természetes elvárás, hogy az approximációja ezt a tulajdonságát megőrizze, azaz

$$f_\gamma^{j+\frac{1}{2}} \geq 0, \quad \text{ahol} \quad f_\gamma^{j+\frac{1}{2}} = \left(f_{i,\gamma}^{j+\frac{1}{2}}\right)_{i=1,n} \in \mathbb{R}^n .$$

teljesüljön. Ekkor (2.8), (2.9) helyett időrétegenként az

$$\mathbf{X}_1 y^{j+1} = \mathbf{X}_2 y^j + f^{j+\frac{1}{2}}; \; j = 0, 1, 2, \ldots . \tag{2.17}$$

$$y^0 \text{ adott} \tag{2.18}$$

feladatot oldjuk meg. Mivel $\mathbf{X}_1^{-1} \geq 0$, így a homogén feladatra megfogalmazott eredmények az inhomogén feladatra is érvényesek maradnak.

Áttérünk a (2.1)–(2.3) feladat véges elemes megoldására (részletesebben [4]). A teljes diszkretizálást két lépésben hajtjuk végre.

a. Először a térbeli diszkretizációt hajtjuk végre, amelynek során a szemidiszkrét approximációt

$$u_n(x,t) = \sum_{i=1}^{n} \alpha_i^n(t)\phi_i^n(x) \tag{2.19}$$

alakban keressük, ahol $\phi_i^n(x)$ $(i = 1, 2, \ldots, n)$ valamely $H_0^1(0, L)$-beli véges elemes bázisfüggvény-rendszer. Az egyelőre ismeretlen $\alpha_i^n(t)$ együttható-függvényeket az

$$\mathbf{M}\frac{d\alpha^n(t)}{dt} + \mathbf{Q}\alpha^n(t) = 0; \; t > 0, \tag{2.20}$$

$$\alpha(0) = \alpha^0 \tag{2.21}$$

Cauchy feladat megoldásával kaphatjuk meg, ahol $\mathbf{M}$ és $\mathbf{Q}$ adott, $n \times n$ méretű konstans mátrixok, az ismeretlen $\alpha^n(t)$ vektor komponensei az $\alpha_i^n(t)$ függvények, α^0 pedig a kezdeti függvény (2.19) szerinti approximációjából származtatott, adott vektor.

b. A (2.20)–(2.21) feladat numerikus megoldására az egyparaméteres (egylépéses) módszert választjuk. Ekkor az

$$\mathbf{M}\frac{\alpha^{j+1} - \alpha^j}{\tau} + \mathbf{Q}(\gamma\alpha^{j+1} + (1-\gamma)\alpha^j) = 0 \tag{2.22}$$

lineáris algebrai egyenletrendszert nyerjük, ahol $\alpha^0 \geq 0$ adott vektor, $0 \leq \gamma \leq 1$ adott szám, továbbá α^j jelöli az $\alpha(t)$ vektor $t_j = j\tau$ időrétegen való approximációját.

Vezessük be a

$$\begin{aligned} \mathbf{X}_1 &:= (\mathbf{M} + \tau\gamma\mathbf{Q}) \\ \mathbf{X}_2 &:= \mathbf{M} - \tau(1-\gamma)\mathbf{Q} \end{aligned} \tag{2.23}$$

és a (2.10) jelöléseket. Nyivánvalóan a (2.23) séma nemnegativitásának szükséges és elégséges feltétele az $\mathbf{X}_1^{-1} \cdot \mathbf{X}_2 \geq 0$ egyenlőtlenség. Tegyük fel, hogy u_0 adott, $L_2(0, L)$-beli függvény és $\Phi_i(x)$ az i-edik lineáris bázisfüggvény. Ekkor

$$\mathbf{M} = (h/6)\begin{bmatrix} 4 & 1 & 0 & \ldots\ldots & 0 \\ 1 & 4 & 1 & & \\ & & \ddots & & \\ & & & 1 & 4 \end{bmatrix}, \quad \mathbf{Q} = (1/h)\begin{bmatrix} 2 & -1 & 0 & \ldots\ldots & 0 \\ -1 & 2 & -1 & & \\ & & \ddots & & \\ & & & -1 & 2 \end{bmatrix},$$

ahol $h = L/(n+1)$, továbbá, $\alpha_i^0 = u_0(x_i)$ $(x_i = ih,\ i = 1, 2, \ldots, n)$. Könnyen látható, hogy ekkor $\mathbf{X}_1$ és $\mathbf{X}_2$ a következő szimmetrikus, egyenletesen kontinuáns mátrixok:

a. ha $q\gamma \not\equiv \frac{1}{6}$, akkor

$$\mathbf{X}_1 = z\,\mathrm{tridiag}[-1; \rho; -1], \tag{2.24}$$

$$\mathbf{X}_2 = \mathrm{tridiag}[s; p; s], \tag{2.25}$$

ahol

$$z := q\gamma - \frac{1}{6}; \quad \rho := \frac{\frac{2}{3} + 2\gamma q}{\gamma q - \frac{1}{6}}; \quad s := \frac{1}{6} + q(1-\gamma); \quad p := \frac{2}{3} - 2q(1-\gamma); \tag{2.26}$$

b. ha $q\gamma = \frac{1}{6}$, akkor

$$\mathbf{X}_1 = \mathbf{E} \text{ és } \mathbf{X}_2 = \mathrm{tridiag}[q; 1-2q; q]. \tag{2.27}$$

(Itt $\mathbf{E}$ az egységmátrix, q a (2.12) alatt értelmezett hányados.) (Megjegyezzük, hogy a fenti előállítás csak $n \geq 3$ esetén érvényes. Az $n = 1$ és $n = 2$ eseteket külön tárgyaljuk.)

A továbbiakban explicit alakban előállítjuk az $\mathbf{X} = \mathbf{X}_1^{-1} \cdot \mathbf{X}_2$ mátrixot. Vezessük be a

$$\theta = \mathrm{arch}(\rho/2) \tag{2.28}$$

jelölést és tegyük fel, hogy $n \geq 3$. A [5] dolgozatban megmutattuk, hogy a fenti numerikus megoldás tetszőleges nemnegatív kezdeti feltétel esetén csak a

$$q\gamma \geq \frac{1}{6} \tag{2.29}$$

feltétel mellett maradhat nemnegatív. Ekkor viszont $\rho > 2$ és így

$$(\mathbf{X}_1^{-1})_{i,j} = \begin{cases} \gamma_{i,j}, & \text{ha } i \leq j \\ \gamma_{j,i}, & \text{ha } i > j \end{cases} \tag{2.30}$$

ahol

$$\gamma_{i,j} = \frac{\operatorname{sh}(i\theta) \cdot \operatorname{sh}(n+1-j)\theta}{z \cdot \operatorname{sh}(\theta) \cdot \operatorname{sh}(n+1)\theta} \tag{2.31}$$

(RÓZSA, [16]).

Ez azt jelenti, hogy az $\mathbf{X}$ szimmetrikus mátrix elemei

$$(\mathbf{X})_{i,j} = (\gamma_{i,j-1} + \gamma_{i,j+1})s + \gamma_{i,j}p; \quad \text{ha } i < j, \tag{2.32}$$

$$(\mathbf{X})_{i,i} = (\gamma_{i-1,i} + \gamma_{i,i+1})s + \gamma_{i,i}p, \quad (i = 1, 2, \ldots, n). \tag{2.33}$$

A későbbiekben szükségünk lesz az $n = 1$ és $n = 2$ esetek tárgyalására is. Közvetlen számolással könnyen belátható, hogy $n = 1$ esetén: $\mathbf{X} = \gamma_{1,1}p$; míg $n = 2$ esetén

$$\mathbf{X} = \begin{bmatrix} p\gamma_{1,1} + s\gamma_{1,2} & s\gamma_{1,1} + p\gamma_{1,2} \\ s\gamma_{2,2} + p\gamma_{2,1} & s\gamma_{2,1} + p\gamma_{2,2} \end{bmatrix}.$$

Áttérünk az $\mathbf{X}$ mátrix struktúrájának vizsgálatára. Tegyük fel, hogy $n \geq 2$.

2.1. LEMMA. *Ha valamely $q > 0$ és $\gamma \in [0, 1]$ esetén a (2.29) egyenlőtlenség teljesül, akkor az $\mathbf{X}$ mátrix valamennyi főátlón kívüli eleme pozitív.*

Bizonyítás. Először tekintsük az $n = 2$ esetet. Ekkor

$$(\mathbf{X})_{1,2} = s\gamma_{1,1} + p\gamma_{1,2} = \frac{\operatorname{sh}\theta}{z \cdot \operatorname{sh} 3\theta}(2s \cdot \operatorname{ch}\theta + p).$$

Mivel $2s \operatorname{ch}\theta + p = \rho s + p = q/\left(\gamma q - \frac{1}{6}\right) > 0$, így az állítás $n = 2$ esetén igaz. Most tegyük fel, hogy $n \geq 3$. A $\gamma q = 1/6$ esetén $\mathbf{X} = \mathbf{E}$ és az állítás triviális. Ha $q\gamma > 1/6$, akkor egyszerű számolással adódik, hogy

$$(\mathbf{X})_{i,j} = \gamma_{i,j}(\rho s + p), \tag{2.34}$$

ami az állításunkat jelenti.

2.2. LEMMA. *Ha valamely $q > 0$ és $\gamma \in [0, 1]$ esetén a (2.29) egyenlőtlenség teljesül és az $\mathbf{X}$ mátrix diagonális elemei közül az első nemnegatív, akkor valamennyi diagonális eleme nemnegatív.*

Bizonyítás. Ha $q\gamma = 1/6$, akkor $\mathbf{X}$ diagonális elemei egyenlők, így az állítás triviális. Hasonlóan nyilvánvaló az állítás az $n = 1$ és $n = 2$ esetekre is. Ezért a

továbbiakban elegendő a $q\gamma > 1/6$ és $n \geq 3$ esetet vizsgálni. A (2.33) összefüggésből látható, hogy $(\mathbf{X})_{1,1} = (\mathbf{X})_{n,n}$. Így elegendő belátni, hogy az

$$(\mathbf{X})_{1,1} = \frac{\operatorname{sh}\theta\,[p\operatorname{sh}(n\theta) + s\operatorname{sh}(n-1)\theta]}{\operatorname{sh}\theta \cdot \operatorname{sh}(n+1)\theta} \geq 0 \tag{2.35}$$

összefüggésből következik az

$$(\mathbf{X})_{i,i} = \frac{\operatorname{sh}(i\theta)\operatorname{sh}(n+1-i)\theta}{\operatorname{sh}\theta\operatorname{sh}(n+1)\theta} \cdot B \tag{2.36}$$

$$B := s\left(\frac{\operatorname{sh}(i-1)\theta}{\operatorname{sh}(i\theta)} + \frac{\operatorname{sh}(n-i)\theta}{\operatorname{sh}(n+1-i)\theta}\right) + p$$

$(i = 2, 3, \ldots, n-1)$ kifejezés nemnegativitása; azaz, ha

$$s\frac{\operatorname{sh}(n-1)\theta}{\operatorname{sh}(n\theta)} + p \geq 0 \tag{2.37}$$

akkor minden $i = 2, 3, \ldots, n-1$ esetén a

$$B \geq 0 \tag{2.38}$$

egyenlőtlenség is teljesül. Mivel $s > 0$, így az $y \mapsto s \cdot y + p$ függvény szigorúan monoton növő függvény, ezért elegendő belátni, hogy

$$\frac{\operatorname{sh}(n-1)\theta}{\operatorname{sh}(n\theta)} \leq \frac{\operatorname{sh}(i-1)\theta}{\operatorname{sh}(i\theta)} + \frac{\operatorname{sh}(n-1)\theta}{\operatorname{sh}(n+1-i)\theta} \tag{2.39}$$

érvényes minden $i = 2, 3, \ldots, n-1$ értékre. Tekintsük a

$$t \mapsto \frac{\operatorname{sh}(t-1)\theta}{\operatorname{sh}(t\theta)} + \frac{\operatorname{sh}(n-t)\theta}{\operatorname{sh}(n+1-t)\theta}; \quad t \in [1, n],\ \theta > 0 \tag{2.40}$$

leképezést! Mivel ez a leképezés differenciálható, a deriváltjából közvetlenül meghatározható, hogy az $[1; (n+1)/2]$ intervallumon szigorúan monoton növekszik, míg az $[(n+1)/2, n]$ intervallumon szigorúan monoton csökken, ami az $(\mathbf{X})_{1,1} = (\mathbf{X})_{n,n}$ összefüggést figyelembevéve éppen a bizonyítandó állítást jelenti.

Mivel

$$u_n(x_i, t_j) = \alpha_i^j, \tag{2.41}$$

ezért a 2.1. Lemma és a 2.2. Lemma eredményeit felhasználva közvetlenül megfogalmazhatjuk a lineáris véges elemes séma valamely adott, rögzített felosztás melletti nemnegativitásának szükséges és elégséges feltételét.

2.1. TÉTEL. *Legyen $n \in \mathbb{N}$ egy rögzített felosztásszám. Ekkor tetszőleges nemnegatív kezdeti függvény esetén a numerikus megoldás pontosan akkor marad nemnegatív, amikor $(\mathbf{X})_{1,1}, \geq 0$, azaz teljesül az*

$$\frac{\operatorname{sh}(n-1)\theta}{\operatorname{sh}(n\theta)} \geq -\frac{p}{s}, \quad \text{ha } n \geq 2; \tag{2.42}$$

$$p \geq 0; \qquad \text{ha } n = 1 \tag{2.43}$$

egyenlőtlenség.

Megjegyzés. Mivel $s > 0$, ezért a fenti feltétel feírható

$$\frac{\operatorname{sh}(n-1)\theta}{\operatorname{sh}(n\theta)} \geq -\frac{p}{s}, \quad n \geq 1 \tag{2.44}$$

alakban is.

Bár a 2.1. Tétel szükséges és elégséges feltételt mond ki, a gyakorlatban nehezen kezelhető. Ennek egyrészt az az oka, hogy a numerikus megoldás során nem egy rácshálón oldjuk meg a feladatot, hanem azok sorozatán és így az n szám rögzítése eltér a valós körülményektől. Másrészt, további problémát jelent a feltétel alakja. Egy olyan feltételrendszer megadása lenne célszerű, amely a megválasztandó lépésközre (azaz a h és τ paraméterre) jelent közvetlen korlátozást.

A továbbiakban áttérünk ezek vizsgálatára. Vegyük észre, hogy

$$\frac{\operatorname{sh}(n-1)\theta}{\operatorname{sh}(n\theta)} = \operatorname{ch}\theta - \operatorname{cth}(n\theta) \cdot \operatorname{sh}\theta$$

Így (2.44) alapján $\mathbf{X}$ elemei pontosan akkor nemnegatívak, ha

$$\operatorname{ch}\theta - \operatorname{cth}(n\theta) \cdot \operatorname{sh}\theta > -\frac{p}{s}. \tag{2.45}$$

Mivel az $a(n) := \operatorname{ch}\theta - \operatorname{cth}(n\theta) \cdot \operatorname{sh}\theta$ sorozat szigorúan monoton növekszik, ezért viszonylag egyszerűen megadhatók feltételek (2.44) tejlesülésére. Ugyanis, ha $a(1) \geq -\frac{p}{s}$, akkor ez szükséges és elégséges feltétele annak, hogy minden n-re (2.44) teljesüljön. Ez a $p \geq 0$ feltételt jelenti.

2.2. TÉTEL. *Tetszőleges $n \in \mathbb{N}$ és kezdeti közelítés esetén a lineáris véges elemes séma nemnegativitásának az*

$$\frac{1}{6\gamma} \leq q \leq \frac{1}{3(1-\gamma)} \tag{2.46}$$

egyenlőtlenség szükséges és elégséges feltétele.

Megjegyzés. A (2.46) feltétel megegyezik az [5] cikkbeli feltétellel, de ott csak az alsó korlát szükségességét mutattuk ki.

Megjegyzés. Ez tétel azt jelenti, hogy minden n esetén a séma pontosan akkor tartja meg a nemnegativitást, ha X_1 M-mátrix, X_2 pedig nemnegatív mátrix. Továbbá, hogy ebben az esetben a nemnegativitást csak a $\gamma \geq 1/3$ sémák őrzik meg.

Megjegyzés. A (2.46) feltételben szereplő felső becslés egyben a séma nemnegativitásának elégséges feltétele tetszőleges $n \geq 2$ esetén.

Tegyük fel, hogy $n \geq 2$. Ekkor a nemnegativitás szükséges és elégséges feltétele az

$$a(2) \geq -\frac{p}{s} \tag{2.47}$$

egyenlőtlenség. Mivel

$$a(2) = \frac{\operatorname{sh}\theta}{\operatorname{sh}(2\theta)} = \frac{1}{2\operatorname{ch}\theta} = \frac{1}{x}, \tag{2.48}$$

így egyszerű számolással nyerhető a következő

2.3. TÉTEL. *Tetszőleges $n \geq 2$ és kezdeti közelítés esetén a lineáris véges elemes séma nemnegativitásának a*

$$\frac{1}{6\gamma} \leq q \leq \frac{3(-1+2\gamma)+\sqrt{9-16\gamma(1-\gamma)}}{12\gamma(1-\gamma)} \tag{2.49}$$

szükséges és elégséges feltétele.

Nyilvánvaló a (2.49) felső becslése tetszőleges $n \geq 3$ esetén egyben a séma nemnegativitásának egy elégséges feltétele is. Láthatóan az felosztások számának növelésével a felső korlátok sorozata is növekszik. A növelhetőség korlátjára vonatkozik a következő

2.4. TÉTEL. *Legyen $n \in \mathbb{N}$ tetszőleges felosztásszám. Ekkor minden nemnegatív kezdeti függvény esetén a numerikus megoldás csak akkor maradhat nemnegatív, ha az adott γ esetén q kielégíti a*

$$\gamma(1-\gamma)q^2 - \frac{5}{6}(1-2\gamma)q + \frac{7}{36} + C \leq 0; \tag{2.50}$$

$$C := 2\sqrt{\gamma q + (1/12)}\left((1-\gamma)q - \frac{1}{3}\right) \tag{2.51}$$

egyenlőtlenséget.

Bizonyítás. Az n növelésével az $a(n)$ sorozat szigorúan monoton módon tart a $\operatorname{ch}\theta - \operatorname{sh}\theta$ határértékhez. Így (2.44) alapján tetszőleges n esetén a nemnegativitás

szükséges feltétele

$$\operatorname{ch}\theta - \operatorname{sh}\theta \geq -\frac{p}{s} \tag{2.52}$$

alakban adható meg. Vegyük észre, hogy

$$\operatorname{ch}\theta - \operatorname{sh}\theta = \exp(-\theta) = \exp(-\operatorname{arch}(\rho/2)) =$$
$$= \exp\ln\left(\frac{\rho}{2} + \sqrt{\left(\frac{\rho}{2}\right)^2 - 1}\right)^{-1} = \left(\frac{\rho}{2} + \sqrt{\left(\frac{\rho}{2}\right)^2 - 1}\right)^{-1}.$$

Így a (2.52) feltételből a jelöléseink figyelembevételével közvetlenül a tétel állítását nyerjük.

Nyilvánvaló, hogy ez a nemnegativitásnak egy olyan szükséges feltételrendszere, amely tovább nem csökkenthető.

Vegyük észre, hogy n növelésével az $a(n)$ felhasználásával megadott elégséges feltételek sorozata olyan gyorsan tart a szükséges feltétel által meghatározott felső korláthoz, mint amilyen gyorsan a $\operatorname{cth}(n\theta)$ sorozat tart az n növelésével 1-hez. Ez a konvergencia viszont igen gyors! Ugyanis

$$\operatorname{cth}(n\theta) = 1 + \frac{2}{(\exp\theta)^{2n} - 1}; \tag{2.53}$$

és mivel $\rho > 2$ miatt

$$\exp\theta = \frac{\rho}{2} + \sqrt{\left(\frac{\rho}{2}\right)^2 - 1} > 1,$$

ezért látható módon a $\operatorname{cth}(n\theta)$ értékei már viszonylag kis n esetén is közel vannak egyhez, azaz viszonylag kis n esetén is elégséges feltétel korlátja közel van a szükséges feltétel határához.

A véges differenciás diszkretizációhoz hasonlóan, a nemnegatív jobboldalú inhomogén diffenciálegyenlet esetén a fenti feltételek mellett a véges elemes diszkretizáció is megtartja a megoldás nemnegativitását. Ez könnyen belátható, ha figyelembevesszük, hogy a (2.29) szükséges feltétel éppen az $X_1^{-1} \geq 0$ tulajdonságot biztosítja.

A elsőfajú peremfeltétellel kitűzött parabolikus feladatok numerikus megoldására vonatkozó nemnegativitási eredmények jól alakalmazhatók a másod- illetve a harmadrendű peremfeltételek esetén is. Tekintsük a továbbiakban a (2.1) egyenletet és a (2.2) kezdeti feltételt az

$$u(0,t) = \frac{\partial u(L,t)}{\partial x} = 0, \quad t > 0 \tag{2.54.}$$

illetve a

$$u(0,t) = \frac{\partial u(L,t)}{\partial x} + cu(L,t) = 0, \quad t \geq 0 \tag{2.55}$$

peremfeltételekkel, ahol $c > 0$ adott állandó.

A továbbiakban azzal a kérdéssel foglalkozunk, hogy a (2.1), (2.2), (2.54) illetve a (2.1) (2.2) (2.55) feladatok 1.3. Következmény szerinti nemnegativitási tulajdonsága milyen feltételrendszer esetén öröklődik át a lineáris véges elemes térbeli illetve az egylépéses sémát alkalmazó időbeli diszkretizációval nyert numerikus megoldásra. A fenti eljárással a szemidiszkretizáció után mindkét feladatra egy (2.20), (2.21) típusú Cauchy-feladatot kapunk, ahol a (2.54) második peremfeltétele esetén a

$$\mathbf{M} = \frac{h}{6}\begin{bmatrix} 4 & 1 & 0 & 0 & & & \\ 1 & 4 & 1 & 0 & & & \\ & & & & & & 0 \\ & & & 0 & 1 & 4 & 1 \\ & & & 0 & 0 & 1 & 2 \end{bmatrix}; \quad \mathbf{Q} = \frac{1}{h}\begin{bmatrix} 2 & -1 & 0 & & & & \\ -1 & 2 & -1 & & & & \\ & & & & & & 0 \\ & & & & 0 & -1 & 2 & -1 \\ & & & & 0 & & -1 & 1 \end{bmatrix}$$

alakú $\mathbb{R}^{n\times n}$-beli; míg a (2.55) harmadik peremfeltétel esetén

$$\mathbf{M} = \frac{h}{6}\begin{bmatrix} 4 & 1 & 0 & 0 & & & \\ 1 & 4 & 1 & 0 & & & \\ & & & & & & 0 \\ & & & 0 & 1 & 4 & 1 \\ & & & 0 & 0 & 1 & 2 \end{bmatrix}; \quad \mathbf{Q} = \frac{1}{h}\begin{bmatrix} 2 & -1 & 0 & & & & \\ -1 & 2 & -1 & & & & \\ & & & & & & \\ & & & & 0 & -1 & 2 & -1 \\ & & & & 0 & & -1 & 1+c \end{bmatrix}$$

alakú $\mathbb{R}^{(n+1)\times(n+1)}$-beli mátrixok.

A fenti feladatok numerikus megoldására ismételten alkalmazzuk az egylépéses, γ-paraméteres sémát. Megőrízve a korábbi jelöléseket, az $\mathbf{X}$ mátrix nemnegativitására vonatkozóan, az $\mathbf{X}_1^{-1}$ és az $\mathbf{X}_2$ mátrixok nemnegativitásának elégséges feltételeit megadva a következő tétel kapható [8]:

2.5. TÉTEL. *Tegyük fel, hogy* a (2.1), (2.2), (2.54) *feladat diszkretizációjára a*

$$\frac{1}{6\gamma} \le q \le \frac{1}{3(1-\gamma)}; \quad 1 \ge \gamma \ge \frac{1}{3} \tag{2.56}$$

feltétel, míg a (2.1), (2.2), (2.55) *feladatéra a*

$$\frac{1}{6\gamma} \le q \le \frac{1}{3(1-\gamma)(1+ch)}, \quad \frac{1+ch}{3+ch} \le \gamma \le 1 \tag{2.57}$$

feltétel teljesül. Ekkor a numerikus séma megőrzi a kezdeti feltételt leíró függvény nemnegativitását.

Megjegyezzük, hogy a fenti tételben (2.56) illetve (2.57) a numerikus megoldás nemnegativitásának elégséges feltételei.

A folytonos megoldás nemnegativitása, mint azt az első fejezetben már láttuk, többdimenziós térbeli feladatokra is érvényes. A továbbiakban az erre vonatkozó lineáris véges elemes numerikus sémára adunk meg olyan elégséges feltételt, amely ezt a tulajdonságot átörökíti.

Tekintsük a következő, síkbeli parabolikus feladatot:

$$\frac{\partial u}{\partial t} - \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right) = 0; \quad 0 < x < L,\ 0 < y < L,\ t > 0 \tag{2.58}$$

$$u(x, y, 0) = u_0(x, y); \quad 0 \le x \le L,\ 0 \le y \le L, \tag{2.59}$$

$$u(x, 0, t) = u(x, L, t) = u(0, y, t) = u(L, y, t) = 0;\ 0 \le x,\ y \le L,\ t > 0. \tag{2.60}$$

A szemidiszkrét megoldás meghatározásához kétféle eljárást választottunk: először a

$$\phi_{ij}(x, y) = \phi_i(x)\phi_j(y);\ i, j = 1, n \tag{2.61}$$

alakú, téglalapokon értelmezett, egydimenziós bázisfüggvények Descartes-szorzatát; másodszor pedig a

$$\Phi_{i,j}(x, y) = \begin{cases} 1 - (x_i - x)/h - (y_j - y)/h & x, y \in D^h_{i,j,1} \\ 1 - (x_i - x)/h & x, y \in D^h_{i,j,2} \\ 1 - (y_j - y)/h & x, y \in D^h_{i,j,3} \\ 1 + (x_i - x)/h + (y_j - y)/h & x, y \in D^h_{i,j,4} \\ 1 + (x_1 - x)/h & x, y \in D^h_{i,j,5} \\ 1 - (y_j - y)/h & x, y \in D^h_{i,j,6} \end{cases} \tag{2.62}$$

alakú bázisfüggvényeket választottunk, ahol $D^h_{i,j,k}$ $(k = 1, \ldots, 6)$ az előző felosztás $D^h_{i,j}$ téglalapelemének további, háromszögekre történő alkalmas felosztása. (Részleteket illetően lásd [12].)

Tekintsük először a (2.61) típusú függvényeket! Mint a korábbiakban, ebben az esetben is a szemidiszkretizáció egy (2.20), (2.21) alakú Cauchy problémát eredményez, ahol $\alpha(t)$ az ismeretlen, n^2 dimenziós vektor, míg $\mathbf{M}$ és $\mathbf{Q}$ a következő, blokk-tridiagonális, $n^2 \times n^2$ méretű mátrixok:

$$\mathbf{M} = h^2\,\mathrm{tridiag}[\mathbf{M}_2, \mathbf{M}_1, \mathbf{M}_2];$$

$$\mathbf{Q} = \mathrm{tridiag}[\mathbf{Q}_2, \mathbf{Q}_1, \mathbf{Q}_2].$$

Itt $\mathbf{M}_1, \mathbf{M}_2, \mathbf{Q}_1, \mathbf{Q}_2 \in \mathbb{R}^{n\times n}$ adott tridiagonális mátrixok és alakjuk

$$\mathbf{M}_1 = \text{tridiag}\left[\frac{1}{9}, \frac{4}{9}, \frac{1}{9}\right],$$

$$\mathbf{M}_2 = \text{tridiag}\left[\frac{1}{36}, \frac{1}{9}, \frac{1}{36}\right],$$

$$\mathbf{Q}_1 = \text{tridiag}\left[-\frac{1}{3}, \frac{8}{3}, -\frac{1}{3}\right],$$

$$\mathbf{Q}_2 = \text{tridiag}\left[-\frac{1}{3}, -\frac{1}{3}, -\frac{1}{3}\right].$$

A fenti Cauchy-feladat numerikus megoldására ismételten alkalmazzuk az egylépéses sémát. Megőrízve a korábbi jelöléseket, ezúttal is az $\mathbf{X}$ mátrix nemnegativitására vonatkozó elégséges feltételeket adunk meg.

Tekintsük azt a legtriviálisabb elégséges feltételrendszert, amikor is a (2.23) jelölések szerinti $\mathbf{X}_2$ mátrix nemnegatív, míg az $\mathbf{X}_1$ mátrix M-mátrix. Könnyen ellenőrizhetően ekkor a következő állítást kapjuk.

2.6. TÉTEL. *Tekintsük a (2.58)–(2.60) feladatot. Ha a (2.61) alakú bázisfüggvényeket alkalmazó véges elemes térbeli- illetve az egylépéses módszert alkalmazó időbeli diszkretizációs eljárás paramétereire teljesül az*

$$\frac{1}{3\gamma} \le q \le \frac{1}{6(1-\gamma)}, \quad 1 \ge \gamma \ge \frac{2}{3} \tag{2.63}$$

feltétel, akkor a numerikus megoldás tetszőleges időrétegen megőrzi a kezdeti függvény nemnegativitását.

Tekintsük most a (2.62) típusú bázisfüggvényeket! Ekkor a (2.20), (2.21) alakú Cauchy problémában a mátrixok a következő alakúak:

$$\mathbf{M} = \text{tridiag}\, h^2\,[\mathbf{M}_3, \mathbf{M}_1, \mathbf{M}_2];$$

$$\mathbf{M}_1 = \text{tridiag}\left[\frac{1}{12}, \frac{1}{2}, \frac{1}{12}\right];$$

$$\mathbf{M}_2 = \text{tridiag}\left[\frac{1}{12}, \frac{1}{12}, 0\right];$$

$$\mathbf{M}_3 = \text{tridiag}\left[0, \frac{1}{12}, \frac{1}{12}\right];$$

$$\mathbf{Q} = \text{tridiag}\,[\mathbf{E}, \mathbf{Q}_1, \mathbf{E}];$$

$$\mathbf{Q}_1 = \text{tridiag}\,[-1, 4, -1].$$

Megismételve a (2.61) típusú függvényeknél ismertetett gondolatmenet, a következő, elégséges feltételrendszert szolgáltató állítás adható meg:

2.7. TÉTEL. *Tegyük fel, hogy a* (2.58)–(2.60) *feladat numerikus megoldására a 2.6. Tételben leírt módszert alkalmazzuk a* (2.62) *bázisfüggvényekkel. Ekkor, ha az eljárás paramétereire teljesül a*

$$\frac{3+\sqrt{14}}{12\gamma} \le q \le \frac{1}{8(1-\gamma)}, \quad \frac{24+8\sqrt{14}}{36+8\sqrt{14}} \le \gamma \le 1 \tag{2.64}$$

feltétel, akkor a módszer tetszőleges időrétegen megőrzi a kezdeti függvény nemnegativitását.

Végezetül megjegyezzük, hogy STOYAN [20] egydimenziós esetre vonatkozó eredményeihez hasonlóan, a (2.63) és (2.64) becslések LORENZ [11] eredményeinek felhasználásával tovább javíthatók [8].

Köszönetnyilvánítás. A szerzők köszönetüket fejezik ki Rózsa Pálnak a másodrendű közelítésekkel kapcsolatos lineáris algebrai problémák megoldásában, illetve Stoyan Gisbertnek a numerikus nemnegativitás kérdéskörében nyújtott segítségéért. Megköszönik továbbá Tóth Jánosnak a cikk elkészítéséhez adott hasznos tanácsait.

IRODALOM

[1] BABUSKA, I., JANIK, T., „The $h-p$ version of the finite element method for parabolic equations", Part I: The p-version in time, *Numer. Method for Partial Diff. Eqs.* **5** (1989), 363–369.

[2] BABUSKA, I., JANIK, T., „The $h-p$ version of the finite element method for parabolic equations", Part II: The $h-p$-version in time, *Numer. Method for Partial Diff. Eqs.* **6** (1990), 343–369.

[3] BERS, L., JOHN, F., SCHECHTER, M., *Partial Differntial Equations* (Interscience Publisher, New York, 1964).

[4] FARAGÓ, I., „Véges elemek módszere lineáris, parabolikus típusú feladatok megoldására", *Alkalmazott Mat. Lapok* **11** (1985), 123–155.

[5] FARAGÓ, I., KOMÁROMI, N., „Nonegativity of the numerical solution of parabolic problems", *Numerical Methods, North-Holland* **59** (1990), 173–179 (Greenspan, Rózsa, ed.).

[6] FARAGÓ, I., „Positivity of the finite difference scheme for the linear parabolic problems", *Differential Equations, North-Holland* **62** (1991), 113–118 (Farkas, M. and Sebestyén, Z., eds.).

[7] FRIEDMAN, A., *Partial Differential Equations of Parabolic Type* (Prentice-Hall, Englewood Cliffs, 1964).

[8] HAROTEN, H. A., „Nonnegativity of the numerical solution of parabolic problems with differential boundary conditions", *Annales Universitatis Sci. Budapest, Sectio Computorica*, (megjelenés alatt).

[9] HAROTEN, H. A., „Some condition of ρ-stability and the non-oscillation of the numerical solution of linear parabolic problem", *Annales Universitatis, Sci. Budapest, Sectio Computorica*, (megjelenés alatt).

[10] LADYZENSKAJA, O. A., *Lineáris és kvázilineáris parabolikus egyenletek* (Nauka, Moszkva, 1967), (oroszul).

[11] LORENZ, J., „Zur Inversmonotonie diskreter Probleme", *Numer. Math.* **27** (1977), 227–238.

[12] MARCSUK, G. I., *A gépi matematika numerikus módszerei* (Műszaki Könyvkiadó, Budapest, 1976).

[13] PFEIL, T., „Time-Monotonicity of the Solutions of Linear Second Order Homogeneous Parabolic Equations", *Ann. Univ. Sci. Budapest, Eötvös Sect. Math.* **36** (1993), 139–146.

[14] PFEIL, T., „On the Time-Monotononicity of the Solutions of Linear Second Order Homogeneous parabolic Equations on the Whole Domain", *Publ. Math. Debrecen* **45** (1994), (közlésre elfogadva).
[15] RICHTMAYER, R., MORTON, K. W., *Difference methods for initial-value problems* (J. Wiley Publ., New York, 1967).
[16] RÓZSA, P., *Lineáris algebra és alkalmazásai* (Műszaki Kiadó, Budapest, 1976).
[17] SAMARSKII, A. A., *Differnciasémák elmélete* (Nauka, Moszkva, 1977), (oroszul).
[18] SIMON, L., BADERKO, E. A., *Másodrendű lineáris parciális differenciálegyenletek* (Tankönyvkiadó, Budapest, 1983).
[19] SMOLLER, J., *Shock Waves and Reaction-Diffusion Equations* (Springer–Verlag, Berlin, 1983).
[20] STOYAN, G., „On a Maximum Principle for Matrices and on Conservation of Monotonicity", *ZAMM* **62** (1982), 375–381.
[21] SZABÓ, B. A., BABUSKA, I., *Finite Element Analysis* (Wiley, New York, 1991).
[22] THOMÈE, V., „Finite Difference Methods for Linear Parabolic Equations", in Handbook of Numerical Analysis (Ciarlet, P. G. and Lions, J. L., eds.), vol. 1 (Nort-Holland, Amsterdam, 1990).
[23] YOUNG, D. M., *Nagy lineáris rendszerek iterációs megoldása* (Műszaki Könyvkiadó, Budapest, 1979).
[24] ZIENKIEWICZ, O. C., MORGAN. K., *Finite Elements and Approximation* (Wiley, New York, 1983).

(Beérkezett: 1993. február 2.)

FARAGÓ ISTVÁN
HAROTEN A. HARITON
PFEIL TAMÁS
EÖTVÖS LORÁND TUDOMÁNYEGYETEM
ALKALMAZOTT ANALÍZIS TANSZÉK
1088 BUDAPEST
MÚZEUM KRT. 6–8.

KOMÁROMI NÁNDOR
AGRÁRTUDOMÁNYI EGYETEM
AGRÁRMARKETING TANSZÉK
2103 GÖDÖLLŐ
PÁTER KÁROLY U. 1.

THE DIFFERENTIAL EQUATION OF THE HEAT TRANSFER AND QUALITATIVE PROPERTIES ITS NUMERICAL SOLUTIONS: I. THE NONNEGATIVITY OF THE FIRST ORDER APPROXIMATIONS

I. FARAGÓ, H. A. HARITON, N. KOMÁROMI AND T. PFEIL

The most important requirement to the numerical methods — besides the convergence — is conserving the most characteristic qualitative properties of the solution of the original problem to the numerical solutions. In the first part of the paper we formulate the most important qualitative properties of the solution of linear second order parabolic problems. We give the necessary and sufficient conditions of conserving the nonnegativity of the numerical solutions. We examine the finite difference method and the finite element method with linear elements. We consider not only the one dimensional problems with first boundary conditions but both the problems woth different boundary conditions and in two dimensional, too.

Alkalmazott Matematikai Lapok 17 (1993) 123–141

A HŐVEZETÉSI EGYENLET ÉS NUMERIKUS MEGOLDÁSÁNAK KVALITATÍV TULAJDONSÁGAI*
II. A MÁSODFOKÚ KÖZELÍTÉS NEMNEGATIVITÁSA, A MAXIMUM ELV ÉS AZ OSZCILLÁCIÓMENTESSÉG

FARAGÓ ISTVÁN, HAROTEN HARITON, KOMÁROMI NÁNDOR, PFEIL TAMÁS

Budapest

Cikkünk második részében — az első részben megfogalmazott célnak megfelelően — további kvalitatív tulajdonságokat vizsgálunk meg. Először a kvadratikus véges elemes módszer nemnegativitására vonatkozó szükséges és elégséges feltételeket fogalmazzuk meg, majd a lineáris véges elemes séma időbeli monotonitásának feltételét adjuk meg, az L_2 és a maximum normákban. Végezetül a numerikus megoldás oszcillációmentességének feltételét adjuk meg a lineáris véges elemes sémákra.

1. Bevezetés

A dolgozat első részében megvizsgáltuk az elsőfokú véges elemes bázisfüggvényekkel nyert térbeli approximáció és az egylépéses véges differenciás időbeli approximáció által kapott numerikus megoldás nemnegativitásának feltételét. (Cikkünk második részében az első rész formuláira „I/sorszám" jelöléssel hivatkozunk.)

A véges elemek módszerének különböző változatai ismeretesek, amelyek a konvergens numerikus megoldássorozat előállításának módszerében térnek el egymástól:

1. a közelítés fokszámának változatlan hagyása mellett a tartomány felosztásának finomításával („h–verzió");

2. a közelítő polinom fokszámának növelésével („p–verzió");

3. a tartomány finomításának és a közelítő polinomok fokszámát egyszerre növeljük („h–p verzió").

... Dolgozatunkban mi csak a h–verzióval foglalkozunk. Ugyanakkor megjegyezzük, hogy a h–p verzió egy viszonylag új és egyre szélesebb körben terjedő irányzat. Elterjedését elsősorban annak köszönheti, hogy a másik két verziótól eltérően exponenciális gyorsaságú konvergenciát biztosít. Ezidáig elsősorban a mechanikai számításokban, és ott is elliptikus feladatokra alkalmazták; a parabolikus feladatokra vonatkozó publikáció még viszonylag kis számban jelent meg [1], [2].

A h–verzió parabolikus feladatokra való alkalmazása (az első fejezetben tárgyalt módon) könnyen láthatóan a véges elemes sémák alkalmazhatóságára szab ki gyakorlati korlátot. Az időrétegek száma ugyanis — a konvergencia feltételei miatt — általában négyzetesen növekszik térbeli felosztás finomításakor, ami viszont gyakran

*A dolgozat a T 4385 számú OTKA kutatási program keretében készült.

a számítási igény jelentős növekedése miatt alkalmazhatatlanná teszi a sémát. Ezért különösen jelentősek azok a sémák, ahol az approximáció pontosságának növelését más módon, például a bázisfüggvények fokszámának növelésével érjük el. Ekkor természetesen újból megfogalmazódnak a korábbi kérdések. A továbbiakban azt a kérdést vizsgáljuk, amikor másodrendű bázisfüggvényeket alkalmazunk az (I/2.1)–(I/2.3) feladat véges elemes megoldására. (Megjegyezzük, hogy ez nem a p–verziót jelenti, hiszen a konvergenciát a másodfokú elemekre alkalmazott h–verzióval érjük el.)

Ebben a részben további kvalitatív tulajdonságok megőrzésének feltételeivel is foglalkozunk. Nevezetesen, a megoldásra vonatkozó maximum-elvet, illetve normájának időbeli monotonitását, továbbá az oszcillációmentességet vizsgáljuk és feltételrendszereket adunk meg ezek biztosítására.

2. A numerikus megoldás nemnegativitása másodrendű bázisfüggvények esetén

Ebben a részben azt a kérdéskört vizsgáljuk meg, hogy a folytonos feladat előzőekben ismertetett nemnegativitási tulajdonsága milyen feltételek mellett öröklődik át a numerikus megoldásra, ha a térbeli approximáció során a másodfokú spline függvényeket alkalmazzuk. Tekintsük tehát ismételten a

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}; \quad x \in (0, L),\ t > 0, \tag{I/2.1}$$

$$u(x, 0) = u_0(x), \quad x \in (0, L); \tag{I/2.2}$$

$$u(0, t) = u(L, t) = 0;\ t \geq 0 \tag{I/2.3}$$

egydimenziós, lineáris , parabolikus típusú feladat numerikus megoldását.

Legyen $h : L/(n+1)$, $\tau \in \mathbb{R}^+$ és osszuk fel a tartományt az

$$\Omega_{h,\tau}\{(x_i, t_j),\ x_i = ih,\ i = 0, 1, \ldots, n+1,\ t_j = j\tau,\ j = 0, 1, 2, \ldots\} \tag{I/2.4}$$

ekvidisztáns ráccsal. A szemidiszkretizáció során térben másodfokú elemekkel közelítve ismételten egy (I/2.) típusú Cauchy feladatot kapunk, amelyet az első részben ismertetett egylépéses véges differenciás módszerrel oldunk meg. Ekkor a teljes diszkretizáció meghatározásához ismételten egy

$$\mathbf{X}_1 y^{j+1} = \mathbf{X}_2 y^j; \quad j = 0, 1, 2, \ldots; \tag{I/2.8}$$

$$y^0 \text{ adott} \tag{I/2.9}$$

típusú lineáris algebrai egyenletrendszer megoldása szükséges, ahol $\mathbf{X}_1$ és $\mathbf{X}_2$ az

alábbi felépítésű mátrixok:

$$\text{(2.1)} \qquad \mathbf{X}_1 = \begin{bmatrix} a & b & & & & & \\ b & c & b & d & & & \\ & b & a & b & & & \\ & d & b & c & b & d & \\ & & & b & a & b & \\ & & & d & b & c & b & d \\ & & & & & b & a & b \\ & & & & & & \dots & \end{bmatrix} \qquad \mathbf{X}_2 = \begin{bmatrix} y & v & & & & & \\ v & w & v & z & & & \\ & v & y & v & & & \\ & z & v & w & v & z & \\ & & & v & y & v & \\ & & & z & v & w & v & z \\ & & & & & v & y & v \\ & & & & & & \dots & \end{bmatrix}$$

$$\text{(2.2)} \qquad \begin{aligned} a &= 16/30 + 16\gamma q/3 & \qquad y &= 16/30 - 16q(1-\gamma)/3 \\ b &= 2/30 - 8\gamma q/3 & \qquad v &= 2/30 + 8q(1-\gamma)/3 \\ c &= 8/30 + 14\gamma q/3 & \qquad w &= 8/30 - 14q(1-\gamma)/3 \\ d &= -1/30 + \gamma q/3 & \qquad z &= -1/30 - q(1-\gamma)/3 \end{aligned}$$

(Megjegyezzük, hogy $\mathbf{X}_1^{-1}$ minden $q = \tau/h^2 > 0$ és $\gamma \in [0,1]$ érték mellett létezik.)

Feladatunk tehát a következő: τ és h mely megválasztása mellett öröklődik át az (I/2.1)-(I/2.3) feladat megoldásának nemnegativitása (I/2.8)-(I/2.9) feladat teljes diszkretizációt szolgáltató megoldására? Ennek nyilvánvalóan szükséges és elégséges feltétele, hogy az $\mathbf{X}_1$ mátrix inverzének és az $\mathbf{X}_2$ mátrixnak a szorzata nemnegatív legyen, azaz az

$$\text{(I/2.10)} \qquad \mathbf{X} = \mathbf{X}_1^{-1} \cdot \mathbf{X}_2$$

mátrixra teljesüljön az

$$\text{(I/2.11)} \qquad \mathbf{X} \geq 0$$

feltétel.

Legyen $\mathbf{X}_1, \mathbf{X}_2 \in \mathbb{R}^{(2n+1)\times(2n+1)}$. (Ez a feltétel elsősorban az $\mathbf{X}_1$ mátrix invertálásánál játszik szerepet. A q és γ paraméterekre vonatkozó vizsgálatot tehát páratlan méretre végezzük el, de hasonló eredménnyel megismételhető párosra is.)

Az $\mathbf{X}_1$ mátrix invertálásához vezessük be a következő jelöléseket:

$$\mathbf{K} = \begin{bmatrix} 0 & 1 & & & \\ 1 & 0 & 1 & & \\ & 1 & 0 & 1 & \\ & & \ddots & & \\ & & & 1 & 0 \end{bmatrix}, \quad \mathbf{K} \in \mathbb{R}^{(n+1)\times(n+1)}$$

$$\mathbf{P} = \begin{bmatrix} 1\,0\,0\,\ldots & & & & & & \\ 0\,0\,1\,\;0\,\;0\,\ldots & & & & & & \\ 0\,0\,0\,\;0\,\;1\,\;0\,\;0\ldots & & & & & & \\ \quad\ddots & & & & & & \\ 0\,0\ldots & & & & & 0\,1 \\ 0\,1\,\;0\,\ldots & & & & & 0\,0 \\ 0\,0\,\;0\,\;1\,\ldots & & & & & 0\,0 \\ \quad\ddots & & & & & \\ 0\,0\,\;0\,\;0\,\ldots & & & & & 1\,0 \end{bmatrix}, \quad \mathbf{P} \in \mathbb{R}^{(2n+1)\times(2n+1)}$$

$$\mathbf{B}_1 = \begin{bmatrix} 1\,0\ldots & & & & \\ 1\,1\,\;0\,\ldots & & & & \\ 0\,1\,\;1\,\;0\,\ldots & & & & \\ 0\,0\,\;1\,\;1\,\;0\,\ldots & & & & \\ \quad\ddots & & & & \\ & & & 0\,1\,1 \\ & & & 0\,1 \end{bmatrix}, \quad \mathbf{B}_1 \in \mathbb{R}^{(n+1)\times n}$$

(Vegyük észre, hogy $\mathbf{B}_1$ téglalap alakú mátrix, továbbá, hogy $\mathbf{P}$ egy olyan permutálómátrix, amelynek sorai rendre az $e_1, e_3, \ldots, e_{2n+1}, e_2, e_4, \ldots, e_{2n}$ bázisvektorok. Tehát pl. e_3 harmadik eleme 1, az összes többi 0.)

Jelölje $\mathbf{T} := \mathbf{X}_1\mathbf{P}$ a következő hipermátrixot:

$$\left[\begin{array}{c|c} \mathbf{A} & \mathbf{B} \\ \hline \mathbf{C} & \mathbf{D} \end{array}\right]$$

ahol

$$\begin{aligned} \mathbf{A} &= a\mathbf{E} \\ \mathbf{B} &= b\mathbf{B}_1 \\ \mathbf{C} &= b\mathbf{B}_1^T \\ \mathbf{D} &= c\mathbf{E} + d\mathbf{K} \end{aligned} \tag{2.3}$$

($\mathbf{E} \in \mathbb{R}^{(n+1)\times(n+1)}$ egységmátrix.)

Mivel $\mathbf{A}^{-1}$ és $\mathbf{T}^{-1}$ létezik, ezért alkalmazható a

$$\mathbf{T}^{-1} = \left[\begin{array}{c|c} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1})^{-1} \\ \hline -(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} & (\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \end{array}\right] \tag{2.4}$$

előállítás (pl.: Rózsa [16]).

Így a keresett inverzre

$$\mathbf{X}_1^{-1} = \mathbf{P}^T\mathbf{T}^{-1} \tag{2.5}$$

Végezzük el a (2.4)–(2.5) számításokat, azaz adjuk meg $\mathbf{X}_1^{-1}$ mátrixokat közvetlen alakban!

Legyen

$$
\begin{aligned}
g &= (b^2 - ad)/a \\
p &= ag/b^2 \\
s_1 &= b^2/a^2 \\
s_2 &= -b/ga \\
s_3 &= 1/g \\
\rho &= (ac - 2b^2)/(b^2 - ad)
\end{aligned}
\tag{2.6}
$$

és

$$
\gamma_{i,j} = \begin{cases} \dfrac{\operatorname{sh} i\theta \cdot \operatorname{sh}(n+1-j)\theta}{\operatorname{sh}\theta \cdot \operatorname{sh}(n+1)\theta}, & i \le j \\ \dfrac{\operatorname{sh} j\theta \cdot \operatorname{sh}(n+1-i)\theta}{\operatorname{sh}\theta \cdot \operatorname{sh}(n+1)\theta}, & i \ge j \end{cases}
\tag{2.7}
$$

ahol

$$
\theta = \operatorname{arch}(\rho/2)
\tag{2.8}
$$

Az $\mathbf{X}_1^{-1}$ mátrix előállítását az 1. mellékletben közöljük. Vegyük észre, hogy a mátrix az alábbi szerkezetű:

- szimmetrikus;
- a p paraméter csak a főátlóban, minden második sorban szerepel;
- az első és az utolsó sortól eltekintve $\mathbf{X}_1^{-1}$ elemeinek kiszámításához egy sorban legfeljebb kettő, illetve négy $\gamma_{i,j}$ érték szükséges, és ez sorról-sorra változik.

Látható, hogy a kvadratikus elemekkel történő közelítés esetén azt a legkézenfekvőbb elégséges feltételrendszert, amelyet a lineáris elemekkel való közelítésnél megadtunk — t.i. hogy az $\mathbf{X}_1$ egy M–mátrix és $\mathbf{X}_2$ nemnegatív mátrix — nem tudjuk kielégíteni, mert q és γ valamennyi értéke mellett a z elem nemnegatív, így az $\mathbf{X}_2 \ge 0$ feltételt nem tudjuk biztosítani.

Igaz azonban az alábbi

2.1. LEMMA. *Az $\mathbf{X} := \mathbf{X}_1^{-1}\mathbf{X}_2$ mátrix nemnegativitásának szükséges feltétele, hogy az $\mathbf{X}_1$ mátrix monoton legyen.*

Bizonyítás. Rövid számolással belátható, hogy a (2.6)-ban definiált g–re érvényes

$$
g = (b^2 - ad)/a = ((1 - 8q\gamma)^2 + 176q^2\gamma^2)/90a,
$$

ami nyilvánvalóan pozitív szám. Ebből következik, hogy s_1 és s_3 nemnegatív.

Az s_2 szám előjelének vizsgálatához tekintsük — például — az $(\mathbf{X})_{1,3}$ elemet:

$$s_2\gamma_{1,1}v + s_1(\gamma_{1,1} + \gamma_{1,2})y + s_2\gamma_{1,2}v \geq 0. \tag{2.9}$$

Tegyük fel, hogy $b \geq 0$. Ekkor azt kapjuk, hogy

$$yb \geq av \tag{2.10}$$

Részletesen kiírva:

$$16(1/30 - q(1-\gamma)/3)) \cdot 2(1/30 - 4q\gamma/3) \geq \\ \geq 16(1/30 + q\gamma/3) \cdot 2(1/30 + 4q(1-\gamma)), \tag{2.11}$$

amiből a számításokat elvégezve

$$0 \geq 1 \tag{2.12}$$

ellentmondásra jutunk.

Tehát szükséges a $b < 0$ feltétel, ami egyenértékű a

$$q > 1/(40\gamma) \tag{2.13}$$

feltétellel.

Mivel $b < 0$ esetén $s_2 > 0$ (lásd 2.6), ezért $\mathbf{X}_1^{-1}$ előállításából (lásd 1. számú melléklet) nyilvánvaló, hogy a (2.13) feltétel teljesülése esetén $\mathbf{X}_1^{-1}$ minden eleme nemnegatív.

Vegyük észre, hogy a $b < 0$ feltétel $\gamma = 0$ esetben nem teljesíthető, így igaz a

2.2. LEMMA. *Ha az* (I/2.1)–(I/2.3) *feladat megoldását az* (I/2.22)–(I/2.23) *sémával számítjuk, akkor a megoldás nemnegativitása másodrendű bázisfüggvényeket alkalmazva* $\gamma = 0$ *esetben nem érhető el.*

A továbbiakban $\mathbf{X}$ elemeinek előjelvizsgálatához $\mathbf{X}_1^{-1}$ és $\mathbf{X}_2$ speciális szerkezete miatt vezessük be az alábbi elnevezéseket:

$$(X)_{k,\ell} \begin{cases} \text{„nemteljes elem”}, & \text{ha } \ell \in \{1, 2, 2n, 2n+1\} \\ \text{„teljes elem”}, & \text{ha } \ell \in \{3, 4, \ldots, 2n-1\} \\ \text{„öt-elem”}, & \text{ha } \ell \text{ páros} \\ \text{„három-elem”}, & \text{ha } \ell \text{ páratlan.} \end{cases}$$

(Például $(\mathbf{X})_{1,1}$ nemteljes három-elem, $(\mathbf{X})_{2,4}$ teljes öt-elem. Az elnevezések arra utalnak, hogy $(\mathbf{X})_{k,\ell}$ milyen skalárszorzat eredménye; azaz $\mathbf{X}_2$ ℓ-edik oszlopa v, y, v

vagy z, v, w, v, z nemzérus elemeket tartalmazza. Nemteljesnek nevezzük pl. az $(\mathbf{X})_{1,1}$ elemet, mert $\mathbf{X}_2$ első oszlopában y, v áll.)

$$(X)_{k,\ell}\begin{cases}\text{„kettő-sor" eleme,} & \text{ha } k \text{ páros vagy 1 vagy } 2n+1\\ \text{„négy–sor" eleme,} & \text{egyébként.}\end{cases}$$

(Az elnevezések arra utalnak, hogy $\mathbf{X}_1^{-1}$ megfelelő sorában az elemek legfeljebb kettő vagy négy $\gamma_{i,j}$ érték összegével adhatók meg.)

Nyilvánvaló, hogy $\mathbf{X}$ szimmetrikussága miatt elegendő — például — a felső triangulust vizsgálni.

A továbbiakban vizsgáljuk meg az egyes elemek nemnegativitásának feltételét. Először a teljes elemeket majd a nemteljes elemeket tekintjük. Külön kell vizsgálnunk a főátló elemeit, mert $(\mathbf{X})_{i,i}$ számításakor $\gamma_{i,j}$ értéke (2.7) képletek szerint változik.

– Teljes három-elem, kettő-sor

Vegyük észre, hogy az első és az utolsó sort ($k = 1$ és $k = 2n + 1$) nem kell külön tárgyalnunk, mivel

$$s_1 = -\frac{bs_2}{a} \quad \text{és} \quad s_2 = -\frac{bs_3}{a}$$

Így $(\mathbf{X})_{1,\ell} \geq 0$ pontosan akkor teljesül, amikor $(\mathbf{X})_{2,\ell} \geq 0$; és ugyanígy $(\mathbf{X})_{2n,\ell} \geq 0$ pontosan akkor igaz, amikor $(\mathbf{X})_{2n+1,\ell} \geq 0$. (A $b < 0$ szükséges feltétel teljesül.)

A vizsgált elem értéke részletesen kiírva:

$$s_2\gamma_{i,j}v + s_1(\gamma_{i,j} + \gamma_{i,j+1})y + s_2\gamma_{i,j+1}v \geq 0 \tag{2.14}$$

azaz

$$av - yb \geq 0 \tag{2.15}$$

Ez megegyezik a (2.13) feltétellel.

A számításokat a többi elemtípusra hasonlóan végezzük el. Elemi, de fáradságos úton belátható, hogy a (2.13) feltétel azonos az $\mathbf{X}$ mátrix következő, főátlón kívüli elemeinek nemnegativitási feltételével:

- teljes három-elem, négy-sor,
- teljes öt-elem, négy-sor,
- nemteljes elemek.

Hasonlóan igazolható továbbá, hogy a teljes öt-elem, kettő-sor nemnegativitási feltétele

$$q \geq 7/60\gamma \tag{2.16}$$

azaz, ha (2.13) teljesül, akkor (2.16) is igaz.

– A főátló elemei

Közvetlen számolással belátható, hogy a főátlóban a következő szimmetria érvényesül:

$$(\mathbf{X})_{1,1} = (\mathbf{X})_{2n+1,2n+1}, \quad (\mathbf{X})_{2,2} = (\mathbf{X})_{2n,2n} \cdots \text{ azaz}$$

$$(\mathbf{X})_{i,i} = (\mathbf{X})_{2n+2-i,2n+2-i} \cdot \tag{2.17}$$

Az elemek kifejtésének összehasonlításával igazolható továbbá, hogy ha $(\mathbf{X})_{1,1} \geq 0$, akkor $(\mathbf{X})_{3,3}, (\mathbf{X})_{5,5} \ldots$ is nemnegatív; illetve $(\mathbf{X})_{2,2} \geq 0$ esetén $(\mathbf{X})_{4,4}, (\mathbf{X})_{6,6} \ldots$ is nemnegatív.

– $(\mathbf{X})_{2,2}$ nemnegativitása

$$s_2 v(2\gamma_{1,1} + \gamma_{1,2}) + s_4(w\gamma_{1,1} + z\gamma_{1,2}) \geq 0. \tag{2.18}$$

Ha $\gamma_{1,1}$ és $\gamma_{1,2}$ értékét behelyettesítjük, akkor

$$\frac{\operatorname{sh}(n-1)\theta}{\operatorname{sh} n\theta}(az - bv) + aw - 2bv \geq 0. \tag{2.19}$$

Megjegyzés. Bár a fenti feltétel szükséges és elégséges, a korábbi feltételekkel szemben nehezen kezelhető, mert:

- a q és a γ paraméteren túl n értékét is tartalmazza;
- explicit egyenlőtlenséget q és γ között $\gamma_{i,j}$ szerkezete miatt nem tudunk megadni.

– $(\mathbf{X})_{1,1}$ nemnegativitása

$$s_1 y(\gamma_{1,1} + p) + s_2 v\gamma_{1,1} \geq 0. \tag{2.20}$$

Felhasználva p és $\gamma_{1,1}$ értékét:

$$\frac{\operatorname{sh} n\theta}{\operatorname{sh}(n+1)\theta}(bva - yb^2) - yad - b^2 y \geq 0. \tag{2.21}$$

A (2.21) feltétel (2.19)-hez hasonló struktúrájú.

Emiatt a fenti feltételek teljesülését numerikusan vizsgáltuk.

Eljárásunk a következő volt:

- rögzítettük n és γ értékét;

- q értékét $\Delta q = 0,01$; $0,001$-es lépésközzel a (2.16) feltételből számítható határtól indulva $q = 25$ értékig növeltük (a felső határt az alább a/ pontban megadott monotonitás indokolta);
- megállapítottuk $(\mathbf{X})_{1,1}$ és $(\mathbf{X})_{2,2}$ előjelét;
- az eljárást megismételtük az $n = 2, 3, \ldots, 50$ és a $\gamma = 0,1$, $0,101$, $0,102$, ..., $0,999$ értékekre.

Számításainkat tíz decimális jegy pontossággal végeztük, és a következőket tapasztaltuk:

Rögzített n és γ mellett tekintsük a

$$q \mapsto (\mathbf{X})_{1,1}(q)$$

és

$$q \mapsto (\mathbf{X})_{2,2}(q); \quad q \in \mathbb{R}^+$$

függvényeket. Ezek a függvények a következő tulajdonságokkal rendelkeznek:

a/ kezdetben szigorúan monoton nőnek, majd szigorúan monoton csökkennek;

b/ egyetlen q_n^*-gal jelölt zérushelyük van.

Tekintsük a továbbiakban az

$$n \mapsto q_n^*$$

sorozatot. Erre teljesül, hogy

c/ szigorúan monoton nő;

d/ felülről korlátos ($n > 20$ esetén a c/ pontban említett szigorú monotonitása csak a hatodik tizedesjegytől mutatható ki.)

Vegyük észre, hogy a (2.19) és (2.21) feltételpár a lineáris bázisfüggvényekre vonatkozó (I/2.42) feltétellel hasonló szerkezetű; n növelésével egyre nagyobb felső határt ad q-ra nézve (rögzített γ mellett). Lineáris bázisfüggvények esetében azonban csak az $(\mathbf{X})_{1,1}$ elem korlátozza q növelését, másodrendű esetben pedig $(\mathbf{X})_{1,1}$ és $(\mathbf{X})_{2,2}$.

A (2.19) és (2.21) egyenlőtlenségekkel meghatározott paraméterhatárokat a 2. mellékletben közöljük.

Az eredményeink alapján megállapíthatjuk, hogy rögzített γ esetén q megválasztását a (2.13) feltétel alulról, a (2.19), illetve (2.21) feltétel pedig felülről korlátozza.

Részben elméleti megfontolásokkal, részben numerikus tapasztalatokkal támasztható alá tehát a következő eredmény:

SZÜKSÉGES ÉS ELÉGSÉGES FELTÉTEL MÁSODRENDŰ BÁZISFÜGGVÉNYEKRE. *Az (I/2.1)–(I/2.3) feladat végeselemes megoldásakor a q és a γ paramétert úgy kell megválasztanunk, hogy a (2.13), (2.19) és (2.21) egyenlőtlenségek teljesüljenek: így a megoldás nemnegativitása biztosítható.*

3. A véges elemes differencia-sémák maximum elve és időbeli monotonitása

Mint azt az első rész első fejezetében megmutattuk, a folytonos feladat u megoldásából származtatott $t \mapsto \|u(\cdot, t)\|$ függvény (ahol $\|\cdot\|$ az $L_2(0, L)$ vagy $C(0, L)$ normák egyikét jelenti) a legnagyobb értékét a $t = 0$ helyen veszi fel; valamint, hogy ez a függvény monoton csökken. A továbbiakban megvizsgáljuk, hogy ezek a tulajdonságok milyen feltételek mellett öröklődnek át a véges elemes numerikus megoldásra. Nevezetesen, olyan sémákat keresünk, amelyekre egyrészt teljesül a maximumelv, másrészt valamivel több is: a numerikus megoldás rögzített időrétegen vett normájának értéke az időrétegek szerint monoton csökkenő. A továbbiakban ezt a tulajdonságot időbeli monotonitásnak nevezzük, megjegyezve, hogy helyenként szokásos norma szerinti kontraktivitásnak is nevezni.

A véges differencia sémákra vonatkozó eredmények megtalálhatók Samarskii [17] és Stoyan [20] munkáiban. Terjedelmi okokból mi nem térünk ki részletesen erre a kérdésre.

Vezessük be a j-edik időrétegbeli véges elemes numerikus megoldásra az $u_h^j(x)$ jelölést, azaz

$$u_h^j(x) = \sum_{i=1}^{n} \alpha_i^j \phi_i^n(x), \tag{3.1}$$

ahol α_i^j az (I/2.20), (I/2.21) Cauchy feladat α^j megoldásvektorának i-edik komponense, $\phi_i^n(x)$ pedig a lineáris bázisfüggvény. Ekkor a fenti tulajdonságok az

$$\|u_h^{j+1}\| \leq |u_h^j\| \tag{3.2}$$

egyenlőtlenséget jelentik.

Először az L_2-beli esetet vizsgáljuk.

3.1. TÉTEL. *Az (I/2.1)–(I/2.3) feladat L_2-normában stabil véges differenciás illetve véges elemes numerikus módszerei L_2-normában időben monoton csökkenőek is.*

Bizonyítás. Tekintsük u_h^j spektrális alakban történő előállítását [22]. Jelölje λ_i $(i = 1, 2, \ldots n)$ az $M^{-1}Q$ mátrix sajátértékeit, $s_i(x)$ $(i = 1, 2, \ldots, n)$ pedig az ortonormált sajátvektor-rendszerét. Ekkor

$$u_h^j(x) = \sum_{i=1}^{n} (u_h^j, s_i) \cdot s_i(x), \tag{3.3}$$

ahol (u_h^j, s_i) az $L_2(0, L)$-beli skaláris szorzatot jelöli. Ekkor

$$u_h^{j+1}(x) = \sum_{i=1}^{n} r_\gamma(\tau\lambda_i)(u_h^j, s_i) \cdot s_i(x), \tag{3.4}$$

ahol

$$r_\gamma(\lambda) := (1+\gamma\lambda)^{-1}(1-(1-\gamma)\lambda) \tag{3.5}$$

az (I/2.22) egyparaméteres módszernek megfelelő függvény. Mivel az L_2-beli stabilitás jól ismert feltétele a

$$\max_{1\le i\le n} |r(\tau\lambda_i)| \le 1, \tag{3.6}$$

összefüggés [17], ezért a Bessel-egyenlőtlenség alkalmazásával az

$$\|u_h^{j+1}\|^2 \le \sum_{i=1}^{n} |r(\tau\lambda_i)(u_h^j, s_i)|^2 \le \sum_{i=1}^{n} |(u_h^j, s_i)|^2 \le \|u_h^j\|^2 \tag{3.7}$$

L_2–norma szerinti becslés nyerhető, ami a tétel állítását jelenti.

Megjegyezzük, hogy a fenti tétel az (I/2.1)–(I/2.3) típusú, időben nem korlátos tartományon kitűzött differenciál-egyenletek numerikus sémáira vonatkozik. Ha a tartomány az időváltozó szerint is korlátos, akkor (3.6) már nem szükséges feltétele az L_2-stabilitásnak; a séma stabil marad

$$\max_{1\le i\le n} |r(\tau\lambda_i)| \le 1 + 0(\tau) \tag{3.6a}$$

esetén is [15]. Ekkor viszont az időbeli monotonitás nem feltétlenül teljesül.

A maximum normabeli becslés előtt tekintsük az

$$\begin{gathered} A_i y_{i-1} - C_i y_i + B_i y_{i+1} = -F_i; \quad i = 1, 2, \ldots, n-1 \\ y_0 = y_n = 0 \end{gathered} \tag{3.8}$$

alakú lineáris algebrai egyenletrendszert, ahol az együtthatókra teljesülnek az alábbi feltételek:

$$|A_i| > 0, \ |B_i| > 0, \ |C_i| - |A_i| - |B_i| > 0. \tag{3.9}$$

Ekkor a megoldásvektorra érvényes az

$$\|y\|_C \le \left\| \frac{F}{D} \right\|_C \tag{3.10}$$

egyenlőtlenség [17], ahol y és $\frac{F}{D}$ jelöli az y_i és $\frac{F_i}{D_i}$ (ahol $D_i = C_i - A_i - B_i$) komponensű vektorokat, továbbá valamely $w \in \mathbb{R}^n$ vektor esetén

$$\|w\|_C = \max_{1\le i\le n} |w_i| \tag{3.11}$$

3.2. TÉTEL. *Tegyük fel, hogy*

$$q \leq \frac{1}{3(1-\gamma)}. \tag{3.12}$$

Ekkor a véges elemes numerikus sémára (3.2) *a maximum normában érvényes.*

Bizonyítás. Tekintsük a véges elemes numerikus sémát. Ekkor időrétegenként egy (3.8) alakú lineáris algebrai egyenletrendszert kapunk az

$$A_i = B_i = -\frac{h}{6} + \frac{\tau\gamma}{h},\ C_i = \frac{2h}{3} + \frac{2\tau\gamma}{h} \tag{3.13}$$

$$F_i^j = \left(\frac{h}{6} + \frac{\tau(1-\gamma)}{h}\right)\alpha_{i-1}^j + \left(\frac{2h}{3} - \frac{2\tau(1-\gamma)}{h}\right)\alpha_i^j + \left(\frac{h}{6} + \frac{\tau(1-\gamma)}{h}\right)\alpha_{i+1}^j$$

értékekkel.

Ha $A_i = B_i = 0$, akkor $\mathbf{X}_1 = h\mathbf{E}$ és ekkor az állítás nyilvánvalóan igaz.

Tegyük fel most, hogy $|A_i| = |B_i| \neq 0$ és alkalmazzuk a (3.10) becslést a feladatra! Könnyen látható, hogy ekkor a (3.12) feltétel mellett (3.9) teljesül. Mivel $D_i = h$, ezért

$$\|\alpha^{j+1}\|_C \leq h^{-1}\|F^j\|_C. \tag{3.14}$$

Másrészt, ugyancsak (3.12) következtében

$$\frac{2h}{3} - \frac{2\tau(1-\gamma)}{h} \geq 0, \tag{3.15}$$

ezért

$$|F_i^j| \leq h \max_{1\leq i\leq n} |\alpha_i^j|. \tag{3.16}$$

Nyilvánvalóan (3.14) és (3.16) együttesen az

$$\|\alpha^{j+1}\|_C \leq \|\alpha^j\|_C \tag{3.17}$$

egyenlőtlenséget adja. Figyelembevéve, hogy

$$u_h^j(x_i) = \alpha_i^j, \tag{3.18}$$

ezért a (3.17) becslés a

$$\max_{1\leq i\leq n} |u_h^{j+1}(x_i)| \leq \max_{1\leq i\leq n} |u_h^j(x_i)| \tag{3.18}$$

összefüggést jelenti. Így a rácshálón a maximumelv teljesül. Végezetül, mivel $u_h^k(x)$ szakaszonként lineáris, ezért a maximumát csak valamely csomópontban veheti fel. Mindez a tétel állítását igazolja.

Vegyük észre, hogy a tételben szereplő (3.12) feltétel megegyezik a véges elemes séma nemnegativitására vonatkozó

$$\frac{1}{6\gamma} \leq q \leq \frac{1}{3(1-\gamma)} \tag{3.19}$$

feltétel felső korlátjával. Ezért a nemnegativitási feltétel mellett a véges elemes séma pozitív értékeken keresztül időben monoton a maximum normában.

4. A numerikus megoldás oszcillációmentessége

Mint arra már az első rész fejezetében (1.4. Következmény) utaltunk, a vizsgált parabolikus probléma megoldása végtelen időtartomány esetén oszcillációmentes. A továbbiakban megvizsgáljuk, hogy a numerikus megoldás mely feltételek mellett őrzi meg ezt a tulajdonságot.

Először definiáljuk a numerikus megoldás oszcillációmentességét! Jelölje $y^j = [y_1^j, y_2^j, \ldots, y_n^j]$ a numerikus megoldást a j-edik időrétegen.

Definíció. Azt mondjuk, hogy a numerikus megoldás (időben) oszcillációmentes, ha léteznek olyan N_i $(i = 1, 2, \ldots, n)$ természetes számok, hogy minden rögzített i $(i = 1, 2, \ldots, n)$ esetén az (y_i^j) $(j = N_i, N_i + 1, \ldots)$ monoton számsorozatok.

Értelemszerűen, egy numerikus megoldást oszcillálónak nevezünk, ha nem oszcillációmentes.

Tekintsük a

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}; \quad x \in (0, L), \ t > 0 \tag{4.1}$$

$$u(0,t) = u(L,t) = 0 \quad t \geq 0, \tag{4.2}$$

$$u(x,0) = u_0(x), \ x \in (0, L) \tag{4.3}$$

feladatot. Lineáris véges elemekkel, illetve az egylépéses módszerrel diszkretizálva újra a jól ismert

$$\left(\mathbf{M} + \tau\gamma\mathbf{Q}\right)\alpha^{j+1} = \left(\mathbf{M} - \tau(1-\gamma)\mathbf{Q}\right)\alpha^j \tag{4.4}$$

feladatot nyerjük, ahol $\mathbf{M}$ és $\mathbf{Q}$ a már ismert, szimmetrikus, pozitív definit, $n \times n$ dimenziós mátrixok.

Jelölje a

$$\lambda \cdot M\alpha + Q\alpha = 0 \tag{4.5}$$

sajátérték-feladat megoldását $\lambda_1, \lambda_2, \ldots, \lambda_n$ és $v_1, v_2, \ldots, v_n$. Mint ismeretes [16], a (v_i) vektorrendszer lineárisan független, valamint a λ_i sajátértékek valósak és negatívak. Így az α^{j+1} és az α^j vektorok előállíthatók az

$$\alpha^{j+1} = \sum_{m=1}^{n} y_m^{j+1} v_m; \quad \alpha^j = \sum_{m=1}^{n} y_m^j v_m \tag{4.6}$$

alakban, ahol az y_m^j és y_m^{j+1} $(m = 1, 2, 3, \ldots, n)$ számokra a (4.6) kifejezés (4.4) egyenletbe történő behelyettesítésével és a (4.5) felhasználásával az

$$y_m^{j+1} = \frac{\frac{1}{\tau} + (1-\gamma)\lambda_m}{\frac{1}{\tau} - \gamma \cdot \lambda_m} y_m^j \tag{4.7}$$

rekurziót kapjuk. Így közvetlenül megadható az L_2-beli stabilitási tartományon belül az oszcillációmentesség egy elégséges feltétele:

$$0 < \frac{\frac{1}{\tau} + (1-\gamma)\lambda_m}{\frac{1}{\tau} - \gamma \cdot \lambda_m} < 1; \quad m = 1, 2, \ldots, n. \tag{4.8}$$

Ez a

$$\begin{aligned} &\tau < \frac{1}{(1-\gamma)\,|\lambda_{\max}|}, && \text{ha } \gamma \in [0,1), \\ &\tau \text{ tetszőleges}, && \text{ha } \gamma = 1 \end{aligned} \tag{4.10}$$

feltételt jelenti, ahol $|\lambda_{\max}| = \max\limits_{1 \le i \le n} |\lambda_i|$.

Nyilvánvalóan a (4.10) becslés $|\lambda_{\max}|$ értékétől függ. Mivel az M és Q mátrixok egyenletesen kontinuáns, szimetrikus mátrixok és sajátértékeik pozitívak, ezért [16]

$$|\lambda_{\max}| = |\lambda_n| = \frac{4\cos^2(h/2)}{h^2\left(1 - \frac{2}{3}\cos^2(h/2)\right)}\,. \tag{4.11}$$

A fenti eredményeket összegezve a következő állítást nyerjük.

4.1. TÉTEL. *Tegyük fel, hogy a valamely rögzített rácsháló paramétereire adott γ esetén teljesül a*

$$\begin{aligned} &q < \frac{1 - \frac{2}{3}\cos^2(h/2)}{4(1-\gamma)\cos^2(h/2)}, && \textit{ha } \gamma \in [0,1), \\ &\tau \textit{ tetszőleges}, && \textit{ha } \gamma = 1. \end{aligned} \tag{4.12}$$

feltétel. Ekkor a véges elemes séma oszcillációmentes.

A fenti becslést hasonlítsuk össze a [24]-ben szereplő becsléssel. Itt a maximális abszolút értékű sajátérték becslésére a következő, bizonyítás nélkül szereplő állítást

alakalmazzák: a globális rendszer abszolút értékben maximális sajátértéke kisebb, mint a lokális rendszerek maximális abszolút értékű sajátértékei. Mivel a felosztásuk ekvidisztáns, ezért az egyes lokális rendszerek megegyeznek és egyszerű számolással a

$$|\lambda_{\max}| < \frac{12}{h^2} \tag{4.13}$$

becslést kapjuk. Ez a $\gamma \in [0, 1)$ esetén a

$$q < \frac{1}{12(1-\gamma)} \tag{4.14}$$

feltételt jelenti. Vegyük észre, hogy a (4.12) feltétel az osztásrészek számának növelésével (azaz h nullához tartásával) felülről tart ezen korláthoz. Következésképpen, minden rögzített rácshálón (4.12) nagyobb korlátot jelent.

Megjegyzés. Ha a (4.2) peremfeltételben az L pontbeli első (Dirichlet-féle) peremfeltételt felcseréljük a második (Neumann-féle) peremfeltételre, akkor a

$$|\lambda_{\max}| < \frac{\pi^2}{L^2}\left(\frac{12}{h^2} - C_1(h)\right), \tag{4.15}$$

becslés kapható [8], ahol

$$C_1(h) = \frac{48 - h^2}{64}\,. \tag{4.16}$$

A fenti sajátérték-becslésből a feladat numerikus megoldásának oszcillációmentességére vonatkozó elégséges feltétel a (4.10) feltétel alapján már közvetlenül megadható.

Köszönetnyilvánítás. A szerzők köszönetüket fejezik ki RÓZSA PÁLnak a másodrendű közelítésekkel kapcsolatos algebrai problémák megoldásában, illetve STOYAN GISBERTnek a numerikus nemnegativitás kérdéskörében nyújtott segítségéért. Megköszönik továbbá TÓTH JÁNOSnak a cikk elkészítéséhez adott hasznos tanácsait.

1. melléklet

Az $\mathbf{X}_1^{-1}$ mátrix elemeit $\mathbf{W}$ és $\mathbf{S}$ mátrixokkal adjuk meg:

$$(\mathbf{X}_1^{-1})_{i,j} = (\mathbf{W})_{i,j} * (\mathbf{S})_{i,j} \ .$$

A $\mathbf{W}$ mátrix elemei (bal oldal):

$$\begin{bmatrix}
\gamma_{1,1}+p & \gamma_{1,1} & \gamma_{1,1}+\gamma_{1,2} & \gamma_{1,2} & \gamma_{1,2}+\gamma_{1,3} \\
\gamma_{1,1} & \gamma_{1,1} & \gamma_{1,1}+\gamma_{1,2} & \gamma_{1,2} & \gamma_{1,2}+\gamma_{1,3} \\
\gamma_{1,1}+ & \gamma_{1,1}+ & \gamma_{1,1}+\gamma_{2,1}+ & \gamma_{1,2}+ & \gamma_{1,1}+\gamma_{2,2}+ \\
+\gamma_{2,1} & +\gamma_{2,1} & +\gamma_{1,2}+\gamma_{2,2}+p & +\gamma_{2,2} & +\gamma_{1,3}+\gamma_{2,3} \\
\dots & \dots & \dots & \dots & \dots \\
\gamma_{i,1} & \gamma_{i,1} & \gamma_{i,1}+\gamma_{i,2} & & \\
\gamma_{i,1}+ & \gamma_{i,1}+ & \gamma_{i,1}+\gamma_{i,2}+ & \dots & \gamma_{i,i} \\
+\gamma_{i+1,1} & +\gamma_{i+1,1} & +\gamma_{i+1,1}+\gamma_{i+1,2} & & \\
\dots & \dots & \dots & \dots & \dots \\
\gamma_{n-1,1}+ & \gamma_{n-1,1}+ & \gamma_{n-1,1}+\gamma_{n-1,2}+ & \gamma_{n-1,2}+ & \gamma_{n-1,2}+\gamma_{n-1,3}+ \\
+\gamma_{n-1,2} & \gamma_{n-1,2} & +\gamma_{n,1}+\gamma_{n,2} & +\gamma_{n,2} & +\gamma_{n,2}+\gamma_{n,3} \\
\gamma_{n,1} & \gamma_{n,1} & \gamma_{n,1}+\gamma_{n,2} & \gamma_{n,2} & \gamma_{n,2}+\gamma_{n,3} \\
\gamma_{n,1} & \gamma_{n,1} & \gamma_{n,1}+\gamma_{n,2} & \gamma_{n,2} & \gamma_{n,2}+\gamma_{n,3}
\end{bmatrix}$$

A $\mathbf{W}$ mátrix elemei (jobb oldal):

$$\begin{bmatrix}
\gamma_{1,3} & \dots & \gamma_{1,n-1}+\gamma_{1,n} & \gamma_{1,n} & \gamma_{1,n} \\
\gamma_{1,3} & \dots & \gamma_{1,n-1}+\gamma_{1,n} & \gamma_{1,n} & \gamma_{1,n} \\
\gamma_{1,2}+\gamma_{2,2} & \dots & \gamma_{1,n-1}+\gamma_{1,n}+ & \gamma_{1,n}+ & \gamma_{1,2}+ \\
 & & +\gamma_{2,n-1}+\gamma_{2,n} & +\gamma_{2,n} & +\gamma_{2,n} \\
\dots & \dots & \dots & & \dots \\
\dots & & \gamma_{i,n-1}+\gamma_{i,n} & \gamma_{i,n} & \gamma_{i,n} \\
\gamma_{i,i}+\gamma_{i+1,1}+\gamma_{i,i+1}+ & \dots & \gamma_{i,n-1}+\gamma_{i,n}+ & \gamma_{i,n}+ & \gamma_{i,n}+ \\
+\gamma_{i+1,i+1}+p & \dots & +\gamma_{i+1,n-1}+\gamma_{i+1,n} & +\gamma_{i+1,n} & \gamma_{i+1,n} \\
\dots & \dots & \dots & & \dots \\
\gamma_{n-1,3}+ & \dots & \dots & & \dots \\
+\gamma_{n,3} & & & & \\
\gamma_{n,3} & \dots & \gamma_{n,n}+\gamma_{n,n-1} & \gamma_{n,n} & \gamma_{n,n} \\
\gamma_{n,3} & \dots & \gamma_{n,n}+\gamma_{n,n-1} & \gamma_{n,n} & \gamma_{n,n}+p
\end{bmatrix}$$

Az S mátrix elemei:

$$(\mathbf{S})_{i,j} = \begin{cases} s_1, & \text{ha } i \text{ páratlan és } j \text{ páratlan} \\ s_2, & \text{ha } i \text{ páratlan és } j \text{ páros} \\ s_2, & \text{ha } i \text{ páros és } j \text{ páratlan} \\ s_3, & \text{ha } i \text{ páros és } j \text{ páros.} \end{cases}$$

2. melléklet

$(\mathbf{X})_{1,1}$ és $(\mathbf{X})_{2,2}$ nemnegativitásának paraméterhatára

	$(\mathbf{X})_{1,1}$		$(\mathbf{X})_{2,2}$	
γ	$q_{\max}$ $(n=2)$	$q_{\max}$ $(n=20)$	$q_{\max}$ $(n=2)$	$q_{\max}$ $(n=20)$
0,6	0,34018	0,34018	0,21515	0,21515
0,65	0,40559	0,40567	0,27257	0,27259
0,7	0,49483	0,49506	0,35581	0,35599
0,75	0,62276	0,62357	0,48115	0,48257
0,8	0,81946	0,82219	0,68261	0,68739
0,85	1,15532	1,16486	1,03453	1,05492
0,9	1,84167	1,87919	1,76083	1,84841
0,91	2,07265	2,12335	2,00543	2,12452
0,92	2,36225	2,43180	2,31195	2,47566
0,93	2,73560	2,83284	2,70684	2,93545
0,94	3,23459	3,37404	3,23423	3,56040
0,95	3,93465	4,14168	3,97344	4,45351
0,96	4,98656	5,30980	5,08325	5,82350
0,97	6,74219	7,28847	6,93395	8,16435
0,98	10,25710	11,32171	10,63674	12,98287
0,99	20,80908	23,72024	21,74710	27,97545

IRODALOM

[1] BABUSKA, I., JANIK, T., „The $h-p$ version of the finite element method for parabolic equations", Part I: The p-version in time, *Numer. Method for Partial Diff. Eqs.* **5** (1989), 363–369.

[2] BABUSKA, I., JANIK, T., „The $h-p$ version of the finite element method for parabolic equations", Part II: The $h-p$-version in time, *Numer. Method for Partial Diff. Eqs.* **6** (1990), 343–369.

[3] BERS, L., JOHN, F., SCHECHTER, M., *Partial Differntial Equations* (Interscience Publisher, New York, 1964).

[4] FARAGÓ, I., „Véges elemek módszere lineáris, parabolikus típusú feladatok megoldására", *Alkalmazott Mat. Lapok* **11** (1985), 123–155.

[5] FARAGÓ, I., KOMÁROMI, N., „Nonegativity of the numerical solution of parabolic problems", *Numerical Methods, North-Holland* **59** (1990), 173–179 (Greenspan, Rózsa, ed.).

[6] FARAGÓ, I., „Positivity of the finite difference scheme for the linear parabolic problems", *Differential equations, North-Holland* **62** (1991), 113–118 (Farkas, M. and Sebestyén, Z., eds.).
[7] FRIEDMAN, A., *Partial Differential Equations of Parabolic Type* (Prentice-Hall, Englewood Cliffs, 1964).
[8] HAROTEN, H. A., „Nonnegativity of the numerical solution of parabolic problems with differential boundary conditions", *Annales Universitatis Sci. Budapest, Sectio Computorica*, (megjelenés alatt).
[9] HAROTEN, H. A., „Some condition of ρ-stability and the non-oscillation of the numerical solution of linear parabolic problem", *Annales Universitatis, Sci. Budapest, Sectio Computorica*, (megjelenés alatt).
[10] LADYZENSKAJA, O. A., *Lineáris és kvázilineáris parabolikus egyenletek* (Nauka, Moszkva, 1967), (oroszul).
[11] LORENZ, J., „Zur Inversmonotonie diskreter Probleme", *Numer. Math.* **27** (1977), 227–238.
[12] MARCSUK, G. I., *A gépi matematika numerikus módszerei* (Műszaki Könyvkiadó, Budapest, 1976).
[13] PFEIL, T., „Time-Monotonicity of the Solutions of Linear Second Order Homogeneous Parabolic Equations", *Ann. Univ. Sci. Budapest, Eötvös Sect. Math.* **36** (1993), 139–146.
[14] PFEIL, T., „On the Time-Monotononicity of the Solutions of Linear Second Order Homogeneous parabolic Equations on the Whole Domain", *Publ. Math. Debrecen* **45** (1994), (közlésre elfogadva).
[15] RICHTMAYER, R., MORTON, K. W., *Difference methods for initial-value problems* (J. Wiley Publ., New York, 1967).
[16] RÓZSA, P., *Lineáris algebra és alkalmazásai* (Műszaki Kiadó, Budapest, 1976).
[17] SAMARSKII, A. A., *Differnciasémák elmélete* (Nauka, Moszkva, 1977), (oroszul).
[18] SIMON, L., BADERKO, E. A., *Másodrendű lineáris parciális differenciálegyenletek* (Tankönyvkiadó, Budapest, 1983).
[19] SMOLLER, J., *Shock Waves and Reaction-Diffusion Equations* (Springer-Verlag, Berlin, 1983).
[20] STOYAN, G., „On a Maximum Principle for Matrices and on Conservation of Monotonicity", *ZAMM* **62** (1982), 375–381.
[21] SZABÓ, B. A., BABUSKA, I., *Finite Element Analysis* (Wiley, New York, 1991).
[22] THOMÈE, V., „Finite Difference Methods for Linear Parabolic Equations", in Handbook of Numerical Analysis (Ciarlet, P. G. and Lions, J. L., eds.), vol. 1 (Nort-Holland, Amsterdam, 1990).
[23] YOUNG, D. M., *Nagy lineáris rendszerek iterációs megoldása* (Műszaki Könyvkiadó, Budapest, 1979).
[24] ZIENKIEWICZ, O. C., MORGAN. K., *Finite Elements and Approximation* (Wiley, New York, 1983).

(Beérkezett: 1993. február 2.)

FARAGÓ ISTVÁN
HAROTEN A. HARITON
PFEIL TAMÁS
EÖTVÖS LORÁND TUDOMÁNYEGYETEM
ALKALMAZOTT ANALÍZIS TANSZÉK
1088 BUDAPEST
MÚZEUM KRT. 6–8.

KOMÁROMI SÁNDOR
AGRÁRTUDOMÁNYI EGYETEM
AGRÁRMARKETING TANSZÉK
2103 GÖDÖLLŐ
PÁTER KÁROLY U. 1.

THE DIFFERENTIAL EQUATION OF THE HEAT TRANSFER AND QUALITATIVE PROPERTIES ITS NUMERICAL SOLUTIONS: II. THE NONNEGATIVITY OF THE SECOND ORDER APPROXIMATION, THE MAXIMUM PRINCIPLE AND THE NONOSCILLATION

I. FARAGÓ, H. A. HARITON, N. KOMÁROMI AND T. PFEIL

In the second part of the paper – accordingly with the purposes formulating in second part — we examine other qualitative properties. First we give the condition of the nonnegativity of the numerical solution arising by use of quadratic finite element method. Then we formulate the condition of the monotonicity of the numerical solution both in L_2 and maximum norms, too. Finally, we give the condition of the nonoscillation to the numerical schemes getting by using of linear finite element method.

Alkalmazott Matematikai Lapok **17** (1993) 143–169

A SZÁLLÍTÁSI FELADAT SZTOCHASZTIKUS VARIÁNSAI

NAGY TAMÁS

Miskolc

A cikk első részében az entrópia programozási feladatot és annak dualitási problémakörét mutatjuk be. Az entrópia programozási feladat olyan nemnegatív vektor keresése, amelynek egy adott pozitív vektortól való eltérése a lehető legkisebb lineáris feltételek fennállása mellett. A két nemnegatív vektor egymástól való eltérésének mérésére a valószínűségeloszlások egymástól való eltérésének Kullback–Leibler által bevezetett mérőszámának általánosításaként nyert eltérésfüggvényt használjuk. A cikkünk fő részében két alkalmazást mutatunk be. Tekintsük a klasszikus szállítási feladatot, de ne írjuk elő az összes feltétel egyenlőség formájában való teljesülését, hanem bizonyos indexekre ne adjunk szigorú előírást, helyette a két oldal fenti eltérését építsük be a célfüggvénybe úgy, hogy az eredeti célfüggvény és az eltérés súlyozott átlaga minimális legyen. A másik alkalmazásnál, az input-ouput táblák előrebecslési feladatánál is hasonlóan jártunk el. A fenti feladatokra hatékony algoritmusokat dolgoztunk ki.

1. Bevezetés

Az információelméletben és a matematikai statisztikában alapvető szerepet játszik a valószínűségeloszlások egymástól való eltérésének Kullback–Leibler [13, 14] által bevezetett mérőszáma, amely információ-divergencia vagy diszkrimináló információ néven ismert. Ennek általánosításaként nyerhető az alábbi eltérésfüggvény, amely két nemnegatív vektor egymástól való eltérésének mérésére szolgál. Legyen $\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \geq \mathbf{0}$ és $\boldsymbol{y} = (y_1, y_2, \ldots, y_n) \geq \mathbf{0}$ vektor. Az $\boldsymbol{x}$ és az $\boldsymbol{y}$ nemnegatív vektorok eltérésének mérésére az alábbi összefüggést használjuk:

$$\boldsymbol{D}(\boldsymbol{x} \parallel \boldsymbol{y}) := \sum_{j=1}^{n} x_j \log \frac{x_j}{y_j} - \sum_{j=1}^{n} x_j + \sum_{j=1}^{n} y_j \,,$$

amelyet $\boldsymbol{D}$-eltérésnek nevezünk. A fent definiált eltérésre vonatkozóan a következő két fontos tulajdonságot említjük meg:

a) *Nemnegativitási tulajdonság*

Ha $\boldsymbol{x} \geq \mathbf{0}$ és $\boldsymbol{y} \geq \mathbf{0}$, akkor $\boldsymbol{D}(\boldsymbol{x} \parallel \boldsymbol{y}) \geq \mathbf{0}$ és egyenlőség akkor és csak akkor áll fenn, ha $x_j = y_j$ minden j indexre.

b) *Konvexitási tulajdonság*

A $\boldsymbol{D}(\boldsymbol{x} \parallel \boldsymbol{y})$ eltérésfüggvény mind az $\boldsymbol{x}$, mind az y változójában konvex.

1.1. Az entrópia programozási feladat

Legyen $\boldsymbol{A}$ $m \times n$–es mátrix, $\boldsymbol{b}$ m–dimenziós vektor, $\boldsymbol{d}$ n–dimenziós pozitív vektor. Az entrópia programozási feladatnak, illetve a duáljának az alábbi matematikai programozási feladatot nevezzük.

Primál entrópia program:

Meghatározandó azon $\boldsymbol{x} \in \mathbb{R}^{(n)}$ vektor, melyre

$$\boldsymbol{D}(\boldsymbol{x} \,\|\, \boldsymbol{d}) \quad \text{minimális}$$

feltéve, hogy

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$$
$$\boldsymbol{x} \geq \boldsymbol{0}\,.$$

Duál entrópia program:

Meghatározandók azon $\boldsymbol{y} \in \mathbb{R}^{(m)}$ és $\boldsymbol{z} \in \mathbb{R}^{(n)}$ vektorok, amelyekre

$$\boldsymbol{y}\boldsymbol{b} - \sum_{j=1}^{n} \exp(z_j) + \sum_{j=1}^{n} d_j \quad \text{maximális}$$

feltéve, hogy

$$z_j = \boldsymbol{y}\boldsymbol{a}_j + \log(d_j) \qquad (j = 1, \ldots, n)\,,$$

ahol $\boldsymbol{a}_j$ az $\boldsymbol{A}$ mátrix j–edik oszlopvektora.

Tehát a primál entrópia programozási feladat olyan nemnegatív vektor keresése, amelynek egy adott pozitív vektorból való $\boldsymbol{D}$–eltérése a lehető legkisebb lineáris feltételek fennállása mellett. A primál célfüggvény részletesebb kifejtése után az entrópia programozási feladatpár az alábbi:

$$\left.\begin{aligned} \boldsymbol{A}\boldsymbol{x} &= \boldsymbol{b} \\ \boldsymbol{x} &\geq \boldsymbol{0} \end{aligned}\right\} \mathbb{P} \qquad \left.\begin{aligned} z_j &= y\boldsymbol{a}_j + \log(d_j) \\ (j &= 1, \ldots, n) \end{aligned}\right\} \mathbb{D}$$

$$\sum_{j=1}^{n} \boldsymbol{x}_j \log \frac{\boldsymbol{x}_j}{d_j} - \sum_{j=1}^{n} \boldsymbol{x}_j + \sum_{j=1}^{n} d_j \;\; \min! \qquad \boldsymbol{y}\boldsymbol{b} - \sum_{j=1}^{n} \exp(z_j) + \sum_{j=1}^{n} d_j \;\; \max!$$

1.2. Az entrópia programozás dualitási problémaköre

Az alábbi lemma az entrópia programozási feladatpár lehetséges megoldásaihoz tartozó célfüggvények értékeire ad összefüggést.

LEMMA. (Az entrópia programozás alaplemmája).
Ha $x \in \mathbb{P}$ *és* $y, z \in \mathbb{D}$, *akkor*

$$D(x \parallel d) \geq yb - \sum_{j=1}^{n} \exp(z_j) + \sum_{j=1}^{n} d_j$$

és egyenlőség akkor és csak akkor teljesül, ha $x_j = \exp(z_j)$ *minden* j *indexre.*

KÖVETKEZMÉNY. (gyenge equilibrium).
Ha $x^* \in \mathbb{P}$ *és* $y^*, z^* \in \mathbb{D}$ *olyan, hogy a két célfüggvény értéke megegyezik, akkor* x^*; y^*, z^* *optimális megoldások.*

Észrevételek.

i) A primál célfüggvény mindig korlátos alulról.
ii) Ha $\mathbb{P}$ konzisztens, akkor a duál célfüggvény felülről korlátos.
iii) Ha $\mathbb{P}$ nem konzisztens, akkor a duál célfüggvény nem korlátos felülről.

DUALITÁSI TÉTEL.

a) *Ha* $\mathbb{P}$ *konzisztens, akkor létezik olyan* x^* *vektor, hogy*

$$D(x^* \parallel d) = \sup_{y,z} \left\{ yb - \sum_{j=1}^{n} \exp(z_j) + \sum_{j=1}^{n} d_j \right\}.$$

b) *A supremum akkor és csak akkor cserélhető fel a maximummal, ha* $\mathbb{P}$ *Slater konzisztens, azaz ha létezik olyan* x, *amelyre* $Ax = b$, $x > 0$.

KÖVETKEZMÉNY. (erős equilibrium).

i) *Ha* $\hat{x}$; $\hat{y}$, $\hat{z}$ *optimális megoldások, akkor*

$$D(\hat{x} \parallel d) = \hat{y}b - \sum_{j=1}^{n} \exp(\hat{z}_j) + \sum_{j=1}^{n} d_j .$$

ii) *Az entrópia programozási primál feladatnak egyetlen optimális megoldása van.*

Az entrópia programozás Lagrange függvénye

Az entrópia programozási feladat optimális megoldása és a feladathoz tartozó Lagrange függvény nyeregpontja közötti összefüggést fejezi ki a következő tétel. Az entrópia programozási feladat Lagrange függvényének az

$$L(x, y) := D(x \parallel d) + y(b - Ax)$$

függvényt nevezzük, amelyet tetszőleges $x \geq 0$ és y esetén értelmezünk. Az x^*, y^* pontot a Lagrange függvény nyeregpontjának mondjuk, ha tetszőleges $x \geq 0$ és y vektorokra

$$L(x^*, y) \leq L(x^*, y^*) \leq L(x, y^*).$$

TÉTEL. *Az x^*, y^* vektorok akkor és csak akkor optimális megoldásai az entrópia programozási feladatpárnak, ha a Lagrange függvénynek nyeregpontja(i).*

Vegyes entrópia programozás

Legyen A $m \times n$-es mátrix, b m-dimenziós vektor, c n-dimenziós vektor. Legyen a $J = \{1, \ldots, n\}$ indexhalmaz két diszjunkt halmazra particionálva, jelölje ezeket J_E, J_N. A vegyes entrópia programozási feladatnak az alábbi matematikai programozási feladatot nevezzük.

Primál vegyes entrópia program:

Meghatározandó azon $x \in \mathbb{R}^{(n)}$ vektor, amelyre a

$$cx + \sum_{j \in J_E} x_j \log x_j - \sum_{j \in J_E} x_j$$

függvény minimális feltéve, hogy

$$\left.\begin{aligned} Ax &= b \\ x &\geq 0 \end{aligned}\right\} \quad \mathbb{P}.$$

Duál vegyes entrópia program:

Meghatározandó azon $y \in \mathbb{R}^{(m)}$ vektor, amelyre az

$$yb - \sum_{j \in J_E} \exp(ya_j - c_j)$$

függvény maximális feltéve, hogy

$$\left.ya_j - c_j \leq 0 \qquad j \in J_N\right\} \quad \mathbb{D}.$$

Megjegyezzük, hogy $J_E = \mathbb{O}$, $(J_N = J)$ esetén a lineáris programozási feladatot nyerjük. Az alábbi lemma a vegyes entrópia programozási feladatpár lehetséges megoldásaihoz tartozó célfüggvények értékeire ad összefüggést.

LEMMA. (A vegyes entrópia programozás alaplemmája).
Ha $x \in \mathbb{P}$ és $y \in \mathbb{D}$, akkor

$$cx + \sum_{j \in J_E} x_j \log x_j - \sum_{j \in J_E} x_j \geq yb - \sum_{j \in J_E} \exp(ya_j - c_j)$$

és egyenlőség akkor és csak akkor teljesül, ha

$$x_j = \exp(ya_j - c_j) \qquad j \in J_E\,,$$
$$x_j(ya_j - c_j) = 0 \qquad j \in J_N\,.$$

DUALITÁSI TÉTEL. *Tegyük fel, hogy* $\mathbb{P}$ *és* $\mathbb{D}$ *konzisztens. Ekkor létezik olyan* $x^* \in \mathbb{P}$ *vektor, hogy*

$$cx^* + \sum_{j\in J_E} x_j^* \log x_j^* - \sum_{j\in J_E} x_j^* \geq \sup_{y\in\mathbb{D}} \left\{ yb - \sum_{j\in J_E} \exp(ya_j - c_j) \right\}$$

és a supremum akkor és csak akkor cserélhető fel a maximummal, ha $\mathbb{P}$ *Slater konzisztens, azaz ha létezik olyan* x, *amelyre* $Ax = b$, $x > 0$.

2. Az entrópia programozás alkalmazása a szállítási feladatra

2.1. A feladat megfogalmazása

Tekintsük a klasszikus szállítási feladatot, amelynél legyenek a termelők (T) kínálatai $(a_1, \ldots, a_m) > 0$, a fogyasztók (F) keresletei $(b_1, \ldots, b_n) > 0$, az i–edik termelő és a j–edik fogyasztó közötti szállítási egységköltség pedig $c_{ij} \geq 0$. Jelölje x_{ij} a szállítási mennyiséget az i–edik termelőtől a j–edik fogyasztóhoz. A szállítási feladat az alábbi matematikai programozási feladattal fogalmazható meg:

Meghatározandók az x_{ij} mennyiségek úgy, hogy a

$$\sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij}$$

mennyiség minimális legyen feltéve, hogy

$$x_{ij} \geq 0, \qquad (i = 1, \ldots, m;\ j = 1, \ldots, n),$$
$$\sum_{j=1}^{n} x_{ij} = a_i, \qquad (i = 1, \ldots, m),$$
$$\sum_{i=1}^{m} x_{ij} = b_j, \qquad (j = 1, \ldots, n).$$

Módosítsuk a feladatot oly módon, hogy a

$$\sum_{j=1}^{n} x_{ij} = a_i, \qquad (i = 1, \ldots, m),$$
$$\sum_{i=1}^{m} x_{ij} = b_j, \qquad (j = 1, \ldots, n)$$

egyenleteket nem írjuk elő minden indexre, hanem csak bizonyosakra. Jelöljék az I_N és a J_N indexhalmazok azon termelők, illetve fogyasztók indexeit, amelyeknél előírjuk az egyenlőséget. Legyenek I_E és J_E azon indexhalmazok, amelyekre nem írjuk elő az egyenlőséget. Ezen indexekre azt követeljük meg, hogy a két oldalon lévő mennyiség D–eltérése minél kisebb legyen és ezt úgy valósítjuk meg, hogy az eltéréseket a célfüggvénybe építjük be az eredeti célfüggvény mellé valamilyen súlyarány figyelembevételével. Legyen λ a D–eltérés súlya és $(1-\lambda)$ pedig az eredeti célfüggvény súlya, ahol $0 < \lambda < 1$. Feladatunkat a fentiek alapján az alábbi matematikai programozási feladat írja le.

Meghatározandók az x_{ij} értékek úgy, hogy az

$$(1-\lambda)\sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij} + \lambda\left[\sum_{i\in I_E} D\left(\sum_{j=1}^{n} x_{ij} \parallel a_i\right) + \sum_{j\in J_E} D\left(\sum_{i=1}^{m} x_{ij} \parallel b_j\right)\right]$$

mennyiség minimális legyen feltéve, hogy

$$\begin{aligned} x_{ij} &\geq 0, \qquad (i = 1,\ldots,m;\ j = 1,\ldots,n),\\ \sum_{j=1}^{n} x_{ij} &= a_i, \qquad i \in I_N,\\ \sum_{i=1}^{m} x_{ij} &= b_j, \qquad j \in J_N. \end{aligned}$$

2.2. A probléma mint entrópia programozási feladat

A fenti feladatot entrópia programozási primál feladat alakjára transzformálhatjuk új változók bevezetésével. Vezessük be az y_i $(i \in I_E)$ és a z_j $(j \in J_E)$ új változókat az alábbiak szerint:

$$\begin{aligned} y_i &= \sum_{j=1}^{n} x_{ij}, \qquad i \in I_E,\\ z_j &= \sum_{i=1}^{m} x_{ij}, \qquad j \in J_E. \end{aligned}$$

Az új változók segítségével a következő feladatot kapjuk:

$$\begin{aligned} x_{ij} &\geq 0, \qquad (i = 1,\ldots,m;\ j = 1,\ldots,n),\\ \sum_{j=1}^{n} x_{ij} &= a_i, \qquad i \in I_N, \qquad (1)\\ \sum_{j=1}^{n} x_{ij} - y_i &= 0, \qquad i \in I_E, \end{aligned}$$

$$(1)\qquad \begin{aligned} \sum_{i=1}^{m} x_{ij} &= b_j, \qquad j \in J_N, \\ \sum_{i=1}^{m} x_{ij} - z_j &= 0, \qquad j \in J_E, \end{aligned}$$

$$(1-\lambda)\sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij} + \lambda\left[\sum_{i\in I_E} D(y_i \parallel a_i) + \sum_{j\in J_E} D(z_j \parallel b_j)\right] =$$

$$= (1-\lambda)\sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij} + \lambda\left[\sum_{i\in I_E} y_i \log\frac{\lambda y_i}{\lambda a_i} - \sum_{i\in I_E} y_i + \sum_{i\in I_E} a_i + \right.$$

$$\left. + \sum_{j\in J_E} z_j \log\frac{\lambda z_j}{\lambda b_j} - \sum_{j\in J_E} z_j + \sum_{j\in J_E} b_j\right] \text{ min!}$$

A fenti célfüggvényben a logaritmus mögötti törteket λ-val bővítettük és ezáltal az y_i, z_j változókról az $\hat{y}_i, \hat{z}_j$ új változókra az alábbi módon térhetünk át:

$$\hat{y}_i = \lambda y_i,$$
$$\hat{z}_j = \lambda z_j.$$

A célfüggvényből a konstans tagokat elhagyva az új változókkal a feladat az alábbi alakot ölti:

$$\begin{aligned} x_{ij} &\geq 0, \qquad (i = 1, \ldots, m;\ j = 1, \ldots, n), \\ \sum_{j=1}^{n} x_{ij} &= a_i, \qquad i \in I_N, \\ \sum_{j=1}^{n} \lambda x_{ij} - \hat{y}_i &= 0, \qquad i \in I_E, \\ \sum_{i=1}^{m} x_{ij} &= b_j, \qquad j \in J_N, \\ \sum_{i=1}^{m} \lambda x_{ij} - \hat{z}_j &= 0, \qquad j \in J_E, \end{aligned}$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n}(1-\lambda)c_{ij}x_{ij} + \sum_{i\in I_E}(-\log\lambda a_i)\hat{y}_i + \sum_{j\in J_E}(-\log\lambda b_j)\hat{z}_j +$$

$$+ \sum_{i\in I_E}\hat{y}_j\log\hat{y}_i - \sum_{i\in I_E}\hat{y}_i + \sum_{j\in J_E}\hat{z}_j\log\hat{z}_j - \sum_{j\in J_E}\hat{z}_j \text{ min!}$$

Látható, hogy ez egy vegyes entrópia programozási primál feladat. Írjuk fel ennek a duál párját. Legyenek a duálváltozók u_i $(i = 1, \dots, m)$ és v_j $(j = 1, \dots, n)$, amelyekre az alábbi feladat vonatkozik:

Meghatározandó u_i $(i = 1, \dots, m)$ és v_j $(j = 1, \dots, n)$, úgy, hogy a

$$(2) \qquad \sum_{i \in I_N} a_i u_i + \sum_{j \in J_N} b_j v_j - \sum_{i \in I_E} \exp(-u_i + \log \lambda a_i) - \sum_{j \in J_E} \exp(-v_j + \log \lambda b_j)$$

függvény maximális, feltéve, hogy

$$(3) \qquad \begin{aligned} u_i + v_j &\le (1-\lambda)c_{ij}, \quad i \in I_N,\ j \in J_N, \\ u_i + \lambda v_j &\le (1-\lambda)c_{ij}, \quad i \in I_N,\ j \in J_E, \\ \lambda u_i + v_j &\le (1-\lambda)c_{ij}, \quad i \in I_E,\ j \in J_N, \\ \lambda u_i + \lambda v_j &\le (1-\lambda)c_{ij}, \quad i \in I_E,\ j \in J_E. \end{aligned}$$

Az entrópia programozási feladatra vonatkozó egyensúlyi összefüggést, vagyis a primál és a duál feladat optimális megoldásaira vonatkozó optimalitási kritériumot esetünkben az alábbi formában írhatjuk:

$$(4) \qquad \begin{aligned} x_{ij}\,[(1-\lambda)c_{ij} - u_i - v_j] &= 0, \quad i \in I_N,\ j \in J_N, \\ x_{ij}\,[(1-\lambda)c_{ij} - u_i - \lambda v_j] &= 0, \quad i \in I_N,\ j \in J_E, \\ x_{ij}\,[(1-\lambda)c_{ij} - \lambda u_i - v_j] &= 0, \quad i \in I_E,\ j \in J_N, \\ x_{ij}\,[(1-\lambda)c_{ij} - \lambda u_i - \lambda v_j] &= 0, \quad i \in I_E,\ j \in J_E, \end{aligned}$$

$$(5) \qquad \begin{aligned} \hat{y}_i &= \exp(-u_i + \log \lambda a_i), \qquad i \in I_E, \\ \hat{z}_j &= \exp(-v_j + \log \lambda b_j), \qquad j \in J_E. \end{aligned}$$

Visszatérve az y_i, z_j változókra és felhasználva az (5) egyensúlyi feltételeket, az (1), (3) és a (4) figyelembevételével az x_{ij} primál és u_i, v_j duál változók optimális értékeinek meghatározására az alábbi rendszer megoldása szolgál:

$$(6) \qquad \begin{aligned} x_{ij} &\ge 0, && (i = 1, \dots, m;\ j = 1, \dots, n), \\ \sum_{j=1}^{n} x_{ij} &= a_i, && i \in I_N, \\ \sum_{j=1}^{n} x_{ij} &= a_i \exp(-u_i), && i \in I_E, \\ \sum_{i=1}^{m} x_{ij} &= b_j, && j \in J_N, \\ \sum_{i=1}^{m} x_{ij} &= b_j \exp(-v_j), && j \in J_E, \end{aligned}$$

$$
\begin{aligned}
u_i + v_j &\le (1-\lambda)c_{ij}, \quad i \in I_N,\ j \in J_N,\\
u_i + \lambda v_j &\le (1-\lambda)c_{ij}, \quad i \in I_N,\ j \in J_E,\\
\lambda u_i + v_j &\le (1-\lambda)c_{ij}, \quad i \in I_E,\ j \in J_N,\\
\lambda u_i + \lambda v_j &\le (1-\lambda)c_{ij}, \quad i \in I_E,\ j \in J_E,
\end{aligned} \tag{7}
$$

$$
\begin{aligned}
x_{ij}\,[(1-\lambda)c_{ij} - u_i - v_j] &= 0, \quad i \in I_N,\ j \in J_N,\\
x_{ij}\,[(1-\lambda)c_{ij} - u_i - \lambda v_j] &= 0, \quad i \in I_N,\ j \in J_E,\\
x_{ij}\,[(1-\lambda)c_{ij} - \lambda u_i - v_j] &= 0, \quad i \in I_E,\ j \in J_N,\\
x_{ij}\,[(1-\lambda)c_{ij} - \lambda u_i - \lambda v_j] &= 0, \quad i \in I_E,\ j \in J_E.
\end{aligned} \tag{8}
$$

2.3. Megoldási algoritmus

A primál és a duál feladat optimális megoldására szolgáló algoritmust az alábbiakban vázolhatjuk. Kiindulunk egy u_i, v_j duál megoldásból, amely a (7)-et kielégíti és megkisérlünk olyan x_{ij} megoldást keresni, amely a (6) primál és a (8) egyensúlyi feltételeket teljesíti. Az algoritmust egy tételben foglaljuk össze és a tételre adandó bizonyítás konstruktív volta adja az iterációs eljárást.

TÉTEL. *Legyenek u_i, v_j duál lehetséges megoldások, amelyek kielégítik* (7)*-et.*

(i) *Vagy meg tudunk adni olyan x_{ij} megoldást, amely kielégíti* (6)*-ot és* (8)*-at,*

(ii) *vagy elő tudunk állítani olyan új u_i, v_j lehetséges duál megoldásokat, amelyekre vonatkozó* (2) *duál célfüggvényérték határozottan nagyobb, mint az előző célfüggvényérték.*

Bizonyítás. Az egyszerűbb tárgyalásmód miatt jelöljük $\hat{c}_{ij}$-vel az $(1-\lambda)c_{ij}$ mennyiséget, valamint vezessük be az u_i, v_j változók helyett az $\hat{u}_i, \hat{v}_j$ változókat az alábbi módon:

$$
\hat{u}_i = \begin{cases} u_i, & \text{ha } i \in I_N \\ \lambda u_i, & \text{ha } i \in I_E, \end{cases} \qquad \hat{v}_j = \begin{cases} v_j, & \text{ha } j \in J_N \\ \lambda v_j, & \text{ha } j \in J_E, \end{cases} \tag{9}
$$

Ekkor a (7) duál feltételrendszer és a (8) egyensúlyi feltételek egyszerűbb alakban írhatók, azaz

$$
\hat{u}_i + \hat{v}_j \le \hat{c}_{ij}, \qquad (i = 1,\dots,m;\ j = 1,\dots,n), \tag{10}
$$

$$
x_{ij}(\hat{c}_{ij} - \hat{u}_i - \hat{v}_j = 0, \qquad (i = 1,\dots,m;\ j = 1,\dots,n). \tag{11}
$$

A (10) összefüggésből egy induló $\hat{u}_i, \hat{v}_j$ egyszerűen előállítható, például a klasszikus szállítási feladatnál használatos sor–oszlop redukció segítségével. Mint tudjuk az x_{ij} meghatározására a (6) és a (11) egyenletrendszer megoldása szolgál. Erre alkalmazhatnánk az általános KŐNIG modellt [9] úgy, hogy a kínálatokat a (6) egyenletek

jobboldalai, azaz a_i $(i \in I_N)$, illetve $a_i \exp(-\hat{u}_i/\lambda)$, $(i \in I_E)$, a keresleteket pedig a b_j $(j \in J_N)$ illetve $b_j \exp(-\hat{v}_j/\lambda)$ $(j \in J_E)$ jelentse. Azonban ezek a kínálatok és keresletek általában nem teljesítik a kereslet–kínálat egyensúlyt, így az általános KŐNIG modell közvetlenül nem használható. Azonban mégis kínálkozik lehetőség az általános KŐNIG modell alkalmazására az x_{ij} értékek meghatározásában, mivel a kereslet–kínálat egyensúly biztosítható. Ha valamely $\hat{u}_i, \hat{v}_j$ duál megoldás, akkor az $\hat{u}_i + \vartheta$, $\hat{v}_j - \vartheta$ is duál megoldás, azaz kielégíti (10)–et, ahol ϑ tetszőleges szám. A ϑ meghatározása pedig úgy történik, hogy a kereslet–kínálat egyensúlya fennálljon. A fentiek alapján az alábbiak szerint konstruáljuk meg az általános KŐNIG modellt:

Legyenek a kínálatok (r_i)

$$r_i = \begin{cases} a_i & i \in I_N \\ a_i \exp[-(\hat{u}_i + \vartheta)/\lambda] & i \in I_E \end{cases},$$

és a keresletek (s_j)

$$s_j = \begin{cases} b_j & j \in J_N \\ b_j \exp[-(\hat{v}_j - \vartheta)/\lambda] & j \in J_E \end{cases},$$

továbbá legyen megengedett a szállítás, ahol

$$\hat{c}_{ij} - \hat{u}_i - \hat{v}_j = 0\,,$$

és letiltjuk a szállítást, ahol

$$\hat{c}_{ij} - \hat{u}_i - \hat{v}_j > 0\,.$$

A ϑ értékének meghatározására a $\sum_{i=1}^{m} r_i = \sum_{j=1}^{n} s_j$ kereslet–kínálat egyensúly teljesítése szolgál, amelyet részletezés nélkül az alábbi összefüggéssel határozhatunk meg:

$$\vartheta = \lambda \log \frac{p_1 + \sqrt{p_1^2 + 4p_2p_3}}{2p_3}\,,$$

ahol

$$p_1 = \sum_{i \in I_N} a_i - \sum_{j \in J_N} b_j\,, \qquad p_2 = \sum_{i \in I_E} a_i \exp(-\hat{u}_i/\lambda)\,,$$

$$p_3 = \sum_{j \in J_E} b_j \exp(-\hat{v}_j/\lambda)\,.$$

(i) Ha a megkonstruált KŐNIG modell megoldható, akkor jelölje x_{ij} a szállítási mennyiségeket. Ez az x_{ij} kielégíti a (6) primál és a (11) optimalitási feltételeket is, így az x_{ij} optimális megoldás.

(ii) Ha a megkonstruált KŐNIG modell nem oldható meg, akkor a KŐNIG tétel [9] értelmében van olyan $P \subset \{1,\dots,m\}$ és $R \subset \{1,\dots,n\}$ indexhalmazpár, amelyekre

$$\sum_{i\in P} r_i > \sum_{j\in R} s_j ,$$

amely részletesebben

$$(12)\qquad \sum_{i\in P\cap I_N} a_i + \sum_{i\in P\cap I_E} a_i \exp[-(\hat{u}_i+\vartheta)/\lambda] > \\ > \sum_{j\in R\cap J_N} b_j + \sum_{j\in R\cap J_E} b_j \exp[-(\hat{v}_j-\vartheta)/\lambda],$$

és az alábbiak érvényesek:

$$\hat{c}_{ij} - \hat{u}_i - \hat{v}_j > 0, \qquad \text{ha } i\in P,\ j\notin R,$$
$$x_{ij} = 0, \qquad \text{ha } i\notin P,\ j\in R.$$

A P és az R halmazok segítségével új $\hat{u}'_i$ és $\hat{v}'_j$ duál megoldásokat konstruálunk az egyelőre ismeretlen ε segítségével:

$$\hat{u}'_i = \begin{cases} \hat{u}_i+\varepsilon, & \text{ha } i\in P \\ \hat{u}, & \text{ha } i\notin P \end{cases}; \qquad \hat{v}'_j = \begin{cases} \hat{v}_j-\varepsilon, & \text{ha } j\in R \\ \hat{v}_j, & \text{ha } j\notin R \end{cases},$$

illetve az u'_i és v'_j változókkal:

$$u'_i = \begin{cases} u_i+\varepsilon, & i\in P\cap I_N \\ u_i+\varepsilon/\lambda, & i\in P\cap I_E \\ u_i, & i\notin P \end{cases}; \qquad \hat{v}'_j = \begin{cases} v_j-\varepsilon, & j\in R\cap J_N \\ v_j-\varepsilon/\lambda, & j\in R\cap J_E \\ v_j, & j\notin R \end{cases}.$$

Az ε értéket két szempont figyelembevételével választjuk:

a) $\hat{u}'_1, \hat{v}'_j$ lehetséges megoldások legyenek,

b) a (2) duál célfüggvény értéke a legnagyobb legyen.

ad a) Legyen $\varepsilon_1 = \min\{\hat{c}_{ij} - \hat{u}_i - \hat{v}_j \mid i\in P,\ j\notin R\} > 0$.

Ha ε olyan, hogy $0 \le \varepsilon \le \varepsilon_1$, akkor $\hat{u}'_i, \hat{v}'_j$ lehetséges megoldások, azaz kielégítik a $\hat{c}_{ij} - \hat{u}'_i - \hat{v}'_j \ge 0$ feltételt, ugyanis

ha $i\in P$ és $j\in R$, akkor $\hat{c}_{ij} - \hat{u}'_i - \hat{v}'_j = \hat{c}_{ij} - \hat{u}_i - \hat{v}_j \ge 0$,

ha $i\notin P$ és $j\notin R$, akkor $\hat{c}_{ij} - \hat{u}'_i - \hat{v}'_j = \hat{c}_{ij} - \hat{u}_i - \hat{v}_j \ge 0$,

ha $i\in P$ és $j\notin R$, akkor $\hat{c}_{ij} - \hat{u}'_i - \hat{v}'_j = \hat{c}_{ij} - \hat{u}_i - \hat{v}_j - \varepsilon \ge 0$, $\varepsilon \le \varepsilon_1$ esetén és

ha $i\notin P$ és $j\in R$, akkor $\hat{c}_{ij} - \hat{u}'_i - \hat{v}'_j = \hat{c}_{ij} - \hat{u}_i - \hat{v}_j + \varepsilon \ge 0$, $\varepsilon \ge 0$ esetén.

ad b) Vizsgáljuk meg ezután a (2) duál célfüggvényt, amely a következő:

$$F(\hat{u},\hat{v}) = \sum_{i\in I_N} a_i(\hat{u}_i + \vartheta) + \sum_{j\in J_N} b_j(\hat{v}_j - \vartheta) - \\ - \sum_{i\in I_E} \lambda a_i \exp[-(\hat{u}_i + \vartheta)/\lambda] - \sum_{j\in J_E} \lambda b_j \exp[-(\hat{v}_j - \vartheta)/\lambda].$$

Jelöljük $\varphi(\varepsilon)$-nal az $F(\hat{u},\hat{v})$ célfüggvény növekedését ε függvényében, amelyre igaz az, hogy

$$\varphi(\varepsilon) = F(\hat{u}',\hat{v}') - F(\hat{u},\hat{v}) = \varepsilon\Big(\sum_{i\in P\cap I_N} a_i - \sum_{j\in R\cap J_N} b_j\Big) + \\ + (1 - \exp(-\varepsilon/\lambda) \sum_{i\in P\cap I_E} \lambda a_i \exp[-(\hat{u}_i + \vartheta)/\lambda] + \\ + (1 - \exp(\varepsilon/\lambda) \sum_{j\in R\cap J_E} \lambda b_j \exp[-(\hat{v}_j - \vartheta)/\lambda].$$

A $\varphi(\varepsilon)$ függvény első deriváltját képezve

$$\varphi'(\varepsilon) = \sum_{i\in P\cap I_N} a_i - \sum_{j\in R\cap J_N} b_j + \exp(-\varepsilon/\lambda) \sum_{i\in P\cap I_E} a_i \exp[-(\hat{u}_i + \vartheta)/\lambda] - \\ - \exp(\varepsilon/\lambda) \sum_{j\in R\cap J_E} b_j \exp[-(\hat{v}_j - \vartheta)/\lambda]).$$

Jelölje q_1, q_2, q_3 a $\varphi'(\varepsilon)$ összefüggésben lévő tagokat az alábbiak szerint:

$$q_1 = \sum_{i\in I_N} a_i - \sum_{j\in R\cap J_N} bj, \qquad q_2 = \sum_{i\in P\cap I_E} a_i \exp[-(\hat{u}_i + \vartheta)/\lambda], \\ q_3 = \sum_{j\in R\cap J_E} b_j \exp[-(\hat{v}_j - \vartheta)/\lambda],$$

és legyen $\delta = \exp(\varepsilon/\lambda)$ így a $\varphi'(\varepsilon) = 0$ egyenlet átrendezéssel az alábbi alakot ölti:

$$q_3\delta^2 - q_1\delta - q_2 = 0.$$

A fenti másodfokú egyenlet pozitív gyökét keresve megállapítható, hogy q_2, q_3 pozitivitása miatt az egyik gyök pozitív, a másik pedig negatív, a KŐNIG tétel miatt (amely a jelöléseinkkel a (12) alapján $q_1 > q_3 - q_2$ formát ölti) pedig a pozitív gyök egynél nagyobb. A keresett pozitív gyök:

$$\delta = \frac{q_1 + \sqrt{q_1^2 + 4q_2q_3}}{2q_3} > 1,$$

így $\varepsilon = \varepsilon_2 = \lambda \log \delta > 0$ választással a duál célfüggvény növekedése a legnagyobb. Ahhoz, hogy a duál célfüggvény értéke a legnagyobb legyen és a duál megengedettség megmaradjon ε-t a következőképpen választjuk:

$$\varepsilon = \min\{\varepsilon_1, \varepsilon_2\}\,.$$

2.4. Az algoritmus, mint megengedett irány módszer

A megoldó algoritmust duál módszernek is nevezhetjük, mivel az eljárás minden lépésében a duál feladat egy lehetséges megoldását határozzuk meg. A (9)-et és a ϑ számmal való transzformációt figyelembevéve és ezt a (2) duál célfüggvénybe behelyettesítve a duál feladat az alábbi alakot ölti:

Maximalizálandó az

$$F(\hat{u}, \hat{v}) = \sum_{i \in I_N} a_i(\hat{u}_i + \vartheta) + \sum_{j \in J_N} b_j(\hat{v}_j - \vartheta) - \sum_{i \in I_E} \lambda a_i \exp[-(\hat{u}_i + \vartheta)/\lambda] - \\ - \sum_{j \in J_E} \lambda b_j \exp[-(\hat{v}_j - \vartheta)/\lambda]$$

függvény feltéve, hogy

$$\hat{u}_i + \hat{v}_j \le \hat{c}_{ij} \qquad (i = 1, \ldots, m;\ j = 1, \ldots, n)\,.$$

Mivel az $F(\hat{u}, \hat{v})$ függvény konkáv, így használhatunk megengedett irány módszert. Legyen $\hat{u}_i, \hat{v}_j$ egy lehetséges megoldás. Jelölje a megengedett irányokat d_i^u $(i = 1, \ldots, m)$ és d_j^v $(j = 1, \ldots, n)$. Jelölje T azon indexpárok halmazát, amelyeknél a duál feltétel egyenlőséggel teljesül, azaz

$$T = \{(i, j) \mid \hat{c}_{ij} - \hat{u}_i - \hat{v}_j = 0\}\,.$$

Mint ismeretes az irányok meghatározására az alábbi lineáris programozási feladat szolgál (ZOUTENDIJK [24]), ahol a modellben a Csebisev normálást használjuk:

$$\begin{aligned} d_i^u + d_j^v &\le 0, && (i, j) \in T\,, \\ -1 \le d_i^u &\le 1, && (i = 1, \ldots, m)\,, \\ -1 \le d_j^v &\le 1, && (j = 1, \ldots, n)\,, \\ \langle \operatorname{grad} F(\hat{u}, \hat{v}) d \rangle & \quad \max! \end{aligned}$$

A $\operatorname{grad} F(\hat{u}, \hat{v})$ vektor koordinátáit a bevezetett kínálati (r_i) és a keresleti (s_j) mennyiségek adják, azaz

$$\operatorname{grad} F(u, v) = (r_1, \ldots, r_m;\ s_1, \ldots, s_n)\,.$$

Mivel a gradiensvektor együtthatói, azaz az iránykereső feladat célfüggvényének együtthatói pozitívak, így a d_i^u, d_j^v irányokra vonatkozó alsó korlátok elhagyhatók, tehát az iránymeghatározó lineáris programozási feladat az alábbi

$$\begin{aligned} d_i^u + d_j^v &\le 0, \qquad (i,j) \in T, \\ d_i^u &\le 1, \qquad (i = 1, \dots, m), \\ d_j^v &\le 1, \qquad (j = 1, \dots, n), \\ \sum_{i=1}^{m} r_i d_i^u + \sum_{j=1}^{n} s_j d_j^v &\quad \max! \end{aligned}$$

Ehhez a lineáris programozási feladathoz tartozó duál feladatot a következőképpen írhatjuk fel. Legyenek ezen feladat változói x_{ij} $(i,j) \in T$, t_i $(i = 1, \dots, m)$, p_j $(j = 1, \dots, n)$, ekkor

$$\begin{aligned} \sum_{j|T} x_{ij} + t_i &= r_i, \qquad (i = 1, \dots, m), \\ \sum_{j|T} x_{ij} + p_j &= s_j, \qquad (j = 1, \dots, n), \\ t_i &\ge 0, \qquad (i = 1, \dots, m), \\ p_j &\ge 0, \qquad (j = 1, \dots, n), \\ x_{ij} &\ge 0, \qquad (i,j) \in T, \\ \sum_{i=1}^{m} t_i + \sum_{j=1}^{n} p_j &\quad \min! \end{aligned}$$

Terjesszük ki az x_{ij} változókat az összes indexpárra és a feltételi egyenletekből a célfüggvényt ekvivalens alakban írva azt kapjuk, hogy

$$\begin{aligned} \sum_{j=1}^{n} x_{ij} + t_i &= r_i, \qquad (i = 1, \dots, m), \\ \sum_{i=1}^{m} x_{ij} + p_j &= s_j, \qquad (j = 1, \dots, n), \\ t_i &\ge 0, \qquad (i = 1, \dots, m), \\ p_j &\ge 0, \qquad (j = 1, \dots, n), \\ x_{ij} &\ge 0, \qquad (i = 1, \dots, m), \ (j = 1, \dots, n), \\ x_{ij} &= 0, \qquad (i,j) \notin T, \\ \sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij} &\quad \max! \end{aligned}$$

Az algoritmus során ezt az általános KŐNIG feladatot oldjuk meg. Amennyiben a KŐNIG feladat nem oldható meg, azaz nem mindegyik t_i és p_j zérus, kiadódnak a P és R indexhalmazok, amelyekből a lineáris programozás optimalitási kritériumának felhasználásával a d_i^u és d_j^v irányok meghatározhatók. Az algoritmus konvergenciáját a lehetséges irányokra kidolgozott elmélet igazolja (ZOUTENDIJK [24]).

3. Az entrópia programozás alkalmazása az input–output táblák előrebecslésére

3.1. Az input–output táblák előrebecslési feladata

Az ágazati kapcsolatok, a közlekedési és a szállítási struktúrák vizsgálatánál gyakran előforduló feladat az input–output táblák előrebecslése. Input–output táblán olyan táblázatot értünk, amelynek elemei adott kibocsátási és befogadóhelyek között áramló mennyiségek számértékeit tartalmazzák. Az előrebecslés feladata az, hogy amennyiben ismerjük a jelenlegi input–output táblát és ismerjük a megváltozott teljes kibocsátási és befogadási mennyiségeket minden kibocsátási és befogadóhelyre, akkor hogyan lehet megbecsülni, prognózist adni az új input–output tábláról. Az $a_{ij} > 0$ szám jelölje az i–edik kibocsátóhelyről a j–edik befogadóhelyre a forgalom értékét. Az a_{ij} értékeket az A mátrixba foglalva ezt nevezzük input–output táblázatnak. A $\sum_{j=1}^{n} a_{ij}$ mennyiség az i–edik kibocsátóhely összes kibocsátása, a $\sum_{i=1}^{m} a_{ij}$ mennyiség a j–edik befogadóhely összes befogadása. Ezeket a mennyiségeket az input–output tábla marginális értékeinek nevezzük. Feladatunk az alábbiakban foglalható össze: Ismerve a jelenlegi

$A = (a_{ij})$ input–output táblát és a megváltozott

$d = (d_1, d_2, \ldots, d_m) > 0$ marginális input

$b = (b_1, b_2, \ldots, b_n) > 0$ marginális output értékeket, becsüljük meg, illetve adjunk prognózist az új $X = (x_{ij})$ input–output tábláról. A fenti feladat megoldására elterjedt eljárás az úgynevezett RAS módszer [6, 20, 22, 23]. Ennél a módszernél az a feltételezés, hogy a jövőbeli forgalmat a jelenlegi forgalom faktorokkal való felszorzásával állítja elő, azaz $x_{ij} = r_i a_{ij} s_j$ alakban keresi a megoldást. Ez a feltételezés adja a módszer nevét, ha az r_i számokból képzett diagonál mátrixot R–rel, az s_j számokból képzett diagonál mátrixot S–el jelöljük, akkor a megoldásra az $X = \text{RAS}$ alak adódik. A RAS módszer mellett szintén elterjedt eljárás a DEMING–STEPHAN [2] által javasolt, a négyzetes kontingenciát minimalizáló hipotézisen alapuló eljárás. Ebben az esetben a hipotézis az, hogy olyan X táblát keresünk, amelyre az X táblának az A táblához viszonyított eltérése minimális, azaz, ha a

$$\sum_{i=1}^{m} \sum_{j=1}^{n} \frac{(x_{ij} - a_{ij})^2}{a_{ij}}$$

függvény értéke minimális. E módszert potenciálok módszerének is nevezik a fenti matematikai programozási feladat egyensúlyi (optimalitási) kritériuma alapján, az X^* ugyanis akkor és csak akkor optimális, ha

$$x_{ij}^* = a_{ij}(u_i + v_j),$$

és az u_i, v_j értékeket nevezik potenciáloknak.

3.2. A RAS mint entrópia programozási feladat

Dolgozatomban az input–output táblák előrebecslésénél az alábbi hipotézissel élek: Azt a táblát tekintem „jó" előrebecslésnek, amelyre az X tábla és az A tábla D–eltérése minimális. Mint láttuk a potenciálok módszerénél is eltérés szerepelt a hipotézisben. Azonban a D–eltérésfüggvény „jobbnak" mutatkozik, ugyanis a D–eltérés nem nagyobb, mint a potenciálok módszerénél alkalmazott eltérés függvény, amely tulajdonképpen a Pearson eltérés. A fenti hipotézissel élve az alábbi matematikai programozási feladat megoldásaként nyerhetjük az $X = (x_{ij})$ input–output táblát:

$$x_{ij} \geq 0, \qquad (i = 1, \ldots, m;\ j = 1, \ldots, n),$$

$$\sum_{j=1}^{n} x_{ij} = d_i, \qquad (i = 1, \ldots, m),$$

$$\sum_{i=1}^{m} x_{ij} = b_j, \qquad (j = 1, \ldots, n),$$

$$D(X \parallel A) = \sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij} \log \frac{x_{ij}}{a_{ij}} - \sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij} + \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij} \text{ min!}$$

A fenti feladat egy entrópia programozási primál feladat. Az entrópia programozás egyensúlyi feltételét figyelembe véve a primál és a duál feladat optimális megoldására az alábbi összefüggés áll fenn:

$$x_{ij} = \exp(u_i + v_j + \log a_{ij})$$

Vezessük be az $r_i = \exp(u_i)$ $(i = 1, \ldots, m)$ és az $s_j = \exp(v_j)$ $(j = 1, \ldots, n)$ jelöléseket, amelyeket felhasználva az alábbiakat kapjuk:

$$x_{ij} = r_i a_{ij} s_j, \qquad (i = 1, \ldots, n;\ j = 1, \ldots, n).$$

Tehát a RAS módszer egy entrópia programozási feladatból is megkapható.

3.3. Az általános GRAVITY modell

A bonyolultabb modelleket (GRAVITY modell) akkor használják, ha ismertek a kibocsátó- és a befogadóhelyek között áramló mennyiségek c_{ij} mozgatási költségei és a jövőbeli margóértékeken kívül a jövőbeli összforgalmi költségszint is adott. Az előrebecslést ennél a feladatnál is azon hipotézis alapján végezzük, hogy az $\boldsymbol{X} = (x_{ij})$ prognózist akkor fogadjuk el „jónak", ha a jelenlegi $\boldsymbol{A} = (a_{ij})$ táblához képest a $\boldsymbol{D}$–eltérés a lehető legkisebb. Legyen az i–edik kibocsátóhely jövőbeli összes kibocsátása $d_i > 0$, a j–edik befogadóhely jövőbeli összes befogadása $b_j > 0$, az i–edik kibocsátóhely és a j–edik befogadóhely között áramló mennyiség mozgatási egységköltsége $c_{ij} \geq 0$, valamint a jövőbeli összforgalmi költségszint K. A fentiek alapján a feladat matematikailag a következőképpen adható meg. Meghatározandók $\boldsymbol{X} = (x_{ij})$ értékek úgy, hogy az

$$
\begin{aligned}
x_{ij} &\geq 0, \qquad (i = 1, \ldots, m;\ j = 1, \ldots, n), \\
\sum_{j=1}^{n} x_{ij} &= d_i, \qquad (i = 1, \ldots, m), \\
\sum_{i=1}^{m} x_{ij} &= b_j, \qquad (j = 1, \ldots, n), \\
\sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} &= K
\end{aligned}
$$

feltételek mellett a

$$
\boldsymbol{D}(\boldsymbol{X} \parallel \boldsymbol{A}) = \sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij} \log \frac{x_{ij}}{a_{ij}} - \sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij} + \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}
$$

célfüggvény értéke minimális legyen.

Módosítsuk feladatunkat az alábbiak szerint. Ne minden kibocsátóhelyre és minden befogadóhelyre írjuk elő a jövőbeli marginális értékeket egyenlőség formájában és ne a teljes összforgalmi költségszintet írjuk elő, hanem csak ezek egy részére írjuk elő az egyenlőséget. Legyen $I = \{1, \ldots, m\}$ indexhalmaz a kibocsátóhelyek indexeinek halmaza $J = \{1, \ldots, n\}$ a befogadóhelyek indexeinek halmaza és $IJ = \{(1,1), (1,2), \ldots, (m,n)\}$ az indexpárok halmaza. Legyen adott az $I_N \subseteq I$ indexhalmaz, amely azon kibocsátóhelyek indexeit jelöli, amelyeknél egyenlőség formájában írjuk elő az összkibocsátás teljesülését és $J_N \subseteq J$ pedig legyen azon befogadóhelyek indexeinek halmaza, amelyeknél az összbefogadást írjuk elő egyenlőség formájában. Legyen adott továbbá az $N \subseteq IJ$ indexpárhalmaz, amelynél a költségszint egyenlőség formájában való teljesülését írjuk elő. Legyenek $I_E = I \setminus I_N$, $J_E = J \setminus J_N$, $E = IJ \setminus N$. Az I_E, J_E indexhalmazokra nem írjuk elő az összkibocsátás ill. az összbefogadás és a marginális értékek egyenlőségét, hanem

azt, hogy a két nemnegatív mennyiség D–eltérése minél kisebb legyen, amelyet úgy valósítunk meg, hogy a célfüggvénybe építjük be ezeket a D–eltéréseket valamilyen súlyszám figyelembevételével. Hasonlóan az E halmazbeli indexpároknál is az összforgalmi költség és az előírt költségszint D–eltérését építjük be a célfüggvénybe. Legyen az N–beli indexpárokra az előírt költségszint K_N, az E–beliekre pedig K_E. Legyen továbbá $\lambda_1, \lambda_2, \lambda_3 > 0$, $\lambda_1 + \lambda_2 + \lambda_3 = 1$ adott súlyszámok, ahol λ_1 az eredeti célfüggvény, λ_2 a marginális értékektől való eltérések, λ_3 pedig a költségszinttől való eltérés súlyszámait jelenti. Feladatunkat az alábbi matematikai programozási feladat írja le.

Meghatározandó x_{ij} úgy, hogy az

$$\begin{aligned} x_{ij} &\geq 0, \qquad (i = 1, \ldots, m;\ j = 1, \ldots, n), \\ \sum_{j=1}^{n} x_{ij} &= d_i, \qquad i \in I_N, \\ \sum_{i=1}^{m} x_{ij} &= b_j, \qquad j \in J_N, \\ \sum\sum_{(i,j)\in N} c_{ij} x_{ij} &= K_N \end{aligned}$$

feltételek mellett a

$$\lambda_1 D(\boldsymbol{X} \parallel \boldsymbol{A}) + \lambda_2 \left[\sum_{i \in I_E} D\left(\sum_{j=1}^{n} x_{ij} \parallel d_i \right) + \sum_{j \in J_E} D\left(\sum_{i=1}^{m} x_{ij} \parallel b_j \right) \right] + \lambda_3 D\left(\sum\sum_{(i,j)\in E} c_{ij} x_{ij} \parallel K_E \right) \tag{13}$$

függvény értéke minimális legyen.

A fenti feladat egy entrópia programozási primál feladat. A (13) célfüggvényt hozzuk a megfelelő alakra, ehhez vezessük be az y_i $(i \in I_E)$, a z_j $(j \in J_E)$ és a t új változókat a következők szerint:

$$\begin{aligned} y_i &= \sum_{j=1}^{n} x_{ij}, \qquad i \in I_E, \\ z_j &= \sum_{i=1}^{m} x_{ij}, \qquad j \in J_E, \\ t &= \sum\sum_{(i,j)\in E} c_{ij} x_{ij}. \end{aligned}$$

Az új változók segítségével az alábbi feladatot nyerjük:

$$
\begin{aligned}
x_{ij} &\geq 0, \qquad (i = 1, \ldots, m;\ j = 1, \ldots, n), \\
\sum_{j=1}^{n} x_{ij} &= d_i, \qquad i \in I_N, \\
\sum_{j=1}^{n} x_{ij} - y_i &= 0, \qquad i \in I_E, \\
\sum_{i=1}^{m} x_{ij} &= b_j, \qquad j \in J_N, \\
\sum_{i=1}^{m} x_{ij} - z_j &= 0, \qquad j \in J_E, \\
\sum\sum_{(i,j)\in N} c_{ij} x_{ij} &= K_N, \\
\sum\sum_{(i,j)\in E} c_{ij} x_{ij} - t &= 0,
\end{aligned}
\tag{14}
$$

$$
\lambda_1 \left[\sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij} \log \frac{\lambda_1 x_{ij}}{\lambda_1 a_{ij}} - \sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij} + \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij} \right] +
$$

$$
+\lambda_2 \left[\sum_{i=I_E} y_i \log \frac{\lambda_2 y_i}{\lambda_2 d_i} - \sum_{i\in I_E} y_i + \sum_{i\in I_E} d_i + \sum_{j\in J_E} z_j \log \frac{\lambda_2 z_j}{\lambda_2 b_j} - \right.
$$

$$
\left. - \sum_{j\in J_E} z_j + \sum_{j\in J_E} b_j \right] + \left[t \log \frac{\lambda_3 t}{\lambda_3 K_E} - t + K_E \right] \ \min!
$$

A célfüggvényben a $\lambda_1, \lambda_2, \lambda_3$ súlyokkal a logaritmus mögötti törteket kibővítettük és ezáltal az x_{ij}, y_i, z_j, t változókról az $\hat{x}_{ij}, \hat{y}_i, \hat{z}_j, \hat{t}$ új változókra az alábbi módon térhetünk át.

$$
\begin{aligned}
\hat{x}_{ij} &= \lambda_1 x_{ij}, \\
\hat{y}_i &= \lambda_2 y_i, \\
\hat{z}_j &= \lambda_2 z_j, \\
\hat{t} &= \lambda_3 t.
\end{aligned}
$$

Az új változókkal a feladat az alábbi alakot ölti:

$$\hat{x}_{ij} \geq 0, \qquad (i = 1, \ldots, m;\ j = 1, \ldots, n),$$

$$\sum_{j=1}^{n} \hat{x}_{ij} = \lambda_1 d_i, \qquad i \in I_N,$$

$$\sum_{j=1}^{n} \lambda_2 \hat{x}_{ij} - \lambda_1 \hat{y}_i = 0, \qquad i \in I_E,$$

$$\sum_{i=1}^{m} \hat{x}_{ij} = \lambda_1 b_j, \qquad j \in J_N,$$

$$\sum_{i=1}^{m} \lambda_2 \hat{x}_{ij} - \lambda_1 \hat{z}_j = 0, \qquad j \in J_E,$$

$$\sum\sum_{(i,j)\in N} c_{ij} \hat{x}_{ij} = \lambda_1 K_N,$$

$$\sum\sum_{(i,j)\in E} \lambda_3 c_{ij} \hat{x}_{ij} - \lambda_1 \hat{t} = 0,$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n} \hat{x}_{ij}(-\log \lambda_1 a_{ij}) + \sum_{i\in I_E} \hat{y}_i(-log\lambda_2 d_i) + \sum_{j\in J_E} \hat{z}_j(-\log \lambda_2 b_j) +$$

$$+\hat{t}(-\log \lambda_3 K_E) + \sum_{i=1}^{m}\sum_{j=1}^{n} \hat{x}_{ij} \log \hat{x}_{ij} - \sum_{i=1}^{m}\sum_{j=1}^{n} \hat{x}_{ij} + \sum_{i\in I_E} \hat{y}_i \log \hat{y}_i -$$

$$- \sum_{i\in I_E} \hat{y}_i + \sum_{j\in J_E} \hat{z}_j \log \hat{z}_j - \sum_{j\in J_E} \hat{z}_j + \hat{t} \log \hat{t} - \hat{t} \ \min!$$

A célfüggvényből a konstans tagokat elhagytuk, így látható, hogy ez egy entrópia programozási primál feladat. Írjuk fel a fenti entrópia programozási primál feladat duál párját.

Meghatározandók azon u_i $(i = 1, \ldots, m)$, v_j $(j = 1, \ldots, n)$, w_N, w_E duál változók, amelyekre a

$$\sum_{i\in I_N} \lambda_1 d_i u_i + \sum_{j\in J_N} \lambda_1 b_j v_j + \lambda_1 K_N w_N -$$

$$- \sum\sum_{(i,j)\in(I_N\times J_N)\cap N} \exp(u_i + v_j + w_N c_{ij} + \log \lambda_1 a_{ij}) -$$

$$- \sum\sum_{(i,j)\in(I_N\times J_N)\cap E} \exp(u_i + v_j + \lambda_3 w_E c_{ij} + \log \lambda_1 a_{ij}) -$$

$$
\begin{aligned}
&- \sum_{(i,j)\in(I_N\times J_E)\cap N}\sum \exp(u_i + \lambda_2 v_j + w_N c_{ij} + \log\lambda_1 a_{ij}) - \\
&- \sum_{(i,j)\in(I_N\times J_E)\cap E}\sum \exp(u_i + \lambda_2 v_j + \lambda_3 w_E c_{ij} + \log\lambda_1 a_{ij}) - \\
&- \sum_{(i,j)\in(I_E\times J_N)\cap N}\sum \exp(\lambda_2 u_i + v_j + w_N c_{ij} + \log\lambda_1 a_{ij}) - \\
&- \sum_{(i,j)\in(I_E\times J_N)\cap E}\sum \exp(\lambda_2 u_i + v_j + \lambda_3 w_E c_{ij} + \log\lambda_1 a_{ij}) - \\
&- \sum_{(i,j)\in(I_E\times J_E)\cap N}\sum \exp(\lambda_2 u_i + \lambda_2 v_j + w_N c_{ij} + \log\lambda_1 a_{ij}) - \\
&- \sum_{(i,j)\in(I_E\times J_E)\cap E}\sum \exp(\lambda_2 u_i + \lambda_2 v_j + \lambda_3 w_E c_{ij} + \log\lambda_1 a_{ij}) - \\
&- \sum_{i\in I_E} \exp(-\lambda_1 u_i + \log\lambda_2 d_i) - \sum_{j\in J_E} \exp(-\lambda_1 v_j + \log\lambda_2 b_j) - \\
&\qquad - \exp(-\lambda_1 w_E + \log\lambda_3 K_E)
\end{aligned}
$$

függvény értéke maximális. Vezessük be az $r_i, s_j, \vartheta_N, \vartheta_E$ pozitív új duál változókat az alábbiak szerint:

$$
r_i = \begin{cases} \exp(u_i) & i \in I_N \\ \exp(\lambda_2 u_i) & i \in I_E \end{cases}, \qquad s_j = \begin{cases} \exp(v_j) & j \in J_N \\ \exp(\lambda_2 v_j) & j \in J_E \end{cases},
$$

$$
\vartheta_N = \exp(w_N), \qquad \vartheta_E = \exp(\lambda_3 w_E).
$$

Ezekkel az új duálváltozókkal a duál feladat célfüggvénye a következő egyszerűbb alakban írható:

$$
\begin{aligned}
&\sum_{i\in I_N} \lambda_1 d_i \log r_i + \sum_{j\in J_N} \lambda_1 b_j \log s_j + \lambda_1 K_N \log\vartheta_N - \sum_{(i,j)\in N}\sum \lambda_1 r_i a_{ij} s_j \vartheta_N^{c_{ij}} - \\
&- \sum_{(i,j)\in E}\sum \lambda_1 r_i a_{ij} s_j \vartheta_E^{c_{ij}} - \sum_{i\in I_E} \lambda_2 d_i r_i^{-\frac{\lambda_1}{\lambda_2}} - \sum_{j\in J_E} \lambda_2 b_j s_j^{-\frac{\lambda_1}{\lambda_2}} - \lambda_3 K_E \vartheta_N^{-\frac{\lambda_1}{\lambda_2}}.
\end{aligned}
$$

Az alábbiakban az egyensúlyi összefüggést írjuk fel, amely a primál és a duál vál-

tozók optimális értékeire vonatkozó feltételeket adja.

$$\hat{x}_{ij} = \lambda_1 r_i a_{ij} s_j \vartheta_N^{c_{ij}}, \qquad (i,j) \in N,$$

$$\hat{x}_{ij} = \lambda_1 r_i a_{ij} s_j \vartheta_E^{c_{ij}}, \qquad (i,j) \in E,$$

$$\hat{y}_i = \lambda_2 d_i r_i^{-\frac{\lambda_1}{\lambda_2}}, \qquad i \in I_E,$$

$$\hat{z}_j = \lambda_2 b_j s_j^{-\frac{\lambda_1}{\lambda_2}}, \qquad j \in J_E,$$

$$\hat{t} = \lambda_3 K_E \vartheta_E^{-\frac{\lambda_1}{\lambda_3}}.$$

Visszatérve az eredeti primál változókra a fenti optimalitási kritérium az alábbi alakot ölti:

$$x_{ij} = r_i a_{ij} s_j \vartheta_N^{c_{ij}}, \qquad (i,j) \in N,$$

$$x_{ij} = r_i a_{ij} s_j \vartheta_E^{c_{ij}}, \qquad (i,j) \in E,$$

$$y_i = d_i r_i^{-\frac{\lambda_1}{\lambda_2}}, \qquad i \in I_E,$$

$$z_j = b_j s_j^{-\frac{\lambda_1}{\lambda_2}}, \qquad j \in J_E,$$

$$t = K_E \vartheta_E^{-\frac{\lambda_1}{\lambda_3}}.$$

3.4. Megoldási algoritmus

A fenti egyensúlyi egyenleteket az entrópia programozási primál feladat feltételi egyenleteibe (14) behelyettesítve, az optimális megoldáspárt az alábbi nemlineáris egyenletrendszer megoldása szolgáltatja:

$$(15)\quad \begin{aligned}
&\sum_{j|N} r_i a_{ij} s_j \vartheta_N^{c_{ij}} + \sum_{j|E} r_i a_{ij} s_j \vartheta_E^{c_{ij}} = d_i, && i \in I_N \\
&\sum_{j|N} r_i a_{ij} s_j \vartheta_N^{c_{ij}} + \sum_{j|E} r_i a_{ij} s_j \vartheta_E^{c_{ij}} = d_i r_i^{-\frac{\lambda_1}{\lambda_2}}, && i \in I_E \\
&\sum_{i|N} r_i a_{ij} s_j \vartheta_N^{c_{ij}} + \sum_{i|E} r_i a_{ij} s_j \vartheta_E^{c_{ij}} = b_j, && j \in J_N \\
&\sum_{i|N} r_i a_{ij} s_j \vartheta_N^{c_{ij}} + \sum_{i|E} r_i a_{ij} s_j \vartheta_E^{c_{ij}} = b_j s_j^{-\frac{\lambda_1}{\lambda_2}} && j \in J_E
\end{aligned}$$

$$\sum_{(i,j)\in N}\sum c_{ij} r_i a_{ij} s_j \vartheta_N^{c_{ij}} = K_N \tag{16}$$

$$\vartheta_E^{-\frac{\lambda_1}{\lambda_2}} \sum_{(i,j)\in E}\sum c_{ij} r_i a_{ij} s_j \vartheta_E^{c_{ij}} = K_E\,. \tag{17}$$

Röviden összefoglalva az algoritmus az alábbi lépésekből áll. Kiindulunk tetszőleges $\vartheta_N > 0$ és $\vartheta_E > 0$ értékekből. A (15) egyenletrendszert megoldjuk és az eredményül kapott r_i, s_j értékeket a (16) és a (17) összefüggésekbe behelyettesítjük. Jelöljük a (16) baloldalát $\tilde{K}_N$-mal, a (17) baloldalát pedig $\tilde{K}_E$-mal. Ezután a (16)-beli és a (17)-beli baloldalakat és a jobboldalakat összehasonlítjuk, majd a két oldal eltérése esetén új ϑ_N és ϑ_E értékeket határozunk meg és eljárásunkat folytatjuk. A ϑ_N és ϑ_E értékeinek meghatározása az alábbiak szerint történik:

ha $\tilde{K}_N = K_N$ és $\tilde{K}_E = K_E$, akkor megállunk,
ha $\tilde{K}_N < K_N$, akkor ϑ_N értékét növeljük, ellenkező esetben csökkentjük,
ha $\tilde{K}_E < K_E$, akkor ϑ_E értékét növeljük, ellenkező esetben csökkentjük.

A továbbiakban a (15) egyenletrendszer megoldási algoritmusát adjuk meg. Kiindulunk az $r_i > 0$ $(i = 1, \ldots, m)$ értékekből és az alábbi (18), (19) összefüggésekkel iteratív eljárással határozzuk meg az r_i és az s_j értékeket.

$$s_j = \frac{b_j}{\sum\limits_{i|N} r_i a_{ij} \vartheta_N^{c_{ij}} + \sum\limits_{i|E} r_i a_{ij} \vartheta_E^{c_{ij}}}, \qquad j \in J_N$$

$$s_j = \left[\frac{b_j}{\sum\limits_{i|N} r_i a_{ij} \vartheta_N^{c_{ij}} + \sum\limits_{i|E} r_i a_{ij} \vartheta_E^{c_{ij}}}\right]^{\frac{\lambda_2}{\lambda_1+\lambda_2}}, \qquad j \in J_E \tag{18}$$

$$r_i = \frac{d_i}{\sum\limits_{j|N} a_{ij} s_j \vartheta_N^{c_{ij}} + \sum\limits_{j|E} a_{ij} s_j \vartheta_E^{c_{ij}}}, \qquad i \in I_N$$

$$r_i = \left[\frac{d_i}{\sum\limits_{j|N} a_{ij} s_j \vartheta_N^{c_{ij}} + \sum\limits_{j|E} a_{ij} s_j \vartheta_E^{c_{ij}}}\right]^{\frac{\lambda_2}{\lambda_1+\lambda_2}}, \qquad i \in I_E\,. \tag{19}$$

Most pedig a költségfüggvények azon tulajdonságait igazoljuk, amelyet az algoritmus megalkotásánál felhasználtunk. Az alábbiakban megmutatjuk, hogy a $K_N(\vartheta_N)$ és a $K_E(\vartheta_E)$ függvények szigorúan monoton növekvő függvények.

Legyenek a K'_N, K'_E költségszintekkel adott modellekhez tartozó optimális duál megoldások $r'_i, s'_j, \vartheta'_N, \vartheta'_E$. Hasonlóan a K''_N, K''_E költségszintű modellekhez tartozó optimális duál megoldások $r''_i, s''_j, \vartheta''_N, \vartheta''_E$.

Tekintsük a duál feladat célfüggvényét:

$$\sum_{i\in I_N}\lambda_1 d_i\log r_i+\sum_{j\in J_N}\lambda_1 b_j\log s_j+\lambda_1 K_N\log\vartheta_N-\sum_{(i,j)\in N}\sum\lambda_1 r_i a_{ij}s_j\vartheta_N^{c_{ij}}-$$
$$-\sum_{(i,j)\in E}\sum\lambda_1 r_i a_{ij}s_j\vartheta_E^{c_{ij}}-\sum_{i\in I_E}\lambda_2 d_i r_i^{-\frac{\lambda_1}{\lambda_2}}-\sum_{j\in J_E}\lambda_2 b_j s_j^{-\frac{\lambda_1}{\lambda_2}}-\lambda_3 K_E\vartheta_E^{-\frac{\lambda_1}{\lambda_3}}.$$

Jelöljük a célfüggvényben szereplő 1., 2., 4., 5., 6. és a 7. tag összegét P-vel. Így a fenti duál célfüggvény a számunkra egyszerűbb

$$P(\boldsymbol{r},\boldsymbol{s},\vartheta_N,\vartheta_E)+\lambda_1 K_N\log\vartheta_N-\lambda_3 K_E\vartheta_E^{-\frac{\lambda_1}{\lambda_3}}$$

alakban írható. Az optimális megoldás definíciója alapján felírhatjuk az alábbi négy egyenlőtlenséget.

$$(20)\quad P(\boldsymbol{r}',\boldsymbol{s}',\vartheta'_N,\vartheta'_E)+\lambda_1 K'_N\log\vartheta'_N-\lambda_3 K'_E\vartheta'^{-\frac{\lambda_1}{\lambda_3}}_E\geq$$
$$\geq P(\boldsymbol{r}'',\boldsymbol{s}'',\vartheta''_N,\vartheta''_E)+\lambda_1 K'_N\log\vartheta''_N-\lambda_3 K'_E\vartheta'^{-\frac{\lambda_1}{\lambda_3}}_E,$$

$$(21)\quad P(\boldsymbol{r}'',\boldsymbol{s}'',\vartheta''_N,\vartheta''_E)+\lambda_1 K''_N\log\vartheta''_N-\lambda_3 K''_E\vartheta''^{-\frac{\lambda_1}{\lambda_3}}_E\geq$$
$$\geq P(\boldsymbol{r}',\boldsymbol{s}',\vartheta'_N,\vartheta'_E)+\lambda_1 K''_N\log\vartheta'_N-\lambda_3 K''_E\vartheta''^{-\frac{\lambda_1}{\lambda_3}}_E,$$

$$(22)\quad P(\boldsymbol{r}',\boldsymbol{s}',\vartheta'_N,\vartheta'_E)+\lambda_1 K'_N\log\vartheta'_N-\lambda_3 K'_E\vartheta'^{-\frac{\lambda_1}{\lambda_3}}_E\geq$$
$$\geq P(\boldsymbol{r}'',\boldsymbol{s}'',\vartheta''_N,\vartheta''_E)+\lambda_1 K'_N\log\vartheta'_N-\lambda_3 K'_E\vartheta''^{-\frac{\lambda_1}{\lambda_3}}_E,$$

$$(23)\quad P(\boldsymbol{r}'',\boldsymbol{s}'',\vartheta''_N,\vartheta''_E)+\lambda_1 K''_N\log\vartheta''_N-\lambda_3 K''_E\vartheta''^{-\frac{\lambda_1}{\lambda_3}}_E\geq$$
$$\geq P(\boldsymbol{r}',\boldsymbol{s}',\vartheta'_N,\vartheta'_E)+\lambda_1 K''_N\log\vartheta''_N-\lambda_3 K''_E\vartheta'^{-\frac{\lambda_1}{\lambda_3}}_E,$$

A (20) és a (22) egyenlőtlenség azt fejezi ki, hogy a K'_N, K'_E költségszintekkel adott feladat esetén az $\boldsymbol{r}', \boldsymbol{s}', \vartheta'_N, \vartheta'_E$ optimális duálváltozókhoz tartozó duál célfüggvényérték nem kisebb bármely duálváltozóhoz tartozó célfüggvényértéknél. A (21) és a (23) egyenlőtlenség pedig azt fejezi ki, hogy a K''_N, K''_E költségszintekkel adott feladat esetén az $\boldsymbol{r}'', \boldsymbol{s}'', \vartheta''_N, \vartheta''_E$ optimális duálváltozókhoz tartozó duál célfüggvényérték nem kisebb bármely duálváltozóhoz tartozó célfüggvényértéknél. Adjuk össze a (20) és a (21) egyenlőtlenséget, a kieső $P(\boldsymbol{r}, \boldsymbol{s}, \vartheta_N, \vartheta_E)$ tagokat már le sem

írva és a $\lambda_3 > 0$–val való osztást elvégezve, valamint az egyszerűsítések után a következőket kapjuk:

$$K'_N \log \vartheta'_N + K''_N \log \vartheta''_N \geq K'_N \log \vartheta''_N + K''_N \log \vartheta'_N ,$$

melyet alkalmasan rendezve kapjuk, hogy

$$(K'_N - K''_N)(\log \vartheta'_N - \log \vartheta''_N) \geq 0 .$$

A fenti egyenlőtlenségből, ha

$$K'_N > K''_N, \quad \text{akkor} \quad \log \vartheta'_N \geq \log \vartheta''_N$$

következik, amelyből a természetes alapú logaritmus monotonitása miatt $\vartheta'_N \geq \vartheta''_N$ is következik. Mivel $\vartheta'_N = \vartheta''_N$ esetén $K'_N = K''_N$, ezért a $\vartheta_N(K_N)$ függvény szigorúan monoton növekvő. Ekkor viszont létezik a $K_N(\vartheta_N)$ inverz függvény, amely szintén szigorúan monoton növekvő.

Ha most a (22) és a (23) egyenlőtlenségeket adjuk össze, akkor az előzőekhez hasonlóan kapjuk, hogy

$$(K'_E - K''_E)\left(\vartheta''_E{}^{-\frac{\lambda_1}{\lambda_3}} - \vartheta'_E{}^{-\frac{\lambda_1}{\lambda_3}}\right) \geq 0 ,$$

amelyből, ha $K'_E > K''_E$, akkor λ_1, λ_3 pozitivitása miatt $\vartheta'_E \geq \vartheta''_E$ is következik. Mivel $\vartheta'_E = \vartheta''_E$ esetén $K'_E = K''_E$, ezért a $\vartheta_E(K_E)$ függvény szigorúan monoton növekvő. Ekkor viszont létezik a $K_E(\vartheta_E)$ inverz függvény, amely szintén szigorún monoton növekvő.

IRODALOM

[1] CSISZÁR I., „Eloszlások eltérésének információ típusú mértékszámai", *I. MTA III. Osztály Közleményei*, vol. 17 (1967), 125–149.

[2] DEMING, W. E. and STEPHAN, F. F., „On a Last Squares Adjustment of a Sampled Frequency Table when the Expected Marginal Totals are known", *Ann. Math. Statist.* **11** (1940), 427–444.

[3] DUFFIN, R. J., PETERSON, E. L. and ZENER, C., *Geometric Programming* (John Wiley, New York, 1966).

[4] DUFFIN, R. J. and PETERSON, E. L., „Duality Theory for Geometric Programming", *SIAM J. Appl. Math.* **14** (1966), 1307–1349.

[5] FORD, L. R. and FULKERSON, D. R., *Flows in Networks* (Princeton University Press, 1962).

[6] FRATAR, J. T., „Vehicular Trip Distribution by Successive Approximations", *Traffic Quarterly* (1954), 53–65.

[7] FRIEDLANDER, D., „A Technique for Estimation a Contingency Table, Given the Marginal Totals and some Supplementary Data", *Journal of the Royal Statistical Society Series A. CXXIV Pt.* 3, (1963), 412–420.

[8] GALE, D., *Theory of Linear Economic Models* (McGraw-Hill, New York, 1960).

[9] KLAFSZKY, E., *Hálózati folyamok* (A Bolyai János Matematikai Társulat kiadványa, Budapest, 1969).

[10] KLAFSZKY, E., „Geometriai Programozás és néhány alkalmazása", *MTA SZTAKI Tanulmányok* No. 8 (1973).

[11] KLAFSZKY, E., „A Theoretical Prediction of the Input Output Tables", *Lecture Notes in Computer Science* No. 4, *Springer Verlag* (1974), 484–492.

[12] KŐNIG, D., *Theorie der endlichen und unendlichen Graphen* (Akad. Verlagsgesellschaft, Leipzig, 1936).

[13] KULLBACK, S. and LEIBLER, P., „On Information and Sufficiency", *Ann. Math. Statist.* **22** (1951), 79–86.

[14] KULLBACK, S., *Information Theory and Statistics* (New York, 1959).

[15] NAGY, T., „Contributions to the Entropy Programming", *Mathematical Optimization*, Theory & Applications, Eisenach, 1990. december 9–13.

[16] NAGY, T., *Entropy Programming and its Applications*, 14th International Symposium on Mathematical Programming, Amsterdam, The Netherlands, August 5–9, 1991.

[17] NAGY, T., *Az entropy programozás és alkalmazásai*, XX. Magyar Operációkutatási Konferencia, Esztergom, 1991. október 7–9.

[18] NAGY, T., *Applications of the Entropy Programming*, 11th International Conference on Mathematical Programming, Mátrafüred, 1992. március 21–26.

[19] ROCKAFELLAR, R. T., *Convex Analysis* (University Press, Princeton, New Yersey, 1970).

[20] SELEJKOVSKIJ, G. B., *Transzportnije osznovanyija kompoziciji gorodszkovo plana* (Gupgor, L., 1963).

[21] STOER, J. and WITZGALL, CH., *Convexity and Optimization in Finite Dimensions I.* (Springer Verlag, 1970).

[22] STONE, R. and BROWN, A., *A Long-Term Growth Model for the British Economy*, Geary, R. C. (szerk): Europe's Future in Figure, Amsterdam, 1966.

[23] STONE, S., BATES, J. and BACHARACH, M., *A Programme for Growth*, Input–Output Relationsships 1954–66 (University of Cambridge, 1963).

[24] ZOUTENDIJK, G., *Mathematical Programming Methods* (North Holland, Amsterdam, 1976).

(Beérkezett: 1992. május 22.)

NAGY TAMÁS
MISKOLCI EGYETEM, MATEMATIKAI INTÉZET
ALKALMAZOTT MATEMATIKAI TANSZÉK
3515 MISKOLC-EGYETEMVÁROS

STOCHASTIC VARIANTS OF THE ENTROPY PROGRAMMING

T. NAGY

In the first part of this paper we present the entropy programming problem and its duality results. The entropy programming is to find a non-negative vector the divergence of which from a given positive vector should be minimal supposing linear constrains. The measure of the divergence between two non-negative vectors is the generalization of the Kullbach–Leibler information. In the main part of this paper we present two applications of this problem.

Consider the transportation problem where some constraints are not required to be exactly satisfied rather the above divergence of the two sides is put into the objective function such a way that the weighted average of the original and the divergence should be minimal. Then we apply the entropy programming problem to the gravity model for trip distribution. Similarly to the above case we put some constraints into the objective. Effecient algorithms were developed for the above problems, these algorithms have very good convergence properties.

Alkalmazott Matematikai Lapok **17** (1993) 171–184

A GOLYÓSMALMI ŐRLEMÉNY SŰRŰSÉGFÜGGVÉNYÉRE FELÍRT INTEGRO-DIFFERENCIÁLEGYENLET MEGOLDHATÓSÁGA ÉS A MEGOLDÁS SPECIÁLIS TULAJDONSÁGAI

MIHÁLYKÓ CSABA

Veszprém

A cikk a szakaszos golyósmalmi őrlés egy matematikai modelljével foglalkozik. Az őrölt anyag szemcseméret szerinti tömegeloszlásának sűrűségfüggvényére felírt integro-differenciálegyenletet vizsgáljuk. Bebizonyítjuk, hogy az egyenlet egyértelmű megoldása sűrűségfüggvény. Továbbá adunk egy olyan konvergens módszert a megoldás közelítő számítására, amely megfelel annak a kívánalomnak, hogy a kapott közelítő megoldás eleget tegyen a (diszkrét) anyagmegmaradás elvének.

1. Bevezetés

A kerámiaiparban széles körben használt eljárás a szakaszos golyósmalmi őrlés. Ez a következőt jelenti: Az őrlendő anyagot beteszik egy forgó, vízszintes hengerbe (malom), azt lezárják. A malomban lévő különböző méretű golyók a malom forgása következtében ide-oda ütődnek és az ütközések során összezúzzák a malomban lévő anyagot. Ez az őrlés.

Az egyik fontos kérdés, hogy időben hogyan változik a malomban lévő anyag szemcseméret szerinti tömegeloszlása. Ennek nyomon követésére alkalmas az anyag szemcseméret szerinti sűrűségfüggvényének vizsgálata.

Vezessük be a következő jelöléseket:
Jelöljük X_0-lal a legkisebb, és X_M-mel a legnagyobb szemcseméretet. $T < \infty$ jelölje az őrlés időtartamát. Legyen $x \in [X_0, X_M]$ és $t \in [0, T]$.

Legyen $v_0(x) : [X_0, X_M] \to \mathbb{R}_0^+$ a beadagolt anyag (kezdeti) sűrűségfüggvénye és $v(x,t) : [X_0, X_M] \times [0,T] \to \mathbb{R}_0^+$ a malomban tartózkodó anyag szemcseméret szerinti sűrűségfüggvénye. Vezessük be továbbá a következő két függvényt: az ún. szelekciós függvényt ($S(x)$), valamint az ún. törési sűrűségfüggvényt ($b(x,y)$). Az $S(x) : [X_0, X_M] \to \mathbb{R}_0^+$ függvény adja meg annak a mértékét, hogy egységnyi idő alatt egységnyi tömegű x méretű szemcséből mennyi törik. A $b(x,y) : H \to \mathbb{R}_0^+$ függvény pedig nem más, mint az y méretű szemcséből töréssel keletkező anyag eloszlásának a sűrűségfüggvénye, ahol $H := \{(x,y) \mid X_0 \le x \le y \le X_M,\ y \ne X_0\}$.

Ezek után tekintsük az őrlési folyamat leírására szolgáló egyik leginkább használatos egyenletet ([1], [3], [4], [5]).

$$\frac{\partial v(x,t)}{\partial t} = -S(x)v(x,t) + \int_x^{X_M} S(z)b(x,z)v(z,t)dz \tag{1.1}$$

$$v(x,0) = v_0(x)$$

ahol $x \in [X_0, X_M]$, $t \in [0,T]$ és $0 \leq X_0 < X_M \leq \infty$, $T < \infty$ valamint az $S(x)$, $b(x,y)$ és $v_0(x)$ függvények adottak.

2. Egzisztencia és unicitás

Az (1.1) egyenlet megoldására vonatkozóan a következő tételt mondhatjuk ki.

2.1. TÉTEL. *Tegyük fel, hogy fennállnak a következő feltételek:*

(i) $S(x)$ *és* $v_0(x)$ *folytonosak az* $[X_0, X_M]$ *intervallumban,* $b(x,y)$ *pedig folytonos a* H *halmazon.*

(ii) $v_0(x)$, $S(x)$ *és* $b(x,y)$ *nemnegatívak.*

(iii) $\forall z \in (X_0, X_M]$-re $\int_{X_0}^{z} b(x,z)dx = 1.$

(iv) $\int_{X_0}^{X_M} v_0(x)dx = 1.$

(v) $B := \sup_{x \in (X_0, X_M]} \int_{x}^{X_M} S(z)b(x,z)dz < \infty.$

Ekkor

1. *Az* (1.1) *integro-differenciálegyenletnek a folytonos függvények körében létezik és egyértelmű a megoldása, és az* t *szerint differenciálható.*
2. *Ez a megoldás nemnegatív.*
3. *A megoldás integrálja az* $[X_0, X_M]$ *intervallumon eggyel egyenlő, azaz* $\forall t \in [0,T]$-re $\int_{X_0}^{X_M} v(x,t)dx = 1.$

Bizonyítás. Az (1.1) egyenlet ekvivalens átalakításával a következő egyenletet kapjuk:

$$v(x,t) = e^{-S(x)t}v(x,0) + \int_0^t e^{S(x)(\tau - t)} \int_x^{X_M} S(z)b(x,z)v(z,\tau)dz\,d\tau\,.$$

Erre az egyenletre alapozva alkalmazhatjuk a Peano-iterációt a következőképpen:

Legyen

$$v_0(x,t) \equiv v_0(x)$$

és $n = 1, 2, \ldots$ esetén

$$v_n(x,t) = e^{-S(x)t}v_0(x) + \int_0^t e^{S(x)(\tau - t)} \int_x^{X_M} S(z)b(x,z)v_{n-1}(z,\tau)dz\,d\tau\,.$$

A tétel 1. állítása a Peano-iteráció szokásos felhasználásával bizonyítható. Ennek leírásától — jól ismert módszer lévén — eltekintünk. A tétel 2. állítása egyenes folyománya az iterációs előállításnak, ugyanis az egyenletben szereplő $v_0(x)$, $S(x)$ $b(x,y)$ függvények nemnegativitása miatt az iterációban rekurzívan megadott $v_n(x,t)$ függvények is nemnegatívak lesznek, így a határértékül adódó $v(x,t)$ megoldás is szükségképpen ilyen. Végezetül, tekintettel a 3. állítás nem szokványos jellegére, annak bizonyítását részletesen megadjuk.

Írjuk fel az (1.1) egyenletet a vele ekvivalens alábbi alakban:

$$v(x,t) = v_0(x) + \int_0^t [-S(x)v(x,t) + \int_x^{X_M} S(z)b(x,z)v(z,\tau)dz]d\tau\,, \quad t \in [0,T]\,.$$

Integráljuk mindkét oldalt X_0-tól X_M-ig, majd az integrálás sorrendjét cseréljük fel és az (iv) feltételt alkalmazzuk:

$$\int_{X_0}^{X_M} v(x,t)dx = \int_{X_0}^{X_M} \left[v_0(x) + \int_0^t \left(-S(x)v(x,\tau) + \int_x^{X_M} S(z)b(x,z)v(z,\tau)dz\right)d\tau\right]dx =$$

$$= \int_{X_0}^{X_M} v_0(x)dx + \int_0^t \int_{X_0}^{X_M} \left[-S(x)v(x,\tau) + \int_x^{X_M} S(z)b(x,z)v(z,\tau)dz\right]dx\,d\tau =$$

$$= 1 - \int_0^t \int_{X_0}^{X_M} S(x)v(x,\tau)dx\,d\tau + I$$

ahol $$I := \int_0^t \int_{X_0}^{X_M} \int_x^{X_M} S(z)b(x,z)v(z,\tau)dz\,dx\,d\tau\,.$$

Ismét felcserélve az integrálás sorrendjét és használva az (iii) feltételt, I-t így alakíthatjuk tovább:

$$I = \int_0^t \int_{X_0}^{X_M} \int_x^{X_M} S(z)b(x,z)v(z,\tau)dx\,dz\,d\tau =$$

$$= \int_0^t \int_{X_0}^{X_M} S(z)v(z,\tau)\left\{\int_{X_0}^z b(x,z)dx\right\}dz\,d\tau = \int_0^t \int_{X_0}^{X_M} S(z)v(z,\tau)dz\,d\tau$$

Innen következik az állításunk. □

KÖVETKEZMÉNY. *A 2.1. Tétel 2. és 3. állítása szerint* $\forall t \in [0,T]$*-re a* $v(\cdot,t)$ *függvény sűrűségfüggvény.*

Az irodalomban széles körben használt ([2], [3], [4]) a következő speciális eset:

$$0 = X_0 < X_M < \infty, \quad S(x) = k \cdot x^q \text{ és } b(x,y) = p \cdot \frac{x^{p-1}}{y^p}, \tag{2.1}$$

ahol $q \geq 0$, $k > 0$ és $p > 0$ valós számok.

KÖVETKEZMÉNY. *Ha* $v_0(x)$ *tetszőleges folytonos sűrűségfüggvény* $[0, X_M]$*-on és a* (2.1) *paraméterek olyanok, hogy* $p, q \in \mathbb{R}$*-re igaz, hogy*

$$\text{vagy } p \geq 1,\ q > 0; \quad \text{vagy pedig } p > 1,\ q = 0 \tag{2.2}$$

akkor az (1.1) *egyenletnek pontosan egy folytonos (t szerint differenciálható) megoldása van.*

Bizonyítás. Ebben az esetben teljesülnek az (i), (ii), (iii) és (iv) feltételek. Az is könnyen ellenőrízhető, hogy az (v) feltételben megadott B pontosan akkor véges, ha érvényes (2.2). Így a 2.1. Tételből következően az egyenletnek létezik és egyértelmű a megoldása, ami egyúttal sűrűségfüggvény is minden t-re a $[0, X_M]$ intervallumban. □

3. Numerikus módszer

Az (1.1) egyenletnek a pontos megoldása csupán néhány speciális esetben ismert [3]. Éppen ezért előtérbe került az egyenlet numerikus megoldása. Azonban, tekintettel arra, hogy sűrűségfüggvényt közelítünk, ezért olyan numerikus módszert keresünk, amely egy, a közelítés interpretációjához szükséges fontos kívánalomnak is eleget tesz. Nevezetesen a 2.1. Tételben szereplő $\int_{X_0}^{X_M} v(x,t)dx = 1$ egyenlőség fennállásához hasonló módon megkívánjuk a konvergencián és a nemnegativitáson felül, hogy a numerikus megoldásra is teljesüljön egy hasonló (diszkrét) megmaradási tétel. Ugyanis ezzel a folyamat diszkrét modelljéhez jutunk.

A cikk e fejezetében egy ilyen módszert ismertetünk.

A numerikus módszer leírásához vezessük be a következő jelöléseket:

Legyenek N, M természetes számok, $\tau := \frac{T}{M}$, $h := \frac{X_M - X_0}{N}$, $t_m := m \cdot \tau$, $x_i = X_0 + ih$, $S_i := S(x_i)$, $b_{ij} := b(x_i, x_j)$ és $v_{im} := v(x_i, t_m)$, ahol $m = 0, 1, \ldots, M$ és $i, j = 0, 1, \ldots, N$.

Az (1.1) egyenlet diszkretizálásából kapjuk az eljárást. A diszkretizációt a következőképpen hajtjuk végre: a t szerinti deriváltat differenciahányadossal, az

integrált pedig (a némileg módosított) trapéz formulával közelítjük. Így a következő közelítő egyenlőséghez jutunk ($m = 0, 1, \ldots, M-1$):

$$
\begin{aligned}
v_{0m+1} &\approx v_{0m} + \tau \sum_{j=1}^{N} S_j b_{0j} v_{jm} \omega_{0j} h \\
v_{im+1} &\approx v_{im} - \tau S_i v_{im} + \tau \sum_{j=i}^{N} S_j b_{ij} v_{jm} \omega_{ij} h \quad (i = 1, 2, \ldots, N-1) \\
v_{Nm+1} &\approx v_{Nm} - \tau S_N v_{Nm}
\end{aligned} \tag{3.1}
$$

ahol

$$
\omega_{ij} = \begin{cases} 1/2, & \text{ha } i = j, \text{ vagy } j = N \\ 1, & \text{ha } i < j \end{cases} \quad (i = 0, 1, \ldots, N-1, \ j = 1, 2, \ldots, N-1).
$$

(Később, az 5. részben tárgyalni fogjuk az elkövetett hiba nagyságrendjét!) A fenti megadásból származó eljárás így írható fel:

$$
\begin{aligned}
y_{i0} &= v_0(x_i) && (i = 0, 1, \ldots, N) \\
y_{0m+1} &= y_{0m} + \tau \sum_{j=1}^{N} S_j b_{0j} y_{jm} \omega_{0j} h \\
y_{im+1} &= y_{im} - \tau S_i y_{im} + \tau \sum_{j=1}^{N} S_j b_{ij} y_{jm} \omega_{ij} h \\
& && (i = 1, \ldots, N-1) \\
y_{Nm+1} &= y_{Nm} - \tau S_N y_{Nm} .
\end{aligned} \tag{3.2}
$$

Ekkor azonban nem biztosítható, hogy az $\{y_{im}\}_{i=0}^{N}$ értékekre teljesül az $\int_{X_0}^{X_M} v(x,t)dx = 1$ egyenlőségnek megfelelő $\sum_{i=0}^{N} y_{im} \gamma_i h$ diszkrét megmaradási tétel. (Esetünkben $\gamma_0 = \gamma_N = \frac{1}{2}$ és $\gamma_1 = \gamma_2 = \cdots = \gamma_{N-1} = 1$, annak megfelelően, hogy a diszkretizálásnál is trapéz formulát használtunk.) Ezért a következő módosítást végezzük a (3.2) rekurzión.

Legyen

$$
\tilde{\omega}_{ij} = \begin{cases} 1/2, & \text{ha } i = j, \text{ vagy } i = 0 \\ 1, & \text{ha } i < j, \ i = 1, \ldots, j; \ j = 1, 2, \ldots, N . \end{cases}
$$

Legyen továbbá

$$\tilde{b}_{ij} = \begin{cases} \dfrac{b_{ij}}{\sum_{k=0}^{j} b_{kj}\tilde{\omega}_{kj}h}, & \text{ha } j \neq 0, N, \quad i = 0, 1, \ldots, j \\ \dfrac{b_{iN}}{\sum_{k=0}^{N-1} b_{kN}\tilde{\omega}_{kN}h}, & \text{ha } j = N, \quad i = 0, 1, \ldots, N. \end{cases} \tag{3.3}$$

Ekkor a módosított rekurziót mátrix alakban a következőképpen írhatjuk:

$$\tilde{y}_0 = \frac{1}{\sum_{i=0}^{N} v_0(x_i)\gamma_i h} \cdot v_0; \tag{3.4}$$

$$\tilde{y}_{m+1} = D \cdot \tilde{y}_m \qquad (m = 0, 1, \ldots, M-1),$$

ahol

$$D = (d_{ij})_{(N+1)\times(N+1)} \qquad \text{és}$$

$$d_{ij} = \begin{cases} 1 - \tau S_i + \tau S_i \tilde{b}_{ii}\omega_{ii}h, & \text{ha } i = j \text{ és } i \neq N \text{ és } i \neq 0 \\ 1, & \text{ha } i = j = 0 \\ \tau S_j \tilde{b}_{ij}\omega_{ij}h, & \text{ha } i < j \\ 1 - \tau S_N, & \text{ha } i = j = N \\ 0, & \text{ha } i > j. \end{cases}$$

4. A diszkrét modell tulajdonságai

Bebizonyítjuk, hogy a kapott $\tilde{y}_{im}$ értékekre teljesül a diszkrét megmaradási tétel.

Legyen $S := \max_{x\in[X_0, X_M]} S(x)$.

4.1. TÉTEL. *Tegyük fel, hogy teljesül az* (ii) *feltétel. Ekkor a* (3.4) *képlettel megadott* $\tilde{y}_{im}$ *értékekre fennáll, hogy* $\sum_{i=0}^{N} \tilde{y}_{im}\gamma_i h = 1$ $(m = 0, 1, \ldots, M)$, *továbbá minden* $\tilde{y}_{im}$ *érték nemnegatív, ha* $\tau \leq \frac{1}{S}$.

Bizonyítás. Az állítás $m = 0$-ra triviális. Ezek után feltehetjük, hogy $k = 0, 1, \ldots, m$-re fennáll, hogy $\sum_{i=0}^{N} \tilde{y}_{ik}\gamma_i h = 1$. Bizonyítsuk $(m+1)$-re!

$$\sum_{i=0}^{N} \tilde{y}_{im+1}\gamma_i h = \sum_{i=0}^{N} \tilde{y}_{im}\gamma_i h - \tau \sum_{i=1}^{N} S_i \tilde{y}_{im}\gamma_i h + \sum_{N}, \quad \text{ahol}$$

$$\sum_N := \tau\gamma_0 h \sum_{j=1}^{N} S_j \tilde{y}_{jm} \tilde{b}_{0j}\omega_{0j} h + \tau \sum_{i=1}^{N-1} \gamma_i h \sum_{j=1}^{N} S_j \tilde{y}_{jm} \tilde{b}_{ij}\omega_{ij} h =$$

$$= \tau\gamma_0 h \sum_{j=1}^{N} S_j \tilde{y}_{jm} \tilde{b}_{0j}\omega_{0j} h + \tau \sum_{j=1}^{N-1} \sum_{i=1}^{j} \gamma_i h S_j \tilde{y}_{jm} \tilde{b}_{ij}\omega_{ij} h +$$

$$+ \tau \sum_{i=1}^{N-1} \gamma_i h S_N \tilde{y}_{Nm} \tilde{b}_{iN}\omega_{iN} h = \tau \sum_{j=1}^{N-1} \sum_{i=0}^{j} \gamma_i h S_j \tilde{y}_{jm} \tilde{b}_{ij}\omega_{ij} h +$$

$$+ \tau \sum_{i=0}^{N-1} \gamma_i h S_N \tilde{y}_{Nm} \tilde{b}_{iN}\omega_{iN} h .$$

Tehát igaz a következő egyenlőség:

$$(4.1) \quad \sum_{i=0}^{N} \tilde{y}_{im+1}\gamma_i h = \sum_{j=0}^{N} \tilde{y}_{jm}\gamma_j h - \sum_{j=1}^{N-1} \tilde{y}_{jm} S_j \tau h \left(\gamma_j - \sum_{i=0}^{j} \gamma_i \tilde{b}_{ij}\omega_{ij} h \right) -$$

$$- \tilde{y}_{Nm} S_N \tau h \left(\gamma_N - \sum_{i=0}^{N-1} \gamma_i \tilde{b}_{iN}\omega_{iN} h \right) .$$

Könnyen látható, hogy ha $j = 1, \ldots, N-1$ és $i = 0, \ldots, N-1$, akkor $\gamma_i\omega_{ij} = \tilde{\omega}_{ij}$, továbbá az is igaz, hogy ha $i = 0, \ldots, N-1$, akkor $\gamma_i\omega_{iN} = \frac{1}{2}\tilde{\omega}_{iN}$. Ezért $j = 1, \ldots, N-1$ esetén

$$(4.2) \quad \sum_{i=0}^{j} \gamma_i \tilde{b}_{ij}\omega_{ij} h = \sum_{i=0}^{j} \gamma_i \frac{b_{ij}\omega_{ij} h}{\sum_{k=0}^{j} b_{kj}\tilde{\omega}_{kj} h} = 1 = \gamma_j \quad \text{és}$$

$$(4.3) \quad \sum_{i=0}^{N-1} \gamma_i \tilde{b}_{iN}\omega_{iN} h = \sum_{i=0}^{N-1} \gamma_i \frac{b_{iN}\omega_{iN} h}{\sum_{k=0}^{N-1} b_{kN}\tilde{\omega}_{kN} h} = \frac{1}{2} = \gamma_N .$$

Ez azonban maga után vonja, hogy $\sum_{i=0}^{N} \tilde{y}_{im+1}\gamma_i h = \sum_{i=0}^{N} \tilde{y}_{im}\gamma_i h = 1$, ez pedig a tétel első állítását adja.

Másrészt, ha $\tau \leq \frac{1}{5}$, akkor a D mátrix nemnegatív, amiből következik, hogy $\tilde{y}_{im}$ is nemnegatív. □

KÖVETKEZMÉNY. *Minden m-re az $\tilde{y}_{im}\gamma_i h$ értékek egy diszkrét eloszlás súlyai.*

Megjegyzés. A D mátrix oszlopösszegnormáját a fentiek alapján könnyen kiszámíthatjuk. Legyen $\tau \leq \frac{1}{5}$, azaz a mátrix minden eleme nemnegatív. Ekkor a D

mátrix j–edik oszlopában lévő elemek összege, D_j a következő:

$$D_j = \begin{cases} 1, & \text{ha } j = 0 \\ 1 - \tau S_j + \tau \sum_{i=0}^{j} S_j \tilde{b}_{ij} \omega_{ij} h, & \text{ha } j = 1, \dots, N-1 \\ 1 - \tau S_N + \tau \sum_{i=0}^{N-1} S_N \tilde{b}_{iN} \omega_{iN} h, & \text{ha } j = N. \end{cases}$$

Felhasználva a (4.2) és (4.3) egyenlőségeket, megkapjuk, hogy

$$D_j = 1 + \frac{\tau}{2} \tilde{b}_{0j} h S_j \quad (j = 1, 2, \dots, N-1) \quad \text{és} \quad D_N = 1 - \frac{\tau}{2} S_N + \frac{\tau}{4} S_N \tilde{b}_{0N} h.$$

Így D oszlopösszegnormája $\|D\|_1 = \max_{j=0,1,\dots,N} D_j \leq 1$, ha a $b(x,y)$ függvény olyan, hogy $b(X_0, y) = 0$. Például a (2.2)–ben említett speciális esetben $p > 1$ esetén ez fennáll.

A módszer tárigénye $O(N^2)$ és a műveletigénye az m–edik vektor kiszámításához szükséges műveleteket tekintve $O(mN^2)$.

A módszer aszimptotikus viselkedéséről a következőt mondhatjuk:

4.2. TÉTEL. *Legyen* $S(x) \geq 0$, $b(x,y) \geq 0$ *és* $S(x) = b(x,y) = 0$ *pontosan akkor, ha* $x = X_0$. *Továbbá legyen* $\tau \leq \frac{1}{S}$. *Ekkor létezik az* $\tilde{y}_\infty = \lim_{m \to \infty} \tilde{y}_m$ *és* $\tilde{y}_\infty = \left(\tilde{y}_{00}, \frac{1}{h} - \frac{1}{2}\tilde{y}_{00}, 0, \dots, 0\right)^T$.

Bizonyítás. A D mátrix felsőháromszög-mátrix, ezért a főátlóbeli elemek lesznek D sajátértékei. Jelöljük a sajátértékeket rendre $\lambda_1, \lambda_2, \dots, \lambda_\ell$–lel, és legyen az egyes sajátértékek multiplicitása $k_1, k_2, \dots, k_\ell$. Ekkor D konstrukciójából fakadóan $\lambda_1 = 1$, $k_1 = 2$ és $\lambda_i \in [0,1)$, $i = 2, \dots, \ell$, azaz az 1 kétszeres sajátérték és a többi sajátérték 1–nél kisebb. Könnyen látható, hogy a $v_0 = (1, 0, \dots, 0)^T$ és a $v_1 = (0, 1, 0, \dots, 0)^T$ vektorok D–nek a $\lambda_1 = 1$ sajátértékekhez tartozó sajátvektorai.

Mint ismert ([6], 179. o.), a D mátrix felírható a következő alakban: $D = V_D J_D V_D^{-1}$, ahol a V_D mátrix a λ_i sajátértékekhez tartozó sajátvektorokból illetve fővektorokból — jelöljük most ezeket $v_0, v_1, v_2, \dots, v_N$–nel — álló mátrix, míg J_D a D mátrix Jordan-féle normálalakja. J_D a következőképpen írható fel:

$$J_D = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & & & \\ \vdots & \vdots & & B_D & \\ 0 & 0 & & & \end{bmatrix}, \text{ ahol } B_D \ (N-1) \times (N-1)\text{-es mátrix}.$$

Mivel B_D spektrálsugara, $\varrho(B_D) := \max_{i=2,\dots,\ell} \lambda_i < 1$, ezért mint ismert, $m \to \infty$ esetén B_D^m tart a nullmátrixhoz.

Legyen most

$$\underline{u}_m := V_D^{-1}\tilde{\underline{y}}_m = (u_{0m}, u_{1m}, \ldots, u_{Nm})^T .$$

Ekkor az $\tilde{\underline{y}}_{m+1} = D\tilde{\underline{y}}_m$ egyenlőségből következik, hogy $\underline{u}_{m+1} = J_D\underline{u}_m$. Ezt többször alkalmazva adódik, hogy

$$\underline{u}_{m+1} = J_D^{m+1}\underline{u}_0 ,$$

amit a következő alakban is írhatunk:

$$\underline{u}_{m+1} = J_D^{m+1}\underline{u}_0 = (u_{00}, u_{10}, (B_D^{m+1}\underline{u}_0^{(3)})^T)^T ,$$

ahol $\underline{u}_0^{(3)} = (u_{20}, \ldots, u_{N0})^T$. Az előbb elmondottak alapján $\lim\limits_{m\to\infty} B_D^{m+1}\underline{u}_0^{(3)} = \underline{0}$, ezért létezik az

$$\underline{u}_\infty := \lim_{m\to\infty} \underline{u}_{m+1} = (u_{00}, u_{10}, 0, \ldots, 0)^T \quad \text{határérték} .$$

Így létezik az

$$\tilde{\underline{y}}_\infty := \lim_{n\to\infty} \tilde{\underline{y}}_m \quad \text{is, és} \quad \tilde{\underline{y}}_\infty V_D\underline{u}_\infty = u_{00}\underline{v}_0 + u_{10}\underline{v}_1 = (u_{00}, u_{10}, 0, \ldots, 0)^T .$$

Megadható az $\tilde{\underline{y}}_\infty$ és az $\tilde{\underline{y}}_0$ között egy kapcsolat, ugyanis

$$\tilde{\underline{y}}_0 = V_D\underline{u}_0 = u_{00}\underline{v}_0 + u_{10}\underline{v}_1 + \cdots + u_{N0}\underline{v}_N ,$$

amiből látható, hogy $\tilde{\underline{y}}_0$ előállításában is, és $\tilde{\underline{y}}_\infty$ előállításában is ugyanaz $\underline{v}_0$ és $\underline{v}_1$ együtthatója. Tekintettel arra, hogy a D mátrix első sorában a főátlón lévő elem kivételével minden elem 0, ezért minden m-re $\tilde{y}_{0m} = \tilde{y}_{00}$, tehát határértékben $\tilde{y}_{0\infty} = u_{00} = \tilde{y}_{00}$.

Végezetül pedig, mivel minden m-re $\sum\limits_{i=0}^{N} \tilde{y}_{im}\gamma_i h = 1$, így a határátmenet miatt $\left(\frac{1}{2}u_{00} + u_{10}\right)h = 1$, azaz $\tilde{\underline{y}}_\infty = \left(\tilde{y}_{00}, \frac{1}{h} - \frac{1}{2}\tilde{y}_{00}, 0, \ldots, 0\right)^T$.

Ezzel a tétel állítását maradéktalanul bebizonyítottuk. □

Megjegyzés. Amennyiben az (1.1) egyenletnek létezik megoldása és fennállnak a 4.2. Tétel feltételei, könnyen igazolhatóan teljesül, hogy $t \in [0, T]$-re $v(X_0, t) = v_0(X_0)$. Ez szoros analógiát mutat az előző tétel eredményével.

5. Stabilitás és konvergencia

Az $S(x)$, $b(x,y)$, $v_0(x)$ függvényekre vonatkozó bizonyos simasági feltételek mellett (ld. [7]) és $\tau \leq \frac{1}{S}$ esetén a (3.4) rekurzióval megadott módszer numerikusan stabil minden véges T-intervallumon. Nevezetesen igaz:

$$\|D\|_\infty \leq 1 + C\tau\,, \tag{5.1}$$

ahol a C konstans az N-től és M-től független.

További simasági feltételek ([7]) teljesülése mellett a módszer konvergens: $m = 0, 1, \ldots, M$ és $i = 0, 1, \ldots, N$ esetén igaz, hogy:

$$|v(x_i, t_m) - \tilde{y}_{im}| \leq C_1 h^2 + C_2 \tau\,,$$

ahol $\tilde{y}_{im}$ a (3.4) rekurzióval számolható közelítő érték és a C_1, C_2 konstansok az N-től és M-től függetlenek.

Ugyanis felhasználva a simasági tulajdonságokat, becsülhetjük a t szerinti derivált és a differenciahányados, valamint az integrál és a trapézformula eltérését. Ezeket összegezve a következő formulát kapjuk a függvényérték és a közelítőérték eltérésének normájára:

$$\|\tilde{z}_{m+1}\|_\infty := \|v_{m+1} - \tilde{z}_{m+1}\|_\infty \leq \|D\|_\infty \|\tilde{y}_m\|_\infty + (C_3 h^2 + C_4 \tau)\tau\,, \tag{5.2}$$

ahol a C_3, C_4 konstansok N-től, M-től függetlenek.

Az (5.2) becslés levezetése során használtuk azt is, hogy a 2.1 Tétel (iii) feltétele miatt a (3.3) képletek nevezője $1 + O(h^2)$. (5.2)-ből kapjuk a következő egyenlőtlenséget:

$$\|\tilde{z}_{m+1}\|_\infty \leq \|D\|_\infty^{m+1} \cdot \|\tilde{z}_0\|_\infty + \sum_{i=0}^{m} \|D\|_\infty^i (C_3 h^2 + C_4 \tau)\tau\,. \tag{5.3}$$

Továbbá belátható a $v_0(x)$ függvényre vonatkozó simasági feltétel alapján, hogy

$$\|\tilde{z}_0\|_\infty \leq C_5 h^2\,, \tag{5.4}$$

ahol a C_5 konstans N-től, M-től független. Összegezve (5.1)-(5.4)-et:

$$\begin{aligned}\|\tilde{z}_{m+1}\|_\infty &\leq (1 + C\tau)^{m+1} \cdot C_5 h^2 + \sum_{i=0}^{m} (1 + C\tau)^i (C_3 h^2 + C_4 \tau)\tau \leq \\ &\leq e^{CT} \cdot \left[C_5 h^2 + T \cdot C_3 h^2 + T \cdot C_4 \tau\right] = C_1 h^2 + C_2 \tau\,.\end{aligned}$$

6. Numerikus eredmények

A módszer pontosságát számítógépes futtatásokkal is ellenőríztük. A szakirodalomban legtöbbször előforduló (2.1) esetből származó néhány példán keresztül illusztráljuk a módszer pontosságát. A $p = q$ esetet vizsgáltuk, mert ekkor a megoldás pontos alakja is ismert, [3] nevezetesen $v(x,t) = e^{-kx^p t}(v_0(x) + kpt\,x^{p-1}R_0(x)$, ahol $R_0(x) := \int_x^{X_M} v_0(z)dz$. A számítások eredményei alátámasztják a fentebb megadott hibabecsléseket.

Az 1. Táblázat a diszkrét maximum normában vett különbségeket tartalmazza, különböző $p(= q)$ esetén. Tekintettel arra, hogy az elméletileg bizonyított konvergencia $C_1h^2+C_2\tau$, ezért $\tau = h^2$ választással futtattuk a módszert néhány h értékre. Mint a táblázat is mutatja, másodrendű konvergenciát kaptunk. A $C := C_1 + C_2$ szorzótényező a $p = q = 2$ esetben $\approx 1,4$, $p = q = 3$ esetben $\approx 1,0$ és a $p = q = 4$ esetben $\approx 0,9$.

1. Táblázat:
A trapéz-szabályon alapuló módszer szerint számolt és a pontos értékek eltérése diszkrét maximum normában
($k = 1$, $v_0(x) = 6x(1-x)$, $T = X_M = 1$)

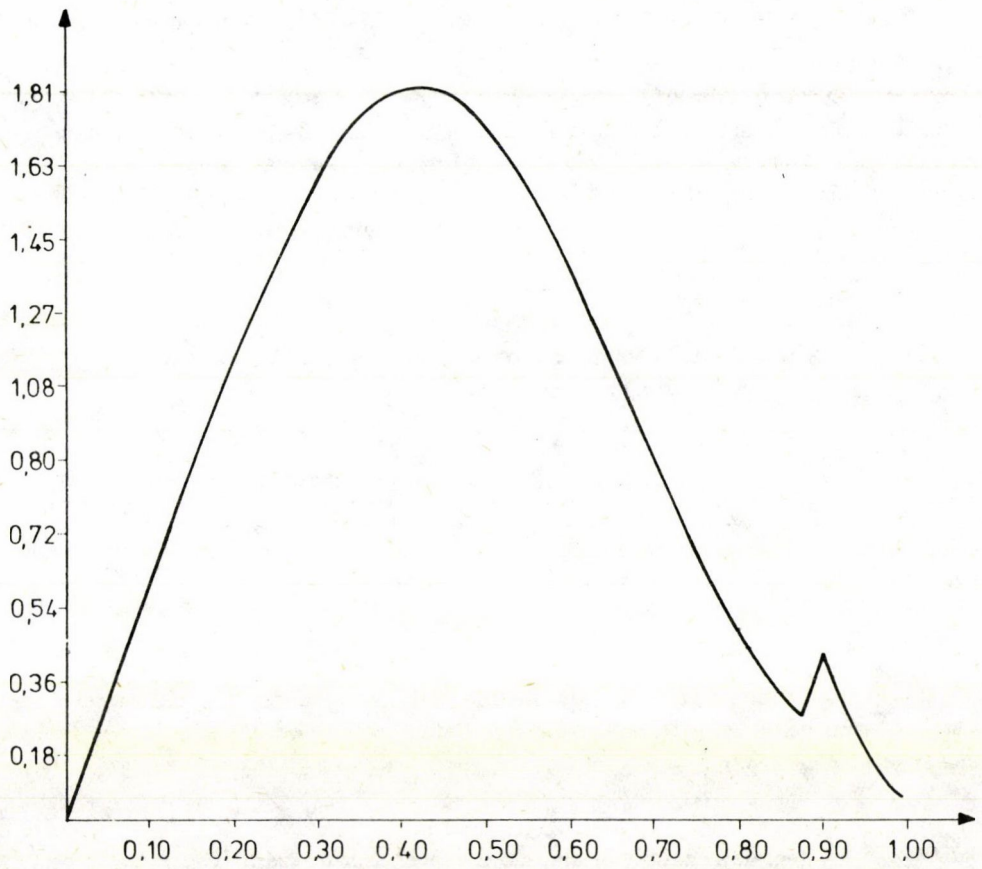
	$\tau = h^2 = 1/400 =$ $= 2,5 \cdot 10^{-3}$	$\tau = h^2 = 1/1600 =$ $= 6,25 \cdot 10^{-4}$	$\tau = h^2 = 1/6400 =$ $= 1,5625 \cdot 10^{-4}$	$\tau = h^2 = 1/14400 \approx$ $\approx 6,944 \cdot 10^{-5}$
$p = q = 2$	$3,56 \cdot 10^{-3}$	$8,91 \cdot 10^{-4}$	$2,23 \cdot 10^{-4}$	$9,90 \cdot 10^{-5}$
$p = q = 3$	$2,61 \cdot 10^{-3}$	$6,54 \cdot 10^{-4}$	$1,63 \cdot 10^{-4}$	$7,26 \cdot 10^{-5}$
$p = q = 4$	$2,37 \cdot 10^{-3}$	$5,89 \cdot 10^{-4}$	$1,47 \cdot 10^{-4}$	$6,54 \cdot 10^{-5}$

Összehasonlításképpen a 2. Táblázatban közöljük ugyanolyan paraméterek mellett egy egyszerűbb — a téglalap szabályon alapuló — módszer által kapott eredményeket.

2. Táblázat:
A téglalap-szabályon alapuló módszer szerint számolt és a pontos értékek eltérése diszkrét maximum normában
($k = 1$, $v_0(x) = 6x(1-x)$, $T = X_M = 1$)

	$\tau = h^2 = 1/400 =$ $= 2,5 \cdot 10^{-3}$	$\tau = h^2 = 1/1600 =$ $= 6,25 \cdot 10^{-4}$	$\tau = h^2 = 1/6400 =$ $= 1,5625 \cdot 10^{-4}$	$\tau = h^2 = 1/14400 \approx$ $\approx 6,944 \cdot 10^{-5}$
$p = q = 2$	$2,25 \cdot 10^{-2}$	$1,25 \cdot 10^{-2}$	$6,50 \cdot 10^{-3}$	$4,41 \cdot 10^{-3}$
$p = q = 3$	$2,25 \cdot 10^{-2}$	$1,14 \cdot 10^{-2}$	$5,70 \cdot 10^{-3}$	$3,82 \cdot 10^{-3}$
$p = q = 4$	$2,24 \cdot 10^{-2}$	$1,16 \cdot 10^{-2}$	$5,90 \cdot 10^{-3}$	$3,94 \cdot 10^{-3}$

Végezetül két ábrát mutatunk be.



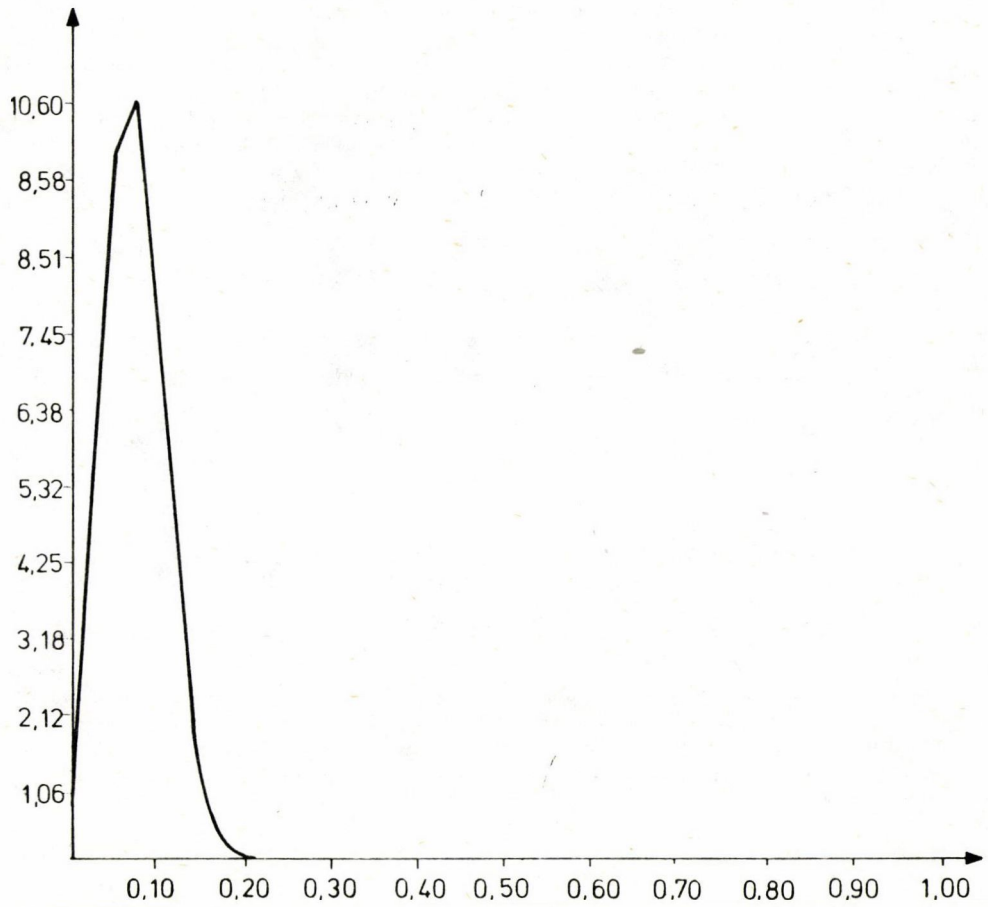
1. ábra: A beadagoláskor nagy szemcsékből álló anyag időbeli változása ($T = 5$)

Az 1. ábrán egy, a gyakorlat számára is fontos jelenséget figyelhetünk meg. Nevezetesen azt, hogy egy olyan kezdeti sűrűségfüggvényből kiindulva, ami ugrófüggvény, időben haladva olyan sűrűségfüggvényhez jutunk, ami egyrészt sima, másrészt eltűnik belőle a kezdeti ugrás okozta nagyméretű eltérés.

A kiindulási sűrűségfüggvény az alábbi:

$$v_0(x) = \begin{cases} 0, & \text{ha } x < \frac{7}{8} \\ 8, & \text{ha } x \geq \frac{7}{8}. \end{cases}$$

A fenti sűrűségfüggvény azt a gyakorlati esetet reprezentálja, amikor a malomba betett anyag homogén eloszlású, de csak nagy szemcsékből áll. Az ábra a $T = 5$ időponthoz tartozó függvénygörbét ábrázolja. A további paraméterek: $p = 2$, $q=3$, $k = 1$, $\tau = h^2 = 0,000625$.



2. ábra: A függvényértékek aszimptotikus viselkedése ($T = 10^3$)

A 2. ábrán a módszer által számolt függvényértékek aszimptotikus viselkedését szemléltetjük.

A 4.2. Tételben bizonyítottuk, hogy a módszer szerint számolt közelítő megoldás határértéke $m \to \infty$ esetén csak az első két koordinátában különbözhet 0-tól. Ezt támasztják alá a számolt függvényértékek is. Esetünkben a paraméterek a következők voltak: $v_0(x) \equiv 1$, $p = 2$, $q = 3$, $k = 1$, $\tau = 1$ és $h = 0,025$ és az ábra a $T = 10^3$ időpillanathoz tartozó függvénygörbét ábrázolja.

IRODALOM

[1] L. G. AUSTIN, R. R. KLIMPEL, „The theory of grinding operations", *Industrial and Engineering Chemistry* **56.**, No. 11, (1964).

[2] V. K. GUPTA, P. C. KAPUR, „A critical appraisal of the discrete size models of grinding kinetics", in Zerkleinern Dechema-Monor, vol. 79 (1976), 447–465.

[3] Y. NAKAJIMA, T. TANAKA, „Solution of batch grinding equation", *Ind. Eng. Process Des. Develop.* **12.**, No 1, (1973).

[4] E. VARGA, B. LAKATOS, „Solution of the batch grinding equation via orthogonal collocation on finite elements", *Proceedings of 5th CAC*, vol. 2 (1989).

[5] T. BLICKLE, S. VERDES, „Model of batch grinding", *Proceedings of 5th CAC*, vol. 2 (1989).

[6] G. MAESS, *Vorlesungen über numerische Mathematik* (Akademie Verlag, Berlin, 1984).

[7] MIHÁLYKÓ CSABA, *A golyósmalmi szakaszos őrlés matematikai modellje*, Diplomadolgozat (ELTE TTK, Budapest, 1992).

(Beérkezett: 1992. november 11.)

MIHÁLYKÓ CSABA
VESZPRÉMI EGYETEM, MATEMATIKAI ÉS SZÁMÍTÁSTECHNIKAI TANSZÉK
8200 VESZPRÉM, EGYETEM U. 10.

SOLUBILITY OF INTEGRODIFFERENTIAL EQUATION FOR THE DENSITY FUNCTION OF BALL MILL GRANULATE, AND SPECIAL PROPERTIES OF SOLUTION

CS. MIHÁLYKÓ

A mathematical model of batch grinding in a ball mill is presented. The integrodifferential equation for the density function of the mass distribution of the particle size is discussed. We conclude that the unique solution of the equation is a density function. Moreover, for the approximating computation, we propose a convergent method that yields a solution satisfying the mass conservation law.

Alkalmazott Matematikai Lapok **17** (1993) 185–194

AZ AFFIN SKÁLÁZÁSI ALGORITMUS MÓDOSÍTÁSAIRÓL*

MÉSZÁROS CSABA

Budapest

A lineáris programozás (LP) számítógépes módszerei között a belső pontos algoritmusok, ezen belül a Karmarkar-típusú módszerek egyre nagyobb teret követelnek maguknak. A témában a közelmúltban több hatékony implementáció is született [1,7,8]. Az LP feladat skálázott feltételi mátrixának nullterére történő vetítés kiszámítása minden Karmarkar-típusú belső pontos algoritmus fő lépése. A dolgozatban erre a lépésre egy eddig még a belső pontos irodalomban nem használt módszert javasolunk, és ezen keresztül megmutatjuk az affin skálázási algoritmus néhány implementációban nagyon hasznos módosítását. A vetítés elvégzésére eddig a legtöbb módszer az $AD_x^2A^T \cdot y = AD_x^2c$ normálegyenletrendszer valamilyen formában történő megoldását használta. Ennek a módszernek a hátránya akkor jelentkezik, ha egy nagy méretű, ritka feltételi mátrixú LP feladatban van sűrű, ill. teljesen kitöltött oszlop. Ekkor a normálegyenletrendszer feltételi mátrixa sűrű, ill. tejesen kitöltött lesz, amely mindenképpen hátrányos. A jelenséget csak speciális technikák alkalmazásával kerülhetjük el, ilyen pl. CHANDRU és KOCHAR eredménye [4] az induló megengedett belső pont keresésére. Módszerünkkel, mely speciálisan a normálegyenletrendszeren keresztül történő megoldást is magában foglalja, ilyen technikákat alkalmazhatunk, melyek különböző pivotválasztási szabályok előírásával adhatók meg.

1. Bevezetés

A Karmarkar algoritmus affin skálázási változatát [2] egyenlőséges feltételekkel megfogalmazott lineáris programozás (LP) feladatra alkalmazzuk:

$$\text{(P)} \qquad \begin{aligned} &\min c \cdot x \\ &Ax = b \\ &x \geq 0\,. \end{aligned}$$

A továbbiakban fetételezzük, hogy az $A \in \mathbb{R}^{m\times n}$ mátrix teljes sorrangú, és rendelkezésre áll egy megengedett x_0 belső pont, azaz: $x_0 > 0$ és $Ax_0 = b$. Legyen a továbbiakban $D_x = \text{diag}(x)$, és $0 < \alpha < 1$ rögzített lépéshossz paraméter. Az algoritmus egy iterációjának lépései röviden a következők [2,11]:

Algoritmus A.1

(1.1) *Duál változók számítása*

$y = \arg\min \|D_xA^Ty - D_xc\|_2$.

*A dolgozat a 2587 és a 2116 számú OTKA szerződések részbeni támogatásával készült.

(1.2) *A kereső irány kiszámítása*
$z = D_x^2(c - A^T y)$.

(1.3) *A lépéshossz meghatározása**

$$h = \frac{\alpha}{\max\limits_{1 \le i \le n} \frac{z_i}{x_i}} .$$

(1.4) *A megállási kritérium ellenőrzése*

(1.5) *Az új megengedett belső pont kiszámítása*
$x^{uj} = x - h \cdot z$.

Az algoritmus (1.1) lépése igényli a legtöbb számítási időt, számítógépes implementációkban a teljes futási idő közel 90 %-át [7]. Belső pontos algoritmusok implementációiban (1.1) megoldására három különböző módszert használtak [6]:

a) A $D_x A^T = QR$ felbontást, ahol $Q \in \mathbb{R}^{n \times n}$ ortogonális mátrix, és $R \in \mathbb{R}^{m \times n}$ pedig a következő alakú: $R = \begin{bmatrix} U \\ 0 \end{bmatrix}$, ahol $U \in \mathbb{R}^{m \times m}$ felső háromszögmátrix. Ekkor (1.1) megoldása a következő trianguláris egyenletrendszer megoldására vezet: $Uy = \hat{c}_1$, ahol $\begin{bmatrix} \hat{c}_1 \\ \hat{c}_2 \end{bmatrix} = Q^T D_x c$. A módszert pl. a [10] implementációban használták.

b) A feladat normálegyenletrendszerének, azaz az $AD_x^2 A^T \cdot y = AD_x^2 c$ egyenlet megoldását. Az egyenletrendszer megoldásához először az $AD_x^2 A^T = LL^T$ Cholesky-felbontást elkészítve két trianguláris egyenletrendszer megoldását kell elvégezni: $Lv = AD_x^2 c$ és $L^T y = v$. A módszert pl. a [8] implementációban használták.

c) A „hibrid" konjugált gradiens módszert a normálegyenletrendszer megoldására. Mivel az $AD_x^2 A^T$ mátrix szimmetrikus, pozitív definit, a konjugált gradiens módszer minden esetben a megoldáshoz konvergál. A konvergencia sebessége azonban nagyban függ a mátrix kondíciószámától. A konvergenciasebesség növelése érdekében az $AD_x^2 A^T$ mátrixra részleges Cholesky-felbontást lehet végezni, azaz $A_N D_x^2 A_N^T = LL^T$ ahol A_N az A oszlopaiból álló nemszinguláris részmátrix. Ekkor az

$$L^{-1} A D_x^2 A^T (L^T)^{-1} v = L^{-1} A D_x^2 c$$

egyenletrendszert kell a konjugált gradiens módszerrel megoldani, majd az $L^T y = v$ trianguláris egyenletrendszert visszahelyettesítéssel. A módszert pl. az [1] implementációban használták.

*A leírásban itt Vanderbei módszerét [11] használtuk. A hányadostesztnél a megengedett poliéder határáig léphetünk el, míg Barns módszerével [2] a poliéderbe írt gömb határáig.

A dolgozatban az algoritmus (1.1) és (1.2) lépését egyszerre végezzük el egy szimmetrikus egyenletrendszer megoldásával:

$$\begin{bmatrix} I & D_x A^T \\ AD_x & 0 \end{bmatrix} \begin{bmatrix} \hat{z} \\ y \end{bmatrix} = \begin{bmatrix} D_x c \\ 0 \end{bmatrix}, \tag{1.6}$$

ahol $z = D_x \hat{z}$. Könnyen látható, hogy az (1.6) megoldásából származó z, y megoldása egyben (1.1) és (1.2)-nek. A módszert a legkisebb négyzetekkel kapcsolatos problémák körében már jó tapasztalatokkal alkalmazták [3]. Könnyen látható, hogy (1.6) a következő ekvivalens formára írható:

$$\begin{bmatrix} D_x^{-2} & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} z \\ y \end{bmatrix} = \begin{bmatrix} c \\ 0 \end{bmatrix}. \tag{1.7}$$

A továbbiakban ezt az alakot fogjuk alkalmazni. Az (1.7) egyenletrendszer megoldásakor használhatunk LU dekompozíciót, azaz a mátrix alsó és felső háromszögmátrixra való felbontását. Ennek egy speciális esete, amikor csak a diagonálisban keresünk pivotelemeket. Ekkor LDL^T felbontást kapunk, ahol L alsó háromszögmátrix, D pedig diagonális mátrix. Megjegyezzük, hogy (1.7) ilyen megoldása mellett speciálisan megkaphatjuk a fentiekben említett *b)* módszert, ha először a D_x^{-2} blokk diagonálisában pivotálunk, majd ezen n lépés után keletkező m egyenletből álló egyenletrendszer éppen az $AD_x^2 A^T \cdot y = AD_x^2 c$ normálegyenletrendszert szolgáltatja.

A második fejezetben felsőkorlátos változók, a harmadikban nem korlátozott változók esetével foglalkozunk. Memutatjuk, hogy VANDERBEI [11,12] módszerei hogyan származnak (1.7) megoldásának különböző pivotválasztási előírásaiból. Ugyanígy következik majd CHANDRU és KOCHAR elegáns módszere az első fázis algoritmusra [4], és ADLER, RESENDE, VEIGA és KARMARKAR „duál" affin skálázási módszerének [1], és a (P) duálisára alkalmazott eredeti algoritmusnak az ekvivalenciája. A negyedik fejezet a slack változók kezelésével, és a módszer implementációiban előnyösen használható egyéb tulajdonságaival foglalkozik.

2. Felső korlátos változók

Vizsgáljuk meg azt az esetet, amikor a (P) feladat változóira kétoldali feltételek vannak, azaz $u \geq x \geq 0$. Ekkor a feladat standard formában az s slack változókkal kiegészítve a következőképpen írható fel:

$$\min \begin{bmatrix} c \\ 0 \end{bmatrix} \begin{bmatrix} x \\ s \end{bmatrix},$$

$$\begin{bmatrix} A & 0 \\ I & I \end{bmatrix} \begin{bmatrix} x \\ s \end{bmatrix} = \begin{bmatrix} b \\ u \end{bmatrix},$$

$$\begin{bmatrix} x \\ s \end{bmatrix} \geq 0.$$

Ekkor az (1.7) egyenletrendszer a következő alakú :

$$\begin{bmatrix} D_x^{-2} & 0 & A^T & I \\ 0 & D_s^{-2} & 0 & I \\ A & 0 & 0 & 0 \\ I & I & 0 & 0 \end{bmatrix} \begin{bmatrix} z \\ z_s \\ y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} c \\ 0 \\ 0 \\ 0 \end{bmatrix}. \tag{2.1}$$

Mivel $s = u - x$, a kereső irány megfelelő komponensei között a következő összefüggés van: $z = -z_s$. Pivotáljunk először a D_s^{-2} diagonálisában, azaz eliminálják a z_s ismeretleneket. Ekkor (2.1)-ből a következő egyenletrendszert kapjuk:

$$\begin{bmatrix} D_x^{-2} & A^T & I \\ A & 0 & 0 \\ I & 0 & -D_s^2 \end{bmatrix} \begin{bmatrix} z \\ y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} c \\ 0 \\ 0 \end{bmatrix}. \tag{2.2}$$

Látható továbbá, hogy a változók felső korlátját kifejező feltételekhez tartozó „duális" változók a következőképpen fejezhetők ki: $y_2 = D_s^{-2} z$. Eliminálják most ezeket a változókat, azaz pivotáljunk a $-D_s^2$ blokk diagonálisában. Ekkor (2.2) a következő alakra transzformálódik:

$$\begin{bmatrix} D_{\hat{x}}^{-2} & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} z \\ y_1 \end{bmatrix} = \begin{bmatrix} c \\ 0 \end{bmatrix}, \tag{2.3}$$

ahol

$$\hat{x}_i = \frac{x_i s_i}{\sqrt{x_i^2 + s_i^2}}, \text{ azaz}$$

$$\hat{x}_i = \frac{x_i(u_i - x_i)}{\sqrt{x_i^2 + (u_i - x_i)^2}}. \tag{2.4}$$

A (2.4) előírás a felső korlátos változók módosított skálázására megegyezik VANDERBEI [11] eredményével. VANDERBEI tovább módosítja (2.4)-et, $\frac{x_i s_i}{\sqrt{x_i^2 + s_i^2}}$ helyett az

$$\hat{x}_i = \min(x_i, s_i) \tag{2.5}$$

előírást javasolja, mivel az egyszerűbb (2.4)-nél, és ha x_i nullához, vagy a felső korlátjához tart, (2.4) és (2.5) ugyanolyan arányban konvergál nullához.

A felső korlátos változók ilyen kezelése mellett a lépéshossz kiszámítását, és a megállási kritérium ellenőrzését is módosítani kell. A $z = -z_s$ összefüggés folytán az (1.3) lépéshossz esetünkben a következőképpen számítható:

$$h = \frac{\alpha}{\max\limits_{1 \le i \le n} \left\{ \frac{z_i}{x_i}, -\frac{z_i}{u_i - x_i} \right\}}.$$

Itt még megemlíthető, hogy a $b_1 \leq ax \leq b_2$ úgynevezett „range" típusú feltételek felső korlátos slack változókkal egyszerűen kezelhetők, mely már régóta ismert és általánosan használt technika a lineáris programozásban.

3. Nem korlátozott változók

Ebben a fejezetben azt az esetet vizsgáljuk meg, amikor nem korlátozott változóink is vannak. Jelölje a szabad változókat x_F, a negatív és pozitív részre való felbontásuk pedig legyen $x_F = x_F^+ - x_F^-$. A (P) feladat ekkor a következő formában írható fel:

$$\text{(PF)} \qquad \min \begin{bmatrix} c \\ c_F \\ -c_F \end{bmatrix} \begin{bmatrix} x \\ x_F^+ \\ x_F^- \end{bmatrix},$$

$$[A \quad F \quad -F] \begin{bmatrix} x \\ x_F^+ \\ x_F^- \end{bmatrix} = b,$$

$$\begin{bmatrix} x \\ x_F^+ \\ x_F^- \end{bmatrix} \geq 0.$$

Az (1.7) egyenletrendszer felírása most a következő lesz:

$$\text{(3.1)} \qquad \begin{bmatrix} D_x^{-2} & 0 & 0 & A^T \\ 0 & D_{x_F^+}^{-2} & 0 & F^T \\ 0 & 0 & D_{x_F^-}^{-2} & -F^T \\ A & F & -F & 0 \end{bmatrix} \begin{bmatrix} z \\ z_{x_F^+} \\ z_{x_F^-} \\ y \end{bmatrix} = \begin{bmatrix} c \\ c_F \\ -c_F \\ 0 \end{bmatrix}.$$

Legyen M_1 az a transzformáció, amelyet balról alkalmazva (3.1)-ben a $-F^T$ blokkot eliminálja oly módon, hogy az F^T blokk sorait a $-F^T$ blokk megfelelő soraihoz adja. Hasonlóan, vezessük be az M_2 transzformációt, melyet jobbról alkalmazva a $-F$ blokkot eliminálja úgy, hogy a $-F$ blokk oszlopaihoz az F blokk megfelelő oszlopait adja hozzá. Tekintsük a (3.1)-gyel ekvivalens következő egyenletrendszet:

$$\text{(3.2)} \qquad M_1 \begin{bmatrix} D_x^{-2} & 0 & 0 & A^T \\ 0 & D_{x_F^+}^{-2} & 0 & F^T \\ 0 & 0 & D_{x_F^-}^{-2} & -F^T \\ A & F & -F & 0 \end{bmatrix} M_2 M_2^{-1} \begin{bmatrix} z \\ z_{x_F^+} \\ z_{x_F^-} \\ y \end{bmatrix} = M_1 \begin{bmatrix} c \\ c_F \\ -c_F \\ 0 \end{bmatrix},$$

melyből a transzformációk elvégzése után a

$$\begin{bmatrix} D_x^{-2} & 0 & 0 & A^T \\ 0 & D_{x_F^+}^{-2} & D_{x_F^+}^{-2} & F^T \\ 0 & D_{x_F^+}^{-2} & D_{x_F^-}^{-2} + D_{x_F^+}^{-2} & 0 \\ A & F & 0 & 0 \end{bmatrix} \begin{bmatrix} w \\ w_{x_F^+} \\ w_{x_F^-} \\ y \end{bmatrix} = \begin{bmatrix} c \\ c_F \\ 0 \\ 0 \end{bmatrix} \tag{3.3}$$

egyenletrendszert kapjuk, ahol

$$\begin{bmatrix} w \\ w_{x_F^+} \\ w_{x_F^-} \\ y \end{bmatrix} = M_2^{-1} \begin{bmatrix} z \\ z_{x_F^+} \\ z_{x_F^-} \\ y \end{bmatrix} . \tag{3.4}$$

A (3.4) egyenletrendszer megoldása könnyen megadható :

$$z = w \, , \tag{3.5}$$

$$z_{x_F^+} = w_{x_F^+} + w_{x_F^-} \, , \tag{3.6}$$

$$z_{x_F^-} = w_{x_F^-} \, . \tag{3.7}$$

Felhasználjuk, hogy ugyanazon szabad változót többféleképpen fel lehet bontani pozitív és negatív részre, azaz a felbontás pozitív és negatív részéhez ugyanazt az értéket adva ugyanazon változó egy más felbontását kapjuk. Hasonló módon a kereső irány azon komponenseit is megváltoztathatjuk, melyek ugyanazon szabad változó pozitív és negatív részéhez tartoznak. Ezek szerint a (3.5)–(3.7) alapján látható, hogy az eredeti kereső irányok helyett a $\hat{z}_{x_F^+} = w_{x_F^+}$ és $\hat{z}_{x_F^-} = 0$ kereső irányokat is választhatjuk, azaz $w_{x_F^-}$ változókra nincs szükségünk. Eliminálunk ezeket (3.3)-ban a $D_{x_F^-}^{-2} + D_{x_F^+}^{-2}$ blokk diagonálisában való pivotálással, miután a következő egyenletrendszerhez jutunk:

$$\begin{bmatrix} D_x^{-2} & 0 & A^T \\ 0 & D_{\hat{x}_F}^{-2} & F^T \\ A & F & 0 \end{bmatrix} \begin{bmatrix} z \\ \hat{z}_{x_F^+} \\ y \end{bmatrix} = \begin{bmatrix} c \\ c_F \\ 0 \end{bmatrix} , \tag{3.8}$$

ahol

$$\hat{x}_{F_i} = \sqrt{x_{F_i}^{+\,2} + x_{F_i}^{-\,2}} \, . \tag{3.9}$$

Tehát csak a nem korlátozott változóknak a skálázása változik, mégpedig (3.9)-nek megfelelően. Az előbbiekben utaltunk arra, hogy egy felbontás pozitív és negatív részét ugyanannyival növelhetjük. Ha így $x_{F_i}^+ \to \infty$ és $x_{F_i}^- \to \infty$, akkor $\hat{x}_{F_i} \to \infty$, azaz határesetben $D_{\hat{x}_F}^{-2} = 0$. Így (3.8) helyett a

$$\begin{bmatrix} D_x^{-2} & 0 & A^T \\ 0 & 0 & F^T \\ A & F & 0 \end{bmatrix} \begin{bmatrix} z \\ \hat{z}_{x_F^+} \\ y \end{bmatrix} = \begin{bmatrix} c \\ c_F \\ 0 \end{bmatrix} \tag{3.10}$$

egyenletrendszer megoldását is alkalmazhatjuk. Itt megjegyezzük, hogy a (3.8) és (3.10) egyenletrendszerek megoldásaként adódó kereső irányok nem szükségképpen egyeznek meg. Könnyen ellenőrizhető, hogy az utóbbi esetben kapott kereső irány megegyezik annál a megközelítésnél adódó kereső iránnyal, amikor a szabad változókat eliminálju k a feladatból és az eredeti módszert alkalmazzuk. Az algoritmus azon pontját is módosítani kell, amely a lépéshosszat meghatározza, mégpedig az (1.3) maximumképzéséből ki kell hagynunk a szabad változókat.

Vanderbei a szabad változókat az első lépésben felső korlátos változóként írja le, majd a felső korláttal $+\infty$-hez, az alsó korláttal $-\infty$-hez tartva határátmenettel, és algebrai azonosságok útján kapja az A.1 agoritmus módosítását a (PF) feladatra [12]:

Algoritmus A.2

A szabad változók kereső irányának számítása

$z_F = (F^T BF)^{-1}(c_F - F^T BAD_x^2 c)$, ahol $B = (AD_x^2 A^T)^{-1}$.

A „duál" változók számítása

$y = BAD_x^2 c + BFz_F$.

A korlátozott változók kereső irányának számítása

$z = D_x^2(c - A^T y)$.

A lépéshossz meghatározása

$$h = \frac{\alpha}{\max\limits_{1 \le i \le n} \frac{z_i}{x_i}} .$$

A megállási kritérium ellenőrzése

Az új megengedett belső pont kiszámítása

$x^{\text{új}} = x - h \cdot z$, $x_F^{\text{új}} = x_F - h \cdot z_F$.

Vanderbei módosított algoritmusa a (3.10) egyenletrendszer megoldásánál a következő pivotválasztási előírásnak felel meg :

Pivotáljunk a D_x^{-2} blokk diagonálisában.

A pivotálás elvégzése után a következő egyenletrendszert kapjuk:

$$\begin{bmatrix} 0 & F^T \\ F & -AD_x^2 A^T \end{bmatrix} \begin{bmatrix} \hat{z}_{x_F^+} \\ y \end{bmatrix} = \begin{bmatrix} c_F \\ AD_x^2 c \end{bmatrix} . \tag{3.11}$$

Pivotáljunk a $-AD_x^2 A^T$ blokkban.

Ekkor (3.11) a következőképpen módosul:

$$[F^T(AD_x^2 A^T)^{-1} F][z_F] = [c_F - F^T(AD_x^2 A^T)^{-1} AD_x^2 c] . \tag{3.12}$$

A (3.12) egyenletrendszer megoldása az A.2 algoritmusnak is egy lépése. A pivotálások végzésénél még könnyen látható, hogy $y = BAD_x^2 c + BFz_F$ és $z = D_x^2(c - A^T y)$, megfelelően az A.2 algoritmusnak.

Vanderbei algoritmusának két speciális esetét taglalja. Az első a

$$\text{(P1)} \qquad \begin{aligned} &\min\ \xi \\ &[A \quad \rho]\begin{bmatrix} x \\ \xi \end{bmatrix} = b \\ &x \geq 0\,, \end{aligned}$$

első fázis feladat esete, ahol $\rho = b - A\zeta$, és $\zeta > 0$ induló megoldása (P1)-nek. Itt egyetlen szabad változónk van, a ξ. Ekkor $F = \rho$, azaz $(F(AD_x^2A^T)^{-1}F^T)^{-1}$ egy skalár szorzó, melyet a kereső irány kiszámításakor elhagyhatunk. Az A.2 algoritmus esetében így a kereső irány: $z = -D_x^2A^T(AD_x^2A^T)^{-1}\rho$, amely az első fázis Chandru és Kochar által módosított változata [4]. A mi interpretációnkban ez az előbbiekből következően egy speciális pivotválasztási stratégia alkalmazását jelenti. Hasonlóan látható ADLER, RESENDE, VEIGA és KARMARKAR „duál" affin skálázási módszere [1], és a (P) duálisára alkalmazott eredeti algoritmus ekvivalenciája, mely Vanderbei módosított algoritmusának másik speciális esete.

4. Megjegyzések

Az (1.7) egyenletrendszer mérete (azaz ismeretleinek száma) csökkenthető, ha nem egyenlőséges feltételek folytán slack változóink vannak. Az LP feladatot a következő formában írhatjuk fel:

$$\text{(PS)} \qquad \begin{aligned} &\min\ c \cdot x \\ &[A \quad -I]\begin{bmatrix} x \\ s \end{bmatrix} = b \\ &\begin{bmatrix} x \\ s \end{bmatrix} \geq 0\,. \end{aligned}$$

Az (1.7) egyenletrendszer ez esetben:

$$\text{(4.1)} \qquad \begin{bmatrix} D_x^{-2} & 0 & A^T \\ 0 & D_s^{-2} & -I \\ A & -I & 0 \end{bmatrix}\begin{bmatrix} z \\ z_s \\ y \end{bmatrix} = \begin{bmatrix} c \\ 0 \\ 0 \end{bmatrix}.$$

Látható, hogy

$$\text{(4.2)} \qquad z_s = D_s^{-2}y\,.$$

A D_s^{-2} blokk diagonálisában pivotálva eliminálják a z_s ismeretleneket, és (4.1) a következő egyenletrendszerré alakul:

$$\text{(4.3)} \qquad \begin{bmatrix} D_x^{-2} & A^T \\ A & -D_s^2 \end{bmatrix}\begin{bmatrix} z \\ y \end{bmatrix} = \begin{bmatrix} c \\ 0 \end{bmatrix}.$$

A módszer használható természetesen akkor is, amikor egyenlőséges és nem egyenlőséges „kevert" feltételeink vannak.

Jelölje M a (4.3) egyenletrendszer együtthatómátrixát, ξ az ismeretlenek vektorát, β pedig a jobboldalt. Az affin skálázási algoritmus iterációi közben tehát $M\xi = \beta$ egyenletrendszereket kell megoldani. Látható, hogy az iterációk közben az egyenletrendszernek csak az együtthatómátrixának diagonálisa változik. Jelölje (x^k, s^k) a k-adik iterációban kapott megoldását (PS)-nek, M^k pedig a k-adik iterációban a (4.3) egyenletrendszer együtthatómátrixát. Ekkor $M^{k+1} = M^k + \Delta^k$, ahol Δ^k diagonális mátrix, és elemei:

$$\Delta_{ii}^k = \begin{cases} \dfrac{1}{(x_i^{k+1})^2} - \dfrac{1}{(x_i^k)^2} & \text{ha } 1 \le i \le n, \\ (s_i^k)^2 - (s_i^{k+1})^2 & \text{ha } n+1 \le i \le n+m\,. \end{cases}$$

Látható, hogy „kis" lépések esetén (azaz ha $\|\Delta^k\|_\infty$ megfelelően kicsi) az $M^{k+1}\xi = \beta$ megoldása helyett iterációs közelítést használhatunk:

$$\xi^{l+1} = (M^k)^{-1}\beta - (M^k)^{-1}\Delta^k\xi^l. \tag{4.4}$$

A ξ^0 kezdeti értéknek jó választás a (4.3) egyenletrendszer előző affin iterációnál kapott megoldása, azaz $\xi^0 = (M^k)^{-1}\beta$. Megjegyezzük, hogy (4.4) akkor és csak akkor konvergens, ha $(M^k)^{-1}\Delta^k$ sajátértékeinek abszolút értéke kisebb mint egy. Ez utóbbi kritériumot ellenőrizhetjük például úgy, hogy az $M^k = LDL^T$ felbontásból, és Δ^k elemeinek ismeretéből $(M^k)^{-1}\Delta^k$ legnagyobb abszolút értékű sajátértékére könnyen adható becslés.

Egy másik lehetőség az $M^k = LDL^T$ felbontás folyamatos felfrissítése. A mi esetünkben az $M^k + \Delta^k = \hat{L}\hat{D}\hat{L}^T$ új felbontást kell meghatározni, mely a FLETCHER–POWEL algoritmus [5] alkalmas módosításával gyorsan számítható. A FLETCHER–POWEL algoritmus belső pontos algoritmusok implementációjában való alkalmazását SHANNO is javasolta [9] dolgozatában.

A lineáris programozás belső pontos algoritmusai közül talán a legtöbb valóban hatékony számítógépes implementáció az affin skálázási algoritmusra született [1,7,8]. A hatékony implementáció megköveteli az eredeti módszerek módosítását, hogy a lineáris programozási feladatok specialitásait kihasználjuk. A dolgozatban ilyen módosításokkal foglalkoztunk a belső pontos irodalomban egy ritkábban használt megközelítési módon keresztül. Ennek a megközelítésnek az az előnye, hogy különböző módosításokat egyszerűen ugyanazon feladat különböző pivotválasztási szabályaiból lehet származtatni. További előnyt jelent a normálegyenletrendszeren keresztül történő megoldással szemben az, hogy a feltételi mátrix ritkásságát és esetleges speciális struktúráját hatásosabban használhatjuk ki.

IRODALOM

[1] ADLER, I., RESENDE, M.G.C., VEIGA, G. and KARMARKAR, N., „An implementation of Karmarkar's algorithm for linear programming", *Mathematical Programming* **44** (1989), 297–335.

[2] BARNES, E.R., „A variation on Karmarkar's algorithm for solving linear programming problems", *Mathematical Programming* **36** (1986), 174–182.

[3] BJÖRK, A., „Methods for sparse linear least squares problems", *Sparse Martix computations* (Bunch, J.R. and Rose D.J., eds.) (Academic Press, INC., New York, 1976), 177–201.

[4] CHANDRU V. and KOCHAR, B.S., „A class of algorithms for linear programming", *Research Memorandum* No. **85-14** (Purdue University, 1985).

[5] FLETCHER, R. and POWEL, M.J.D., „On the modification of LDL^T factorizations", *Mathematics of Computation* **28** (1974), 1067–1087.

[6] LUSTIG, I. and MARSTEN, R.E and SALTZMAN, M.J. and SHANNO, D.F. and SUBRAMANIAN, R., „Interior point methods for linear programming", *Interfaces* **20:4** (1990), 105–116.

[7] MARSTEN, R.E and SALTZMAN, M.J. and SHANNO, D.F. and PIERCE, G.S. and BALLINTIJN, J.F., *ORSA Journal on Computing* **1** (1989), 287–287.

[8] MONMA C.L. and MORTON, A.J., „Computational experience with a dual affine variant of Karmarkar's method for linear programming", *Operations Research Letters* **6** (1987), 261–267.

[9] SHANNO, D.F., „Computing Karmarkar projection quickly", *Mathematical Programming* **41** (1988), 61–71.

[10] TOMLIN, J.A., „An experimenal approach to Karmarkar's projective method for linear programming", *Mathematical Programming Study* **31** (1987), 175–191.

[11] VANDERBEI, R.J. and MEKETON, M.S. and FREEDMAN, B.A.A., „A modification of Karmarkar's linear programming algorithm", *Algorithmica* **1** (1986), 395–407.

[12] VANDERBEI, R.J., „Affine-scaling for linear programs with free variables", *Mathematical Programming* **43** (1989), 31–41.

(Beérkezett: 1992. november 10.)

MÉSZÁROS CSABA
SZÁMÍTÁSTECHNIKAI ÉS AUTOMATIZÁLÁSI KUTATÓ INTÉZET
1111 BUDAPEST, KENDE U. 13–17.

ON THE MODIFICATIONS OF THE AFFINE SCALING ALGORITHM FOR LINEAR PROGRAMMING

CS. MÉSZÁROS

The most really effective implementation of interior point methods for linear programming problems is given for the affine scaling algorithm. The effective implementation require more modifications of the original methods. In the present paper we show a new approach to compute the projection on the scaling matrix of the linear programming problem. This approach has a advantageous property, that more modification rises from various pivot searching method.

Alkalmazott Matematikai Lapok **17** (1993) 195–210

SOROK A MATHIEU FÜGGVÉNYEK SAJÁTÉRTÉKEINEK KISZÁMÍTÁSÁHOZ

NÉMETH GÉZA

Budapest

A cikk sorelőállításokat tartalmaz a Mathieu függvények első hat sajátértékére. A képletek segítségével a sajátértéket igen gyorsan tíz decimális jegy pontossággal kiszámíthatjuk.

1. Bevezetés

Az

$$y'' + (a - 2q\cos 2x)y = 0 \tag{1}$$

differenciál egyenlet megoldásait nevezik Mathieu függvényeknek. A π és 2π periodikus megoldásokat az alábbi sorokkal adják meg [1]

$$ce_{2n}(x,q) = \sum_{r=0}^{\infty} A_{2r}^{(2n)} \cos 2rx, \tag{2}$$

$$ce_{2n+1}(x,q) = \sum_{r=0}^{\infty} A_{2r+1}^{(2n+1)} \cos(2r+1)x, \tag{3}$$

$$n = 0, 1, 2, \ldots,$$

$$se_{2n+1}(x,q) = \sum_{r=0}^{\infty} B_{2r+1}^{(2n+1)} \sin(2r+1)x, \tag{4}$$

$$se_{2n+2}(x,q) = \sum_{r=0}^{\infty} B_{2r+2}^{(2n+2)} \sin(2r+2)x. \tag{5}$$

Egy adott q értékhez az egyenletben szereplő „a" szám a sajátérték. A sajátértékeket a q függvényeként meg lehet határozni hatványsorokkal. Így pl. $n = 0$ esetén

$$a_0(q) = -\frac{q^2}{2} + \frac{7q^4}{128} - \frac{29q^6}{2304} + \ldots, \quad q \to 0, \tag{6}$$

$$a_0(q) = -2q + 2q^{\frac{1}{2}} - \frac{1}{4} + \ldots, \qquad q \to \infty. \tag{7}$$

Hasonló sorok léteznek a többi sajátértékre is. A (6) és (7) sorok csak kis pontosságot tudnak biztosítani a_0 kiszámításánál. Jelen cikkben a sajátértékekre nagypontosságú sorokat határozunk meg.

2. Az új sorok

Ismeretes, hogy a_0 sora kis q esetén konvergál, de $q_0 = 1,468$ esetén már divergál. Sőt a $q \to \infty$-re vonatkozó sorra, hogy milyen $q_1 \geq q$ tartományban lenne konvergens, egyáltalán nem létezik állítás. De különben is felmerül a kérdés, hogy az átmeneti tartományban $q_0 \leq q \leq q_1$-ben hogyan lehet kiszámítani a_0 értékét. Hasonló problémák vannak persze a többi sajátértékre. Aszimptotikus típusú közelítéseket szeretnénk meghatározni, ezért a hatványsort és a $q \to \infty$ aszimptotikus sort írtuk át jobban konvergáló sorba — Csebisev sorba — és az átmeneti tartományra külön határoztunk meg sorfejtést.

Tehát a $0 \leq q < \infty$ szakaszt három részre bontottuk $(0, \alpha)$, (α, β) és (β, ∞) szakaszokra és tekintjük az alábbi sorokat

$$a(q) = \sum_{k=0}^{\infty} c_k T_k^* \left(\frac{q}{\alpha}\right), \qquad 0 \leq q \leq \alpha, \tag{11}$$

$$a(q) = 2q \sum_{k=0}^{\infty} d_k T_k^* \left(\frac{q-\alpha}{\beta-\alpha}\right), \qquad \alpha \leq q \leq \beta, \tag{12}$$

$$a(q) = -2q \sum_{k=0}^{\infty} f_k T_k^* \left(\left(\frac{\beta}{q}\right)^{\frac{1}{2}}\right), \qquad q \geq \beta. \tag{13}$$

A három sor meghatározásához meg kell adni az α és β állandókat. E számok kiválasztásánál azt a célt próbáltuk követni, hogy kb. ugyanannyi tag szerepeljen tíz jegy pontosságig a sorokban. Néha ez nem volt elérhető, ilyenkor arra törekedtünk, hogy legalább egyik sorban se szerepeljen túl sok tag a másikhoz képest a tíz jegy pontosságra vonatkozóan. Tulajdonképpen a (13) sor konvergenciáját $a(q)$ hatványsorának szingularitása határozza meg. Ez pl. a_0 esetében ismert. A további sajátértékekre ezt a kérdést G. BLANCH vizsgálta [2]. Meg kell jegyezni, hogy α, β értékénél nincs szükség nagy pontosságra. Mi is egész értékre kerekítettük a numerikus számítással nyert vagy táblázatból vett értékeket.

A sajátértéket, mint ismeretes lánctörteket tartalmazó transcendens egyenletből lehet meghatározni. Így $a_0(q)$ az

$$a = \cfrac{2q}{\cfrac{a-4}{q} - \cfrac{1}{\cfrac{a-16}{q} - \cfrac{1}{\cfrac{a-36}{q} - \cfrac{1}{\cfrac{a-64}{q} - \ldots}}}}$$

egyenlet gyöke. Ezt az egyenletet pl. a Newton féle iterációval lehet megoldani. A számítás elég sok munkával jár a hosszú lánctörtek számítása miatt. Alkalmas kezdőérték segítségével a számítás természetesen redukálható.

A sorok együtthatóit (11) esetében direkt konverzióval nyertük (a hatványsort átírtuk Csebisev sorba). A (12) és (13) sorok együtthatóit interpolációval határoztuk meg. Az interpolációs pontok száma 15 és 30 között volt (ez függött α és β aktuális értékétől).

Az (1) egyenlet páros megoldásainak (2) és (3) sajátértékeit, szokás szerint, a_{2n} és a_{2n+1}-gyel jelöltük ($n = 0, 1, 2, \ldots$), a páratlan megoldások (4) és (5) sajátértékeit pedig b_{2n} és b_{2n+1}-gyel jelöltük ($n = 1, 2, \ldots$). Az aktuális sorok együtthatóit a_n-hez $c_k^{(n)}$, $d_k^{(n)}$ és $f_k^{(n)}$ jelöljük, hasonlóan b_n-hez a $g_k^{(n)}$, $h_k^{(n)}$, $\ell_k^{(n)}$ jelöléseket használjuk.

A számításokat SHARP PC–E500 kalkulátorral végeztem el. Néhány számítás elvégzéséhez Magyari Zoltán és Réti Sándor egy PC AT – 286 gépet bocsátottak a rendelkezésemre. Szivességükért fogadják köszönetemet.

3. Táblázatok

A nyert sorok együtthatóit táblázatosan adjuk meg a (11), (12) és (13) képletek szeint. A (11) jelölésben kivételt képez a_0 esete, amikor a sorból előre kiemeltük a „$-q^2/2$" szorzót

$$a_0(q) = -\frac{q^2}{2}\sum_{k=0}^{\infty} c_k^{(0)} T_{2k}(q/2), \quad 0 \le q \le 2.$$

Továbbá kivétel a_2, a_4, a_6, b_2, b_4, b_6, mert soruk q^2 szerint halad és így a Csebisev sorban $T_k^*(q/\alpha)$ helyett $T_{2k}(q/\alpha)$ áll értelemszerűen. A (12) soroknál kivételt képez, hogy kényelmi okok miatt a sor előtt „$-2q$" áll „$2q$" helyett a_0 és a_1 valamint b_1 és b_2 esetén.

Az α és β számok értéke az alábbi táblázatból veendő:

	a_n		b_n	
n	α	β	α	β
0	2	8	–	–
1	4	16	4	16
2	2	8	8	32
3	4	16	4	16
4	8	32	8	32
5	12	36	12	36
6	18	54	18	72

k	$c_k^{(0)}$	k	$d_k^{(0)}$	k	$f_k^{(0)}$
0	0,8618429376	0	0,5519085607	0	0,8293360635
1	−,1187244731	1	,1350312940	1	−,1685781503
2	, 161308012	2	−, 295872941	2	, 21183915
3	−, 26849457	3	, 68444126	3	, 364222
4	, 4947118	4	−, 15535049	4	, 51980
5	−, 969742	5	, 3313499	5	, 18903
6	, 198233	6	−, 628172	6	, 6392
7	−, 41773	7	, 91927	7	, 1382
8	, 9009	8	−, 3362	3	−, 7
9	−, 1978	9	−, 4645	9	−, 125
10	, 441	10	, 2577	10	−, 32
1	−, 99	1	−, 975	1	, 6
2	, 23	2	, 310	2	, 5
3	−, 5	3	−, 87	3	−, 0
4	, 1	4	−, 22	4	−, 1
		5	−, 5		
		6	, 1		

1. Táblázat: a_0

k	$c_k^{(1)}$	k	$d_k^{(1)}$	k	$f_k^{(1)}$
0	2,0126384394	0	0,0632309682	0	0,6404567067
1	,6593632251	1	,2770321901	1	−,3542413632
2	−,3598424492	2	−, 594167549	2	, 53989763
3	−, 3086001	3	, 132034491	3	, 1031992
4	, 64507856	4	−, 27579365	4	, 72532
5	−, 550598	5	, 4975958	5	, 15035
6	−, 2547337	6	−, 666334	6	, 5588
7	, 48700	7	, 38722	7	, 2032
8	, 127149	8	−, 1227	8	, 555
9	−, 3767	9	, 11252	9	, 69
10	−, 7121	10	−, 11037	10	−, 25
1	, 286	1	, 6744	1	−, 16
2	, 427	2	−, 3257	2	−, 3
3	−, 22	3	, 1346	3	, 1
4	−, 27	4	−, 491	4	, 1
5	, 2	5	, 160		
6	, 2	6	−, 46		
		7	, 11		
		8	−, 2		

2. Táblázat: a_1

k	$c_k^{(2)}$	k	$d_k^{(2)}$	k	$f_k^{(2)}$
0	4,6328201244	0	0,8113106790	0	0,2069711114
1	,5805330781	1	−,3598234546	1	−,7586666815
2	−, 455593127	2	, 763221832	2	, 379381411
3	, 56327227	3	−, 288172328	3	, 47513667
4	−, 8961428	4	, 110823530	4	, 15280411
5	, 1602925	4	−, 38374531	5	, 3838077
6	−, 307525	6	, 12971353	6	−, 107694
7	, 61847	7	−, 4459701	7	, 683581
8	−, 12867	8	, 1526943	8	−, 313368
9	, 2746	9	−, 513648	9	−, 27657
10	−, 598	10	−, 171583	10	, 36441
1	−, 132	1	−, 57379	1	, 15783
2	, 30	2	, 19147	2	−, 301
3	, 7	3	−, 6362	3	−, 1715
4	−, 2	4	−, 2111	4	−, 103
		5	−, 701	5	, 137
		6	, 233	6	−, 61
		7	−, 77	7	−, 41
		8	, 26	8	, 19
		9	−, 9	9	, 13
		20	, 3	20	−, 3
		1	−, 1	1	−, 2
				2	, 1

3. Táblázat: a_2

k	$c_k^{(3)}$	k	$d_k^{(3)}$	k	$f_k^{(3)}$
0	9,6095410398	0	0,8362760150	0	0,2090277808
1	,8360155908	1	−,3738828941	1	−,7604324744
2	,2323395865	2	, 730860254	2	, 325497950
3	−, 5416740	3	−, 299579511	3	, 25394613
4	−, 66102670	4	, 133704198	4	, 7344275
5	, 441026	5	−, 48712768	5	, 2772374
6	, 2552327	6	, 15978357	6	, 878596
7	−, 48180	7	−, 5483482	7	, 135306
8	−, 127122	8	, 1944686	8	−, 68212
9	, 3770	9	−, 642161	9	−, 53121
10	, 7121	10	, 196979	10	−, 18019
1	−, 286	1	−, 61922	1	−, 725
2	−, 427	2	, 20609	2	, 1978
3	, 22	3	−, 6701	3	, 689
4	, 27	4	−, 2069	4	−, 52
5	−, 2	5	−, 688	5	−, 71
6	−, 2	6	, 248	6	, 13
		7	−, 88	7	, 17
		8	, 30	8	−, 1
		9	−, 10	9	−, 4
		20	, 3	20	−, 1
		1	−, 1	1	, 1

4. Táblázat: a_3

k	$c_k^{(4)}$	k	$d_k^{(4)}$	k	$f_k^{(4)}$
0	17,5456448256	0	0,7374219591	0	0,2714879706
1	1,6540060688	1	−,3757603541	1	−,7047727746
2	, 786344060	2	, 558056903	2	, 247232862
3	−, 280569565	3	−, 165611080	3	, 11096832
4	, 22087319	4	, 100010939	4	, 1567987
5	, 3939582	5	−, 51428819	5	, 429871
6	−, 1390895	6	, 18625044	6	, 167818
7	, 105627	7	−, 4987321	7	, 67682
8	, 36022	8	, 1385547	8	, 23204
9	−, 11439	9	−, 628614	9	, 5372
10	, 642	10	, 294478	10	−, 8
1	, 401	1	−, 84159	1	−, 723
2	−, 112	2	, 5266	2	−, 358
3	, 3	3	, 3505	3	−, 70
4	, 5	4	, 945	4	, 16
5	−, 1	5	−, 1634	5	, 14
		6	, 445	6	, 2
		7	, 128	7	−, 1
		8	−, 87	8	−, 1
		9	, 20		
		20	−, 31		
		1	−, 8		
		2	, 3		
		3	−, 2		

5. Táblázat: a_4

k	$c_k^{(5)}$	k	$d_k^{(5)}$	k	$f_k^{(5)}$
0	26,4631020212	0	0,8537293545	0	0,1735737837
1	2,0425582316	1	−,2999013555	1	−,7933991112
2	,6444282558	2	, 376867139	2	, 343910649
3	, 674750292	3	−, 147371284	3	, 20747441
4	−, 25966388	4	, 76449709	4	, 4109602
5	−, 62567036	5	−, 25792319	5	, 1423766
6	−, 10424747	6	, 5534295	6	, 598815
7	, 2459069	7	−, 1134455	7	, 238533
8	, 1172168	8	, 440952	8	, 75348
9	, 73031	9	−, 136535	9	, 12340
10	−, 80769	10	, 104	10	−, 4309
1	−, 25217	1	, 12195	1	−, 4585
2	, 1547	2	−, 708	2	−, 1954
3	, 2679	3	−, 1628	3	−, 374
4	, 482	4	, 321	4	, 73
5	−, 142	5	, 98	5	, 68
6	−, 81	6	, 21	6	, 11
7	−, 5	7	−, 50	7	−, 6
8	, 7	8	, 11	8	−, 2
9	, 2	9	, 5	9	, 1
		20	−, 2	20	, 1

6. Táblázat: a_5

k	$c_k^{(6)}$	k	$d_k^{(6)}$	k	$f_k^{(6)}$
0	38,9036846084	0	0,8167691187	0	0,1988875467
1	3,1320776133	1	−,3040444058	1	−,7713066561
2	,2318792459	2	, 319615144	2	, 312094697
3	−, 74332864	3	−, 98559499	3	, 15880460
4	−, 103723047	4	, 69363772	4	, 2249405
5	, 8362604	5	−, 31763376	5	, 543050
6	, 2836760	6	, 7943705	6	, 196102
7	−, 211963	7	−, 701770	7	, 84449
8	−, 151053	8	−, 55193	8	, 35809
9	, 14609	9	−, 189647	9	, 13183
10	, 7021	10	, 79572	10	, 3606
1	−, 713	1	−, 18057	1	, 350
2	−, 394	2	−, 22787	2	−, 324
3	, 45	3	, 3235	3	−, 240
4	, 22	4	−, 3295	4	−, 85
5	−, 3	5	−, 1371	5	−, 10
6	−, 1	6	−, 255	6	, 6
		7	, 253	7	, 4
		9	−, 35		
		9	−, 72		
		20	, 10		
		1	−, 14		
		2	−, 6		
		3	−, 1		
		4	, 1		

7. Táblázat: a_6

k	$g_k^{(1)}$	k	$h_k^{(1)}$	k	$\ell_k^{(1)}$
0	−1,5084107502	0	0,6730841244	0	0,8780131278
1	−2,6448971858	1	,1051661248	1	−,1209678172
2	−0,1194616129	2	−, 256005071	2	, 10281837
3	, 151799120	3	, 69950558	3	, 63607
4	−, 17239977	4	−, 20301860	4	, 2285
5	, 1297329	5	, 6134246	5	−, 105
6	, 41351	6	−, 1909256	6	−, 105
7	−, 38103	7	, 607644	7	−, 46
8	, 8128		−, 146595	8	−, 16
9	−, 1012	9	, 64353	9	−, 3
10	, 33	10	−, 21232		
1	, 20	1	, 7041		
2	−, 6	2	−, 2342		
3	, 1	3	, 781		
		4	−, 260		
		5	, 89		
		6	−, 29		
		7	, 10		
		8	−, 3		
		9	, 1		

8. Táblázat: b_1

k	$g_k^{(2)}$	k	$h_k^{(2)}$	k	$\ell_k^{(2)}$
0	1,7107017982	0	0,3137673386	0	0,7424294251
1	−2,1852438073	1	,2175546476	1	−,2549909877
2	, 934065118	2	−, 521679534	2	, 26097665
3	−, 92615212	3	, 140077866	3	, 311784
4	, 11834357	4	−, 39878129	3	, 10459
5	−, 1710792	5	, 11823599	5	, 484
6	, 266111	6	−, 3623490	6	, 16
7	−, 43456	7	, 1142343	7	−, 6
8	, 7347	8	−, 348751	8	−, 4
9	−, 1275	9	, 121248	9	−, 2
10	, 226	10	−, 40388	10	−, 1
1	−, 41	1	, 13561		
2	, 7	2	−, 4570		
3	−, 1	3	, 1541		
		4	−, 519		
		5	, 174		
		6	−, 58		
		7	, 50		
		8	−, 7		
		9	, 2		
		20	−, 1		

9. Táblázat: b_2

k	$g_k^{(3)}$	k	$h_k^{(3)}$	k	$\ell_k^{(3)}$
0	9,1338698334	0	0,5100467592	0	0,4164416672
1	,1457303933	1	−,4661035646	1	−,5691123991
2	−, 44975566	2	,1225801096	2	0, 148501674
3	−, 148902755	3	−, 389277213	3	, 4151019
4	, 18063785	4	, 132106199	4	, 2510
5	−, 1209295	5	−, 45157281	5	−, 170186
6	−, 47760	6	, 15276425	6	−, 88124
7	, 37785	7	−, 5120511	7	−, 29411
8	−, 8107	8	, 1709606	8	−, 4997
9	, 1014	9	−, 570443	9	, 1085
10	−, 33	10	, 190367	10	, 1036
1	−, 20	1	−, 63515	1	, 249
2	, 6	2	, 21181	2	−, 51
3	−, 1	3	−, 7061	3	−, 48
		4	, 2353	4	−, 6
		5	−, 784	5	, 7
		6	, 261	6	, 2
		7	−, 87	7	−, 0
		8	, 29	8	−, 1
		9	−, 10		
		20	, 3		
		1	−, 1		

10. Táblázat: b_3

k	$g_k^{(4)}$	k	$h_k^{(4)}$	k	$\ell_k^{(4)}$
0	16,6859052087	0	0,4835006304	0	0,4209683550
1	,5814341081	1	−,4358389140	1	−,5652166888
2	−, 934157793	2	,1058250436	2	, 142011699
3	, 96540461	3	−, 319175759	3	, 4134821
4	−, 11981881	4	, 108949988	4	, 291588
5	, 1712548	5	−, 38726452	5	, 15961
6	−, 265979	6	, 13610801	6	−, 5935
7	, 43447	7	−, 4650327	7	−, 4479
8	−, 7347	8	, 1552321	8	−, 2034
9	, 1275	9	−, 513222	9	−, 683
10	−, 226	10	, 170030	10	−, 144
1	, 41	1	−, 56668	1	, 5
2	−, 7	2	, 18942	2	, 19
3	, 1	3	−, 6321	3	, 8
		4	, 2102	4	, 1
		5	−, 697	5	−, 1
		6	, 232		
		7	−, 77		
		8	, 26		
		9	−, 9		
		20	, 3		
		1	−, 1		

11. Táblázat: b_4

k	$g_k^{(5)}$	k	$h_k^{(5)}$	k	$\ell_k^{(5)}$
0	25,942677941	0	0,6572705514	0	0,3090004259
1	1,200351771	1	−,3781778942	1	−,6702255834
2	,213098232	2	, 747025186	2	, 215188185
3	−, 51489835	3	−, 193444521	3	, 8086458
4	−, 5969315	4	, 56904477	4	, 678277
5	, 1210522	5	−, 16768391	5	, 22319
6	, 199787	6	, 4688754	6	−, 32177
7	−, 68907	7	−, 1247158	7	−, 21366
8	−, 219	8	, 325231	8	−, 9440
9	, 2758	9	−, 85652	9	−, 3018
10	−, 241	10	, 22941	10	−, 518
1	−, 106	1	−, 6166	1	, 114
2	, 21	2	, 1643	2	, 126
3	, 3	3	−, 435	3	, 45
4	−, 1	4	, 116	4	, 3
		5	−, 31	5	, 5
		6	, 9	6	−, 2
		7	−, 2		
		8	, 1		

12. Táblázat: b_5

k	$g_k^{(6)}$	k	$h_k^{(6)}$	k	$\ell_k^{(6)}$
0	38,1067446781	0	0,5350559818	0	0,3949388499
1	1,9983235869	1	−,4331854370	1	−,5900358901
2	−,1244434678	2	, 983105123	2	, 154556942
3	−, 115541222	3	−, 280830149	3	, 4620497
4	, 38201167	4	, 97361822	4	, 347651
5	−, 5591903	5	−, 37478975	5	, 35107
6	, 560117	6	, 14454509	6	, 3993
7	−, 8320	7	−, 5245322	7	, 379
8	, 17321	8	, 1752478	8	−, 36
9	−, 4335	9	−, 544820	9	−, 52
10	, 642	10	, 164610	10	−, 29
11	, 39	1	−, 51712	1	−, 13
12	−, 10	2	, 17611	2	−, 4
13	, 4	3	−, 6294	3	−, 1
		4	, 2198		
		5	, 714		
		6	−, 217		
		7	−, 66		
		8	, 22		
		9	, 9		
		20	−, 3		
		1	, 1		

13. Táblázat: b_6

IRODALOM

[1] ABRAMOVITZ, M., STEGUN, I. A. (eds.), „Handbook of Mathematical Functions", *NBS Appl. Math. Ser.*, vol. 55 (US Govt. Print. Off., Washington, D.C., 1964).

[2] BLANCH, G., „Numerical aspects of Mathieu eigenvalues", *Rend. Circ. mat. Palermo* (2) **15** (1966), 51–97.

(Beérkezett: 1991. december 21.)

NÉMETH GÉZA
KFKI MÉRÉS- ÉS SZÁMÍTÁSTECHNIKAI KUTATÓ INTÉZET
1525 BUDAPEST, 114, POSTAFIÓK 49.

SERIES APPROXIMATIONS FOR THE EIGENVALUES OF MATHHIEN FUNCTIONS

G. NÉMETH

In this paper series approximations for the eigenvalues of Mathieu functions are given.

A kiadásért felelős az ELTE TTK dékánja
Szedte a KLTE Informatikai és Számítóközpont Kiadvány Szerkesztő Csoportja
és nyomta az ELTE Sokszorosítóüzeme
Felelős vezető: Arató Tamás
Budapest, 1994. – ELTE 94278
Megjelent: 19,2 (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: 425 Ft

TARTALOMJEGYZÉK

Bálint Erzsébet és Deák István, Párhuzamos számítógépek: optimalizálási programok 1
Csendes Tibor, Egy intervallum-aritmetikán alapuló algoritmus a szinthalmazok korlátainak megkeresésére 19
Simon L. Péter és Farkas Henrik, Polinomok gyökstruktúrájának vizsgálata a parametrikus reprezentáció módszerével 41
Csébfalvi A., Nemlineáris útkövető módszer tartószerkezetek stabilitásvizsgálatára, I. Reguláris pontok 57
Csébfalvi A., Nemlineáris útkövető módszer tartószerkezetek stabilitásvizsgálatára, II. Elágazási és határpontok 71
Szirmai Jenő, Néhány tércsoport optimális gömbkitöltése 87
Faragó István, Haroten Hariton, Komáromi Nándor és Pfeil Tamás, A hővezetési egyenlet és numerikus megoldásának kvalitatív tulajdonságai, I. Az elsőfokú közelítések nemnegativitása 101
Faragó István, Haroten Hariton, Komáromi Nándor és Pfeil Tamás, A hővezetési egyenlet és numerikus megoldásának kvalitatív tulajdonságai, II. A másodfokú közelítés nemnegativitása, a maximum elv és az oszcillációmentesség 123
Nagy Tamás, A szállítási feladat sztochasztikus variánsai 143
Mihálykó Csaba, A golyósmalmi őrlemény sűrűségfüggvényére felírt integro-differenciálegyenlet megoldhatósága és a megoldás speciális tulajdonságai 171
Mészáros Csaba, Az affin skálázási algoritmus módosításairól 185
Németh Géza, Sorok a Mathieu függvények sajátértékeinek kiszámításához 195

INDEX

Bálint, E. and *Deák, I.,* Parallel computers: optimization software 1
Csendes, T., An interval method for bounding level sets 19
Simon, P. L. and *Farkas, H.,* The investigation of the root structure of polynomials with the parametric representation method 41
Csébfalvi, A., Nonlinear path-following method for stability of structures I. Regular points . 57
Csébfalvi, A., Nonlinear path-following method for stability of structures II. Bifurcation and limit points 71
Szirmai, J., Optimale kugelpackungen unter einigen raumgruppen 87
Faragó, I., Hariton, H. A., Komáromi, N. and *Pfeil, T.,* The differential equation of the heat transfer and qualitative properties its numerical solutions: I. The nonnegativity of the first order approximations 101
Faragó, I., Hariton, H. A., Komáromi, N. and *Pfeil, T.,* The differential equation of the heat transfer and qualitative properties its numerical solutions: II. The nonnegativity of the second order approximation, the maximum principle and the nonoscillation 123
Nagy, T., Stochastic variants of the entropy programming 143
Mihálykó, Cs., Solubility of integrodifferential equation for the density function of ball mill granulate, and special properties of solution 171
Mészáros, Cs., On the modifications of the affine scaling algorithm for linear programming . 185
Németh, G., Series approximations for the eigenvalues of Mathhien functions 195

Alkalmazott matematikai lapok

1993/3–4

A MAGYAR TUDOMÁNYOS AKADÉMIA
MATEMATIKAI ÉS FIZIKAI TUDOMÁNYOK
OSZTÁLYÁNAK KÖZLEMÉNYEI

17.

KÖTET

ALKALMAZOTT MATEMATIKAI LAPOK

A MAGYAR TUDOMÁNYOS AKADÉMIA MATEMATIKAI ÉS FIZIKAI TUDOMÁNYOK OSZTÁLYÁNAK KÖZLEMÉNYEI

ALAPÍTOTTÁK

KALMÁR LÁSZLÓ, TANDORI KÁROLY, PRÉKOPA ANDRÁS, ARATÓ MÁTYÁS

FŐSZERKESZTŐ

BENCZÚR ANDRÁS

FŐSZERKESZTŐ-HELYETTESEK

DEMETROVICS JÁNOS, FARKAS MIKLÓS

FELELŐS SZERKESZTŐ

SZÁNTAI TAMÁS

XVII. kötet 3–4. szám

Szerkesztőség és kiadóhivatal: 1027 Budapest, Fő utca 68.

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
1027 Budapest, Fő utca 68.

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 850 forint. Megrendelések a szerkesztőség címén lehetségesek.

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 17 (1993) 211–223

EGY ALGORITMUS A KÖLTSÉGTERVEZÉSI FELADAT MEGOLDÁSÁRA TEVÉKENYSÉG-ÉLŰ TERVÜTEM HÁLÓN, (CPM/COST FELADAT)*

KLAFSZKY EMIL ÉS HAJDU MIKLÓS

Budapest

Cikkünkben egy új algoritmust adunk az először KELLEY és WALKER, később FULKERSON által megoldott költségtervezési feladatra. A feladat KELLEY és WALKER nyomán 'Critical Path Method', röviden CPM/cost feladatnak nevezik.

Eljárásunkat KELLEY és FULKERSON algoritmusaira alapozva fejlesztettük ki. A megoldás során a maximális átfutási időhöz tartozó optimális megoldásról térünk át egy kisebb átfutási időhöz tartozó optimális megoldásra. Módszerünk a mi megítélésünk szerint egyszerűbb, és könnyebben programozható az eddigieknél. Kulcsszavak: hálótechnika, költségoptimalizálás, CPM/cost feladat.

1. Bevezetés

A CPM/cost feladat első ismertetése és megoldása KELLEY és WALKER cikkében található [3]. KELLEY [4] munkájának eredménye egy, a lineáris programozás primál-duál algoritmusán alapuló algoritmus. A feladat folyam algoritmussal történő megoldása RAY FULKERSON [2] cikkében található.

Eljárásunk kidolgozásánál KELLEY [4] és FULKERSON [2] eredményeit fejlesztettük tovább. Míg FULKERSON modellje a háló éleit megduplázta, és ezen a bővített ismételte a maximális folyam feladatot, mi KELLEY [4] gondolatát felhasználva, elértük, hogy a hálóhoz új éleket ne kelljen hozzáadni. Ez mind elméleti mind gyakorlati sebességnövekedést okoz a feladat megoldásánál. Ezen túl az algoritmus véleményünk szerint egyszerűbb és könnyebben programozható az eddigieknél.

Cikkünkben feltételezzük, hogy az olvasó tisztában van a hálózati folyamokkal, (FORD–FULKERSON [1]) és a tervütem hálókkal kapcsolatos alapfogalmakkal, és alapfeladatokkal. Ezek a digráf, a folyam és a vágás fogalmai, valamint a maximális folyam minimális vágás, és a CPM/time feladatok. Ezek tárgyalására ezen dolgozatban nincs lehetőségünk.

* Készült az OTKA F4112 pályázata támogatásával.

2. A feladat modellje, és megoldása

Legyen adott egy irányított gráf. Csak egy kezdő és végpontja lehet. A kezdőpont legyen s, a végpont legyen t. A gráfban hurok és a csomópontok között egynél több él nem megengedett. A csomópontok halmaza legyen N, az élek halmaza $\mathcal{A}$. Az élek azonosítása azzal a két csomóponttal történik, melyeket az él összeköt. A gráf jelölése legyen $[N, \mathcal{A}]$. Ebben a tervütem hálóban a gráf élei tevékenységeket reprezentálnak, a csomópontok eseményeket. Az i-edik esemény jelentése a tervütem hálóban a következő: Minden i-be futó tevékenységnek i bekövetkeztére be kell fejeződnie, és minden i-ből kifutó tevékenység legkorábban i bekövetkeztekor kezdődhet el. Ha egy tevékenység idejét általánosan τ-val jelöljük, az események bekövetkeztét π-vel, akkor a fentiekből rögtön adódik a következő feltétel:

$$\pi_j - \pi_i \geq \tau_{ij} \quad \forall (i,j) \in \mathcal{A}. \tag{1}$$

Legyen adott minden τ_{ij} tevékenység időre egy alsó és felső időkorlát, melyek által meghatározott intervallumban kell τ_{ij}-nek lennie. Az alsó időkorlátot rohamidőnek, a felső időkorlátot normálidőnek nevezzük. Az (i,j) élhez tartozó roham és normál idő jele legyen a_{ij} és b_{ij}. A következő feltétel tehát:

$$a_{ij} \leq \tau_{ij} \leq b_{ij} \quad \forall (i,j) \in \mathcal{A}. \tag{2}$$

Az (1) feltétel betartásával a gráfon több különböző, a feltételt kielégítő π_i rendszer adható meg egy adott τ_{ij} rendszerhez. Azt a π_i rendszert, melyre π_i a lehető legkisebb minden i-re, de kielégíti (1)-et, minimális időpolitikának nevezzük. Az ehhez a rendszerhez tartozó π_t értéket p-vel jelöljük, és a tervütem háló átfutási vagy megvalósulási idejének nevezzük, feltéve hogy $\pi_s = 0$.

$$\pi_s = 0 \tag{3}$$

$$\pi_t = p. \tag{4}$$

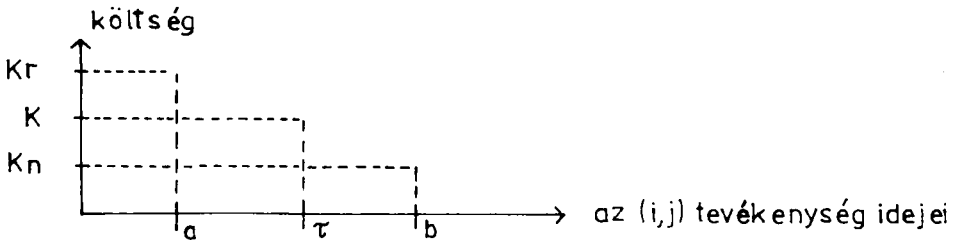
Egy adott τ_{ij} rendszerhez tartozó minimális π_i rendszert tekintve létezik olyan s-ből t-be vezető út, amely mentén (1) egyenlőséggel teljesül. Ezt az utat kritikus útnak nevezzük. Legyen adott minden (i,j) tevékenység normálidejéhez egy Kn_{ij} normál költség és egy c_{ij} költségtényező, mely megmutatja, hogy a tevékenység idejének egy napos gyorsításával mennyivel emelkedik a megvalósítási költség. Ennek ismeretében meghatározható a roham időhöz tartozó ún. roham költség (Kr_{ij}):

$$Kr_{ij} = Kn_{ij} + c_{ij}(b_{ij} - a_{ij}).$$

További megkötés c_{ij}-re, hogy nemnegatív legyen, azaz

$$Kr_{ij} \geq Kn_{ij} \quad \forall (i,j) \in \mathcal{A}.$$

Az imént leírtakat az 1. ábra szemlélteti.



1. ábra

Egy olyan tervütem hálón, ahol a τ_{ij} nem konkrét, hanem (2)-nek megfelelően egy alsó és felső időkorlát között mozoghat, többféle átfutási idő érhető el. A normálidőkkel $\{b_{ij}\}$ számolt hálón a π_t átfutási idő az elérhető maximális, jele legyen $p_{\max}$. A rohamidőkkel számolt hálón, a π_t átfutási idő az elérhető minimális, jele legyen $p_{\min}$. A τ_{ij} tevékenység idők változtatásával, adott τ_{ij} rendszerekhez minden olyan p átfutási idő elérhető, ahol:

$$p_{\min} \leq p \leq p_{\max}. \tag{5}$$

A tevékenység idők változtatásával azonos p átfutási időt is többféleképpen lehet elérni. Nyilvánvaló, ha a tevékenységek költségtényezője kölönbözik, úgy azonos p átfutási idejű megoldáshoz különböző megvalósítási költségek tartoznak. Mi egy adott átfutási időhöz tartozó legolcsóbb megoldást keresünk. Az 1. ábra segítségével felírhatjuk a következő célfüggvényt:

$$\min \left(\sum_A \{ Kn_{ij} + (b_{ij} - \tau_{ij}) c_{ij} \} \right).$$

Mivel Kn_{ij} és $(b_{ij} c_{ij})$ konstansok, ezért az előzővel egyenértékű célfüggvény a következő:

$$\max \left(\sum_A c_{ij} \tau_{ij} \right). \tag{6}$$

Az előbbiekben elmondottak alapján összefoglalva a feladat tehát a következő:

Keresendő az a π_i és τ_{ij} rendszer, amely kielégíti az alábbi feltételeket:

$$\begin{aligned}
&\pi_j - \pi_i \geq \tau_{ij} && \forall (i,j) \in \mathcal{A} \\
&a_{ij} \leq \tau_{ij} \leq b_{ij} && \forall (i,j) \in \mathcal{A} \\
&\pi_s = 0 \\
&\pi_t = p \\
&p_{\min} \leq p \leq p_{\max}
\end{aligned}$$

és maximalizálja a következő célfüggvényt

$$\max\left(\sum_{A} c_{ij}\tau_{ij}\right).$$

Ezt a feladatot a CPM/cost feladat primál feladatának nevezzük. A feladathoz rendeljük hozzá az alábbi duál feladatot.

Legyen φ_{ij} az s pontból a t pontba irányuló folyamrendszer, melynek nagysága θ. Keresendő az $[N, \mathcal{A}]$ digráfon az a φ_{ij} folyam, melyre

$$\theta p + \sum_{\substack{c_{ij}>\varphi_{ij}\\(i,j)\in A}} (c_{ij}-\varphi_{ij})b_{ij} - \sum_{\substack{\varphi_{ij}>c_{ij}\\(i,j)\in A}} (\varphi_{ij}-c_{ij})a_{ij} \to \text{minimális}.$$

A duál feladat célfüggvényének első tagját átalakítjuk, hogy a később kimondott tételek könnyebben bizonyíthatók legyenek. Felhasználva a $\pi_t = p$ és a $\pi_s = 0$ összefüggéseket

$$\theta p = (\pi_s(-\theta)) + (\pi_2 0) + \cdots + (\pi_j 0) + \cdots + (\pi_t\theta) =$$

$$= \sum_{j\in N}\pi_j\left[\sum_{\substack{j\\(i,j)\in A}}\varphi_{ij} - \sum_{\substack{j\\(i,j)\in A}}\varphi_{ji}\right] = \sum_{j\in N}\pi_j\sum_{\substack{j\\(i,j)\in A}}\varphi_{ij} - \sum_{i\in N}\pi_i\sum_{\substack{i\\(i,j)\in A}}\varphi_{ij} =$$

$$= \sum_{j\in N}\pi_j\varphi_{ij} - \sum_{i\in N}\pi_i\varphi_{ij} = \sum_{A}(\pi_j-\pi_i)\varphi_{ij}.$$

A célfüggvény az átalakítás után tehát a következő:

$$\sum_{A}(\pi_j-\pi_i)\varphi_{ij} + \sum_{\substack{c_{ij}>\varphi_{ij}\\(i,j)\in A}} (c_{ij}-\varphi_{ij})b_{ij} - \sum_{\substack{\varphi_{ij}>c_{ij}\\(i,j)\in A}} (\varphi_{ij}-c_{ij})a_{ij} \to \min.$$

A primál és duál feladat között szoros összefüggés van, melyet az alábbi lemmában mondunk ki.

LEMMA. *Ha létezik a primál feladatnak megfelelő τ és π politika, valamint egy φ folyam, akkor a*

$$\sum_{A} c_{ij}\tau_{ij} \le \sum_{A}(\pi_j-\pi_i)\varphi_{ij} + \sum_{\substack{c_{ij}>\varphi_{ij}\\(i,j)\in A}} (c_{ij}-\varphi_{ij})b_{ij} - \sum_{\substack{\varphi_{ij}>c_{ij}\\(i,j)\in A}} (\varphi_{ij}-c_{ij})a_{ij}$$

és egyenlőség akkor és csak akkor van, ha

1°	$\pi_j - \pi_i$	$> \tau_{ij}$	*esetén*	$\varphi_{ij} = 0$
2°	b_{ij}	$> \tau_{ij}$	*esetén*	$\varphi_{ij} \ge c_{ij}$
3°	a_{ij}	$< \tau_{ij}$	*esetén*	$\varphi_{ij} \ge c_{ij}$.

Bizonyítás. Egy olyan élen, ahol $\varphi_{ij} < c_{ij}$ ott a lemma szerint

$$c_{ij}\tau_{ij} \leq (\pi_j - \pi_i)\varphi_{ij} + (c_{ij} - \varphi_{ij})b_{ij}$$

A $(\pi_j - \pi_i)$ és b_{ij} helyére a nálunk kisebb vagy egyenlő τ_{ij}-t írva, az egyenlőtlenségből egyenlőség lesz, tehát az eredeti állítás igaz. Egyenlőség pedig csak akkor van, ha

1°°	$\varphi_{ij} > 0$	esetén	$\pi_j - \pi_i = \tau_{ij}$
2°°	$c_{ij} > \varphi_{ij}$	esetén	$b_{ij} = \tau_{ij}$.

Egy olyan élen, ahol $\varphi_{ij} < c_{ij}$ ott a lemma szerint

$$c_{ij}\tau_{ij} \leq (\pi_j - \pi_i)\varphi_{ij} - (\varphi_{ij} - c_{ij})a_{ij}.$$

A $(\pi_j - \pi_i)$ helyére a nála kisebb vagy egyenlő τ_{ij}-t írva, és az a_{ij} helyére a nála nagyobb vagy egyenlő τ_{ij}-t írva, (azonban ez nagyobb szám kivonását jelenti, mint az eredeti kivonásban) az egyenlőtlenségből egyenlőség lesz, tehát az eredeti állítás igaz. Egyenlőség pedig csak akkor van, ha

1°°	$\varphi_{ij} > 0$	esetén	$\pi_j - \pi_i = \tau_{ij}$
3°°	$c_{ij} < \varphi_{ij}$	esetén	$a_{ij} = \tau_{ij}$.

Összegezve az előzőeket, a lemmában egyenlőség akkor és csak akkor van, ha

1°°	$\varphi_{ij} > 0$	esetén	$\pi_j - \pi_i = \tau_{ij}$
2°°	$c_{ij} > \varphi_{ij}$	esetén	$b_{ij} = \tau_{ij}$
3°°	$c_{ij} < \varphi_{ij}$	esetén	$a_{ij} = \tau_{ij}$.

Megfordítva,

1°	$\pi_j - \pi_i = \tau_{ij}$	esetén	$\varphi_{ij} = 0$
2°	$b_{ij} > \tau_{ij}$	esetén	$\varphi_{ij} \geq c_{ij}$
3°	$a_{ij} < \tau_{ij}$	esetén	$\varphi_{ij} \leq c_{ij}$.

Ez éppen a lemma állítása.

A lemmának egy lényeges következménye van, amit az alábbiakban ismertetünk.

TÉTEL (gyenge equilibrium). *Ha létezik a primál feladatot tetszőleges p átfutási idő mellett kielégítő π_i és τ_{ij} politika, és φ_{ij} a tervütemhálón, valamint a primál feladat célfüggvény értéke (P) egyenlő a duál feladat célfüggvényével (D), azaz $P = D$, akkor mindkét feladat megoldása optimális.*

Bizonyítás (indirekt úton). Tegyük fel, hogy $P = D$, de létezik egy olyan P^* megoldás, hogy $P^* > P$. Ekkor $P^* > P = D$ de ez a lemma szerint ellentmondás.

Tegyük fel, hogy $P = D$, de létezik egy olyan jobb duál megoldás (D^*), hogy $D^* < D$. Ekkor $D^* < D = P$ de ez a lemma szerint ellentmondás, így ezzel a tétel bizonyítva van.

TÉTEL (dualitási tétel). *Tetszőleges p átfutási időhöz ($p_{\min} \leq p \leq p_{\max}$) megadható olyan π, τ és φ rendszer, melyeknél a célfüggvények értékei egyenlők, azaz optimálisak.*

A tétel bizonyítása konstruktív, azaz megadja az algoritmust is.

Bizonyítás. A $p = p_{\max}$ átfutási időhöz tartozó megoldás triviálisan áll elő. A π_i rendszer legyen a $\tau_{ij} = b_{ij}$ értékekből előállított időpolitika, ekkor $\pi_t = p_{\max}$. A φ_{ij} legyen zérus minden élen. Ez egy optimális megoldás, hiszen a primál és a duál célfüggvény értéke azonos, $\sum_A c_{ij} b_{ij}$.

Ha valamely p-re van optimális megoldás, akkor megadható olyan p^*, π^*, τ^* és ϕ^*, amely lemmát teljesíti, azaz szintén optimális, és $p^* < p$. A továbbiakban ezt az állítást igazoljuk.

Az állítás tehát azt mondja ki, hogyha van egy p átfutási időhöz tartozó optimális megoldás, akkor áttérhetünk egy kisebb átfutási időhöz tartozó optimális megoldásra. Mivel a maximális átfutási időhöz tartozó optimális megoldást ismerjük, ezért az összes p átfutási időhöz tartozó optimális megoldást meg lehet adni a $p_{\min}$, $p_{\max}$ intervallumban. Ezt kétlépéses konstrukcióval látjuk be.

Az első lépésben úgy növeljük a φ folyamot φ^*-ra, hogy a lemmában megfogalmazott dualitási feltételek érvényesek maradjanak, azaz a megoldás továbbra is optimális legyen.

A második lépésben csökkentjük a p átfutási időt úgy, hogy a dualitási feltételek továbbra is teljesüljenek, azaz a megoldás továbbra is optimális legyen.

Első lépés:

Vizsgáljuk meg, hogy egy élen mely dualitási feltételek milyen kombinációban teljesülhetnek. A „+" jelzi azt, hogy az adott dualitási feltétel teljesül, a „−" jelzi, hogy nem.

A lehetséges nyolc variációból csak a $(+++)$ és a $(++-)$ nem szerepel, ugyanis egy optimális megoldáshoz tartozó pozitív költségtényezőjű élen ez a kombináció nem fordulhat elő. Ha mindhárom feltétel teljesülne ugyanis egy élen, akkor az azt jelentené, hogy a τ_{ij} tevékenység idő a normál és a rohamidő közt van, és a $\pi_j - \pi_i > \tau_{ij}$ azaz a tevékenységnek van tartalékideje. Ebben az esetben azonban, ($c_{ij} \geq 0$ esetén) ha megnöveljük τ_{ij}-t egy időegységgel, akkor egy jobb megoldást kapunk ugyanahhoz a p-hez, következésképpen egy olyan él, amelyre $c_{ij} \geq 0$ és amelyen mindhárom feltétel teljesül, nem tartozhat egyetlen optimális megoldáshoz sem. Ugyanez az okoskodás zárja ki a $(++-)$ kombinációt is. Ha $c_{ij} = 0$ és a megoldás optimális, akkor az algoritmus mechanizmusa garantálja, hogy egy nemkritikus úton levő élre csak a $(+--)$ és $(+-+)$ kombinációk teljesülhetnek.

Attól függően, hogy mely feltételek premisszái teljesülnek egy élen, az ott átmenő folyamokról bizonyos információkat kaphatunk. Ezeket az információkat a negyedik oszlopban tüntettük fel. Az ötödik oszlopban az éleket soroltuk be öt

csoportba ($\mathcal{A}1 - \mathcal{A}5$), aszerint, hogy rajtuk mely egyensúlyi feltételek teljesülnek.

Feltételek premisszái			Folyam információk	Élek besorolása	κ_{ij}	κ_{ji}
1°	2°	3°				
+	−	+	$\varphi_{ij} = 0$	$\mathcal{A}1$	0	0
+	−	−	$\varphi_{ij} = 0$	$\mathcal{A}1$	ua.	ua.
−	+	+	$\varphi_{ij} = c_{ij}$	$\mathcal{A}2$	0	0
−	+	−	$\varphi_{ij} \geq c_{ij}$	$\mathcal{A}3$	∞	$\varphi_{ij} - c_{ij}$
−	−	+	$\varphi_{ij} \leq c_{ij}$	$\mathcal{A}4$	$c_{ij} - \varphi_{ij}$	φ_{ij}
−	−	−	$\varphi_{ij} \geq 0$	$\mathcal{A}5$	∞	φ_{ij}

Az első lépésben a folyamot kell úgy növelni, hogy a megoldás továbbra is optimális legyen, azaz a dualitási feltételek minden élre teljesüljenek. Ha a folyamokat úgy növeljük, hogy az élek továbbra is ugyanabban az él osztályban maradnak, akkor a megoldás továbbra is optimális lesz. A folyam információkból nyert adatok segítségével meg lehet mondani, hogy egy adott élen mennyivel növelhető, ill. csökkenthető a folyam. Az ilyen okoskodással előállított él kapacitásokat mutatja a hatodik és hetedik oszlop.

A fenti kapacitásokkal megadott hálózaton maximális folyamot keresve, ψ_{ij} folyamot kapunk. Ezt hozzáadva az eredeti φ_{ij} folyamhoz az új folyamérték az éleken

$$\varphi^*_{ij} = \varphi_{ij} + \psi_{ij} \quad \forall (i,j) \in \mathcal{A}.$$

Ezzel az első lépést, melynek célja az, hogy a folyamot növeljük a dualitási feltételek betartásával, befejeztük.

Második lépés:

A megnövelt φ^*_{ij} folyamhoz keresünk egy olyan π^*_i, τ^*_{ij}, $p^* < p$ rendszert, hogy a dualitási feltételek teljesüljenek, azaz a megoldás optimális maradjon. Ezalatt azt értjük, hogy a π értékeket úgy változtatjuk, hogy az él a rajta levő folyamnak megfelelő élosztályba kerüljön. Egyes élek besorolása ezután természetesen megváltozik, mint ahogy azonnal megváltozik egyes tevékenységek tevékenységideje is, hiszen az optimális megoldásnak megfelelően $\tau_{ij} = \min\{\pi_j - \pi_i; b_{ij}\} \quad \forall (i,j) \in \mathcal{A}$.

Az elő lépésben maximális folyamot kerestünk. Ekkor létezik egy olyan (S,T) vágás, melynek élei telítettek. A vágásban $\mathcal{A}1$, $\mathcal{A}2$, és $\mathcal{A}4$ típusú élek lehetnek. A vágásban visszafelé mindegyik éltípus előforduhat. Ezeken a vágásban levő (i,j) éleken csökkentjük a π_j potenciálokat egy alkalmasan választott δ értékkel. A vágásban visszafelé menő (i,j) élen az i esemény potenciálját csökkentjük δ értékkel.

A δ értékek meghatározása az alábbi módon történik.

Ha az él vágásban van:

$\mathcal{A}1$ típusú él esetén $\delta_1 := \min\{\pi_j - \pi_i - b_{ij} : (i,j) \in (S,T), (i,j) \in A1\}$
ekkor az él $\mathcal{A}4$ típusú lesz. A π_j ennél kisebb csökkentése esetén az él $\mathcal{A}1$ típusú marad.

$\mathcal{A}2$ típusú él esetén $\delta_2 := \min\{\pi_j - \pi_i - a_{ij} : (i,j) \in (S,T), (i,j) \in A2\}$
ekkor az él $\mathcal{A}3$ típusú lesz. A π_j ennél kisebb csökkentése esetén az él $\mathcal{A}2$ típusú marad.

$\mathcal{A}4$ típusú él esetén $\delta_4 := \min\{\pi_j - \pi_i - a_{ij} : (i,j) \in (S,T), (i,j) \in A4\}$
ekkor az él $\mathcal{A}3$ típusú lesz. A π_j ennél kisebb csökkentése esetén az él $\mathcal{A}2$ típusú lesz.

Ha az él a vágásban visszafelé megy:

$\mathcal{A}1$ típusú él esetén δ_{1^*} bármekkora lehet az él $\mathcal{A}1$ típusú marad.

$\mathcal{A}2$ típusú él esetén $\delta_{2^*} := \min\{b_j - \pi_j + \pi_i : (i,j) \in (T,S), (i,j) \in A2\}$
ekkor az él $\mathcal{A}4$ típusú lesz. A π_i ennél kisebb csökkentése esetén az él $\mathcal{A}2$ típusú marad. Ennél nagyobb csökkentés nem engedhető meg, mert úgy az él nem a folyamnak megfelelő időértékeket fogja felvenni, azaz a megoldás nem lesz optimális.

$\mathcal{A}3$ típusú él esetén $\delta_{3^*} := \min\{b_j - \pi_j + \pi_i : (i,j) \in (T,S),\ (i,j) \in A3\}$
ekkor az él $\mathcal{A}4$ típusú lesz. A π_i ennél kisebb csökkentése esetén az él $\mathcal{A}2$ típusú lesz. Ennél nagyobb csökkentés nem engedhető meg, mert úgy az él nem a folyamnak megfelelő időértékeket fogja felvenni, azaz a megoldás nem lesz optimális.

$\mathcal{A}4$ típusú él esetén δ_{4^*} bármekkora lehet az él $\mathcal{A}1$ típusú lesz.

$\mathcal{A}5$ típusú él esetén δ_{5^*} bármekkora lehet az él $\mathcal{A}1$ típusú lesz.

Ezen adatok ismeretében δ a következő módon határozható meg.

$$\delta := \min\{\delta_1, \delta_2, \delta_4, \delta_{2^*}, \delta_{3^*}\}.$$

Az így előállított δ érték határozottan nagyobb mint zérus. Legyen az új π_i^* potenciálrendszer az alábbiak szerint meghatározva:

$$\pi_i^* := \begin{cases} \pi_i & \text{ha } i \in S, \\ \pi_i - \delta & \text{ha } i \in T. \end{cases}$$

Az olyan éleken, ahol mindkét csomópont az S, vagy mindkét csomópont a T ponthalmazban volt, a dualitási feltételek automatikusan teljesülnek. A vágásban levő éleken pedig azért teljesülnek, mert δ meghatározásánál pontosan az volt a cél, hogy ezek a feltételek továbbra is teljesüljenek.

Így $p^* = p - \delta$ lesz, egy (i,j) tevékenységidő pedig az alábbi képletből határozható meg:

$$\tau_{ij}^* = \min\{\pi_j^* - \pi_i^*; b_{ij}\} \quad \forall (i,j) \in \mathcal{A}.$$

Azok az $\mathcal{A}4$ és $\mathcal{A}5$ típusú élek, melyek a vágásban visszafelé mennek, már $\delta = 1$ esetén is $\mathcal{A}1$ típusú éllé válnak. Ez azt jelenti, hogy az (i,j) tevékenység, amely eddig kritikus volt, azaz π_j-t ez a tevékenység határozta meg, a továbbiakban nem lesz kritikus. Ha a j csomóponthoz ezután nem vezet kritikus út, akkor $\mathcal{A}4$ és $\mathcal{A}5$ visszamenő vágásbeli éleknél a δ-t csak zérusnak lehet választani. Ez esetben

azonban nem garantálható, hogy az algoritmus nem áll le valahol mielőtt elérné a $p_{\min}$ átfutási időt azzal, hogy p-t nem lehet tovább csökkenteni. A δ értéket ezeknél az éleknél csak akkor lehet tetszőlegesen nagynak venni, amennyiben az algoritmus automatizmusa garantálja, hogy a j csomóponthoz vezet egy másik kritikus út is. Ezt az automatizmust könnyen igazolni lehet. Mivel az adott $(i, j) \in \mathcal{A}$ él a vágásban visszafelé megy, azaz a visszamenő (j, i) él a telített (φ_{ij} folyam folyik rajta), ezért a j csomópontba valamennyi (legalább φ_{ij}) folyamnak be kellett folynia. Mivel folyam csak kritikus éleken lehet, ezért a j csomópontba egy másik csomópontból is vezetett kritikus út, nem csak i-ből. Azaz az algoritmus automatikusan teljesíti azt az elvárást hogy egy $\mathcal{A}4$ és $\mathcal{A}5$ típusú él csak akkor lehet visszafelé a vágásban, ha j csomópontba máshonnan is vezet kritikus út.

Ezzel a bizonyítás második lépése kész.

Ezután az első lépésre visszatérve a folyam ismét növelhető, majd a második lépésben a p átfutási idő csökkenthető. Ezeket a lépéseket addig kell felváltva ismételni, amíg a kívánt átfutási időt el nem érjük. Célszerű a feladat megoldásának kezdetén $p_{\max}$ átfutási idő kiszámítása mellett, a $p_{\min}$ átfutási időt is kiszámítani az a_{ij} értékből. Így ellenőrizhető, hogy az általunk megadott átfutási idő egyáltalán megvalósítható-e.

Ezzel a tétel konstruktívan bizonyítva van.

Az alábbiakban a tétel egy fontos következményét ismertetjük.

TÉTEL (erős equilibrium). *Ha létezik optimális primált kielégítő (P) megoldás és optimális duál (D) megoldás, akkor azok értéke egyenlő.*

Bizonyítás. A dualitási tétel értelmében van optimális primál megoldás (P^*) és minimális duál megoldás (D^*) amelyek optimálisak, azaz $(P^*) = (D^*)$. Mivel (P) és (D) is optimálisak, ezért $(P^*) = (P)$ és $(D^*) = (D)$, de ekkor $(P) = (D)$.

3. Mintafeladat az algoritmus illusztrálására

Adott az ábrán látható tervütem háló, a táblázatban megadott normál- és rohamidő adatokkal. Adottak még a tevékenységekhez rendelt költségtényezők (c_{ij}).

	A	B	C	D	E	F	G	H
b_{ij}	4	7	3	5	2	10	7	2
a_{ij}	3	5	2	3	2	8	5	1
c_{ij}	10	12	5	8	0	6	14	10

Keressük a $p_{\min}$ értékhez tartozó optimális megoldást.

0. lépés ($p_{\max}$ és $p_{\min}$) meghatározása

él	A	B	C	D	E	F	G	H
τ_{ij}	4	7	3	5	2	10	7	2

csomóp.	1	2	3	4	5
π_i	0	4	3	9	16

$$p_{\max} = \pi_t = 16$$

él	A	B	C	D	E	F	G	H
τ_{ij}	3	5	2	3	2	8	5	1

csomóp.	1	2	3	4	5
π_i	0	3	2	6	11

$$p_{\min} = \pi_t = 11$$

Vegyük kiindulásképpen a $\tau_{ij} = b_{ij}$ időértékeket, valamint az ezekből számított eseményidőket, és a $\varphi_{ij} = 0$ folyamot. Ez a $p_{\max}$ átfutási értékhez tartozó optimális megoldás.

1. lépés (folyam növelés)

él	A	B	C	D	E	F	G	H
élek oszt. $\mathcal{A}1 - \mathcal{A}5$	4	1	4	4	1	1	4	1
kapacitások κ_{ij}	10	0	5	8	0	0	14	0
κ_{ji}	0	0	0	0	0	0	0	0
folyamok ψ_{ij}	8	0	0	8	0	0	8	0
új folyam $\varphi_{ij} + \psi_{ij}$	8	0	0	8	0	8	0	8

2. lépés (π_i, τ_{ij}, p csökkentése)

élek a vágásban	B	E	H	F	D
δ értékek	2	4	11	2	2
δ min	2				

csomópontok	1	2	3	4	5
régi π értékek	0	4	3	9	19
új π értékek	0	4	3	7	14

él	A	B	C	D	E	F	G	H
új tevékenységidő	4	7	3	3	2	10	7	2

Az új átfutási idő $p = 14$.
A költségek az előző átfutási időhöz képest 16 egységgel növekedtek, azaz időegységenként 8 költség egységgel.

1*. lépés (folyam növelés)

él	A	B	C	D	E	F	G	H
élek osztályozása	4	4	4	3	1	4	4	1
kapacitások κ_{ij}	2	12	5	∞	0	6	6	0
κ_{ji}	8	0	0	0	0	0	8	0
folyamok ψ_{ij}	2	6	0	0	0	2	6	0
új folyam $\varphi_{ij} + \psi_{ij}$	10	6	0	8	0	2	14	0

2*. lépés (π_i, τ_{ij}, p csökkentése)

élek a vágásban	A	G	H	D^*
δ értékek	1	2	9	2
δ min			1	

csomópontok	1	2	3	4	5
régi π értékek	0	4	3	7	14
új π értékek	0	3	3	7	13

él	A	B	C	D	E	F	G	H
új tevékenységidő	3	7	3	4	2	10	6	2

A D él a vágásban visszafelé megy.
Az új átfutási idő $p = 13$.
A költségek az előző átfutási időhöz képest 16 egységgel növekedtek, azaz egy időegység alatt 16 pénzegységgel.

1**. lépés (folyam növelés)

él	A	B	C	D	E	F	G	H
élek osztályozása	3	4	4	2	1	4	2	1
kapacitások κ_{ij}	∞	6	5	0	0	4	0	0
κ_{ji}	0	6	0	0	0	2	0	0
folyamok ψ_{ij}	4	0	0	0	0	4	0	0
új folyam $\varphi_{ij} + \psi_{ij}$	14	6	0	8	0	6	14	0

2**. lépés (π_i, τ_{ij}, p csökkentése)

élek a vágásban	F	G	H
δ értékek	2	1	6
δ min		1	

csomópontok	1	2	3	4	5
régi π értékek	0	3	3	7	13
új π értékek	0	3	3	7	12

él	A	B	C	D	E	F	G	H
új tevékenységidő	3	7	3	4	2	9	5	2

Az új átfutási idő $p = 12$.
A költségek az előző átfutási időhöz képest 20 egységgel növekedtek, azaz időegységenként 20 pénzegységgel.

1***. lépés (folyam növelés)

él	A	B	C	D	E	F	G	H
élek osztályozása	3	4	4	2	1	2	3	1
kapacitások κ_{ij}	∞	6	5	0	0	0	∞	0
κ_{ji}	4	6	0	0	0	0	0	0
folyamok ψ_{ij}	0	6	0	0	0	0	6	0
új folyam $\varphi_{ij} + \psi_{ij}$	14	12	0	8	0	6	20	0

2***. lépés (π_i, τ_{ij}, p csökkentése)

élek a vágásban	F	D	B	E	H
δ értékek	1	1	2	2	7
δ min	1				

csomópontok	1	2	3	4	5
régi π értékek	0	3	3	7	12
új π értékek	0	3	3	6	11

él	A	B	C	D	E	F	G	H
új tevékenységidő	3	6	3	3	2	8	5	2

Az új átfutási idő $p = 11$.
A költségek az előző átfutási időhöz képest 26 egységgel növekedtek, azaz időegységenként 26 pénzegységgel.
Mivel $p = p_{\min} = 11$, ezért a feladatot megoldottuk.

4. Összegzés

A dolgozatban egy új eljárást adtunk a CPM/cost feladatra. Az eljárásunk a hálózati folyamok elméletén alapul, és felhasználja a lineáris programozás dualitási elméletét. A megoldásunk elméletileg gyorsabb a FULKERSON (1961) által közöltnél, mert vele ellentétben mi nem duplázzuk meg az élek számát. Az algoritmusunk a maximális folyam feladat ismétlődő végrehajtásán alapul. A megoldás során annyiszor kell megismételni a folyam feladatot, ahány törés van a költséggörbében a maximális átfutási idő és az általunk megadott átfutási idő között.

Mivel a maximális feladattal kapcsolatban egyre újabb és jobb algoritmusok látnak napvilágot, ezért a feladathoz szükséges lépésszám alapvetően a maximális folyam feladat megoldására felhasználható algoritmustól függ. A feladat megoldása során bármely létező algoritmus felhasználható. Ha valaki gyors programot akar készíteni a feladatra, javasoljuk, hogy az irodalomban közölt algoritmusokat a gyakorlatban is vizsgálja meg, mert azok elméleti gyorsasága a tervütem háló speciális szerkezete miatt nem biztos hogy érvényesül.

IRODALOM

[1] FORD, L. R. and FULKERSON, D. R., „Maximal Flow Through A Network", *Canadian J. Math.* **8** (1956), 399–404.

[2] FULKERSON, D. R., „A Network Flow Computation for Project Cost Curves", *Management Sci.* **7** (1961), 167–178.

[3] KELLEY, J. E. and WALKER, M.R., *Critical Path Planning and Scheduling* (Proc. of Eastern Joint Computer Conf, Boston, 1959).

[4] KELLEY, J. E., „Critical Path Planning and Scheduling: Mathematical Basis", *Op. Res.* 9 (1961), 296–320.

(Beérkezett: 1992. október 20.)

KLAFSZKY EMIL
BUDAPESTI MŰSZAKI EGYETEM
ÉPÍTÉSKIVITELEZÉSI TANSZÉK
1111 BUDAPEST, MŰEGYETEM RKP. 3. KII/17.
HAJDU MIKLÓS
BUDAPESTI MŰSZAKI EGYETEM
ÉPÍTÉSKIVITELEZÉSI TANSZÉK
1111 BUDAPEST, MŰEGYETEM RKP. 3. KII/17.

A NEW CPM TIME–COST TRADE–OFFS ALGORITHM

E. KLAFSZKY and M. HAJDU

In this paper we give a fast, new algorithm for CPM Time-Cost Trade-Offs problem. Starting from an optimal solution according to the maximal project duration, we can reach the optimal solution according to a smaller project duration. The algorithm is based on network flow theory.

Alkalmazott Matematikai Lapok **17** (1993) 225–239

A KÖLTSÉGTERVEZÉSI FELADAT MEGOLDÁSA KÜLÖNBÖZŐ FÜGGŐSÉGI KAPCSOLATOKAT TARTALMAZÓ TEVÉKENYSÉG CSOMÓPONTÚ HÁLÓ ESETÉN (MPM/COST FELADAT)

HAJDU MIKLÓS

Budapest

Cikkünkben az először KELLEY és WALKER [5] később FULKERSON [3] által megoldott CPM/cost feladatot terjesztjük ki az eltérő függőségi kapcsolatokat tartalmazó tevékenység csomópontú hálóra. A tevékenységek között csak minimális eltávolodás jellegű kapcsolatokat engedjük meg. Ezek a Start–Start-t (SSt), Finish–Start-t (FSt), Start–Finish-t (SFt), Finish–Finish-t (FFt). A feladat megoldása a hálózati folyamok elméletén alapul.

1. Bevezetés

A dolgozat címében említett hálótípust a hazai és egyes európai országok gyakorlatában MPM (Metra Potential Method) hálónak nevezik, míg az angolszász szakirodalom a Precedence Diagramming Method (PDM) elnevezést használja. Az MPM (PDM) hálótechnika korai verziói ROY [6] és FONDAHL [2] nevéhez fűződnek. Roy munkássága nyomán került be az MPM elnevezés a köztudatba. Fondahl eredményeit felhasználva az IBM egyik munkacsoportja J. D. CRAIG [1] vezetésével fejlesztette ki azt, az általuk PDM hálónak nevezett technikát, amelyet ma a gyakorlatban használnak és MPM ill. PDM technikának neveznek. Erre a hálótípusra terjesztjük ki a dolgozatban a költségoptimalizálási feladatot, melyet először KELLEY és WALKER [5] oldottak meg tevékenység–él típusú hálóra és melyet CPM (Critical Path Method) néven publikáltak. A CPM feladatra később KELLEY [4] és FULKERSON [3] is közöltek új megoldást. Dolgozatunkban az MPM hálóra oldjuk meg a költségtervezési feladatot, melyre a továbbiakban MPM/cost néven hivatkozunk.

2. A feladat modellje, és megoldása

Adott egy $[N, \mathcal{A}]$ irányított gráf. A gráfnak egy kezdő és egy végpontja lehet, és minden $i \in N$ ponton kell útnak vezetnie a kezdő pontból (s) a végpontba (t). Kettős illetve többszörös élek megengedettek, hurok nem. A csomópontok tevékenységeket reprezentálnak, az i-edik tevékenység időtartama τ_i. A tevékenységek végzése időben folyamatosan, lineárisan történik, a tevékenységek megszakítása nem

megengedett. Az élek a tevékenységek között technológiai és szervezési kapcsolatok leírására szolgálnak. Tetszőleges két tevékenység között az alábbi kapcsolatok engedhetők meg:

Start–Start–z_{ij} (SSz_{ij}); Finish–Start–z_{ij} (FSz_{ij});
Finish-Finish–z_{ij} (FFz_{ij}); Start–Finish–(SFz_{ij}).

A kapcsolatok az i tevékenység kezdete (vége) és a j tevékenység kezdete (vége) közti minimális távolságot adják meg.

- SSz_{ij} – az i tevékenység kezdete és a j tevékenység kezdete között legalább z_{ij} vagy annál több időnek kell eltelnie.
- FSz_{ij} – az i tevékenység befejezése és a j tevékenység kezdete között legalább z_{ij} vagy annál több időnek kell eltelnie.
- SFz_{ij} – az i tevékenység kezdete és a j követő tevékenység befejezése között legalább z_{ij} vagy annál több időnek kell eltelnie.
- FFz_{ij} – az i tevékenység befejezte és a következő tevékenység befejezte között legalább z_{ij} vagy annál több időnek kell eltelnie.

A fentiekben felsorolt kapcsolatokat minimális kapcsolatoknak nevezzük, mert z_{ij} a tevékenységek kitüntetett pontja közti minimális távolságot jelenti. A hálótervezésben használják még az ún. maximális kapcsolatokat is, ahol szintén az előbb felsorolt kapcsolatokat lehet alkalmazni azzal a különbséggel, hogy ott a z_{ij} a tevékenységek kitüntetett pontjai közti megengedhető maximális távolságot jelenti. A maximális kapcsolatokról csak a teljesség kedvéért teszünk említést, mert a költségtervezési feladatot olyan hálóra terjesztjük ki, amelyikben maximális kapcsolat használata nem megengedett.

A tevékenységekre egy alsó és egy felső időkorlát adott, a roham és a normálidő. Jelük a_i és b_i ($a_i \leq b_i$). Legyen adott minden i tevékenység normál idejéhez egy Kn_i normál költség, amely megmutatja, hogy a tevékenység mennyibe kerül, ha megvalósításának ideje b_i. Adott továbbá minden i tevékenységhez egy c_i költségtényező, mely megmutatja, hogy a tevékenység idejének egy napos gyorsításával mennyivel emelkedik a megvalósítási költség. Ezek ismeretében meghatározható a roham időhöz tartozó ún. roham költség:

$$Kr_i = Kn_i + c_i(b_i - a_i).$$

További megkötés c_i-re, hogy nemnegatív legyen, azaz

$$Kr_i \geq Kn_i \quad \forall (i) \in N.$$

A tényleges tevékenység idő τ_i. A tényleges tevékenységidőnek a roham- és a normálidő közé kell esnie

$$a_i \leq \tau_i \leq b_i \quad \forall (i) \in N. \tag{1}$$

Az i tevékenység kezdetét jelöljük π_{iS} a befejését π_{iF} értékkel. Mivel a tevékenységek időben folyamatosak, ezért π_{iS} meghatározza π_{iF} értéket, és viszont. A tevékenységek közti minimális kapcsolatok alapján az alábbi feltételek írhatók fel.

$$\pi_{jS} - \pi_{iS} \geq z_{ij} \quad \forall (i,j) \in \mathcal{A} \text{ és } \mathrm{SS}z_{ij} \tag{2a}$$

$$\pi_{jF} - \pi_{iF} \geq z_{ij} \quad \forall (i,j) \in \mathcal{A} \text{ és } \mathrm{FF}z_{ij} \tag{2b}$$

$$\pi_{jF} - \pi_{iS} \geq z_{ij} \quad \forall (i,j) \in \mathcal{A} \text{ és } \mathrm{SF}z_{ij} \tag{2c}$$

$$\pi_{jS} - \pi_{iF} \geq z_{ij} \quad \forall (i,j) \in \mathcal{A} \text{ és } \mathrm{FS}z_{ij}. \tag{2d}$$

Kiküszöbölve a tevékenységek befejeztét a feltételekből, és bevezetve a $\tau^F_{i,j}$ és $\tau^F_{j,j}$ jelöléseket a (2a–d) feltételek a következőképpen módosulnak.

$$\pi_j - \pi_i + \tau^F_{j,j} - \tau^F_{i,j} \geq z_{ij} \quad \forall (i,j) \in \mathcal{A} \tag{2}$$

ahol π_i, π_j a tevékenységek kezdete,

$$\tau^F_{j,j} = \begin{cases} \tau_i & \text{Ha az } (i,j) \text{ kapcsolat a } j\text{-edik tevékenység végébe fut,} \\ & \text{azaz a kapcsolat SF vagy FF jellegű.} \\ 0 & \text{Ha az } (i,j) \text{ kapcsolat a } j\text{-edik tevékenység elejébe fut,} \\ & \text{azaz a kapcsolat SS vagy FS jellegű.} \end{cases}$$

$$\tau^F_{i,j} = \begin{cases} \tau_j & \text{Ha az } (i,j) \text{ kapcsolat a } i\text{-edik tevékenység végéből fut,} \\ & \text{azaz a kapcsolat FS vagy FF jellegű.} \\ 0 & \text{Ha az } (i,j) \text{ kapcsolat a } j\text{-edik tevékenység elejéből fut ki,} \\ & \text{azaz a kapcsolat SS vagy FS jellegű.} \end{cases}$$

Definíció. Egy adott kapcsolatot egy adott tevékenységre nézve befejezés típusú kapcsolatnak nevezünk, ha a kapcsolat a tevékenység befejeztéből fut ki, vagy oda érkezik be. Befutó kapcsolatoknál befejezés típusú a SF és a FF kapcsolat, kifutó kapcsolatoknál befejezés típusú a FS és FF kapcsolat.

A definíció ismeretében $\tau^F_{i,j}$ úgyis megfogalmazható, hogy ha a kapcsolat a tevékenységre nézve befejezés típusú akkor $\tau^F_{i,j} = \tau_i$ egyébként zérus.

A start tevékenység kezdete legyen zérus az utolsó tevékenység befejezte legyen p

$$\pi_s = 0 \tag{3}$$

$$\pi_t + \tau_t = p. \tag{4}$$

Egy olyan tervütem hálón, ahol a τ_i nem konkrét, hanem (1)-nek megfelelően egy alsó és felső időkorlát között mozoghat, többféle átfutási idő érhető el. Magától értetődően, az általunk megkívánt p átfutási időnek az elérhető minimális és maximális átfutási idő közt kell lennie

$$p_{\min} \leq p \leq p_{\max}. \tag{5}$$

A tevékenység idők változtatásával azonos p átfutási időt is többféleképpen lehet elérni. Nyilvánvaló, ha a tevékenységek költségtényezője különbözik, úgy azonos p átfutási idejű megoldáshoz különböző megvalósítási költségek tartoznak. Mi egy adott átfutási időhöz tartozó legolcsóbb megoldást keressük. Az I/1 ábra segítségével felírhatjuk a következő célfüggvényt:

$$\min \left(\sum_N \{Kn_i + (b_i - \tau_i)c_i\} \right). \tag{6a}$$

Mivel Kn_i és $(b_i c_i)$ konstansok, ezért az előzővel egyenértékű célfüggvény a következő:

$$\max \left(\sum_N (c_i \tau_i) \right). \tag{6}$$

Az előbbiekben elmondottak alapján összefoglalva a fejezet címében megfogalmazott feladat matematikai modellje a következő:

$$\begin{aligned}
&(1) && a_i \leq \tau_i \leq b_i && \forall (i) \in N \\
&(2) && \pi_j - \pi_i + \tau^F_{j_{ij}} - \tau^F_{i_{ij}} \geq z_{ij} && \forall (i,j) \in \mathcal{A} \\
&(3) && \pi_s = 0 && \\
&(4) && \pi_t + t_t = p && \\
&(5) && p_{\min} \leq p \leq p_{\max} && \\
&(6)=(1^*) && \max \left(\sum_N (c_i \tau_i) \right). &&
\end{aligned}$$

Mielőtt a feladathoz egy duál feladatot rendelnénk, néhány új fogalmat vezetünk be.

Definíció. Egy (i,j) élen átmenő φ_{ij} folyamot az i csomópontra nézve befejezés típusúnak nevezünk, és φ^F_{ij}-vel jelöljük, ha a (i,j) élen levő kapcsolat az i csomópont végéből fut ki, azaz a kapcsolat i-re nézve befejezés típusú ($\mathrm{FS}z_{ij}$ vagy $\mathrm{FF}z_{ij}$.)

Definíció. Egy (j,i) élen átmenő φ_{ji} folyamot az i csomópontra nézve befejezés típusúnak nevezünk, és φ^F_{ji}-vel jelöljük, ha a (j,i) élen levő kapcsolat az i csomópont végébe fut be, azaz a kapcsolat i-re nézve befejezés típusú (SFz_{ji}) vagy FFZ_{ji}.)

Adjunk hozzá a hálózathoz egy a t tevékenységből az s tevékenységbe menő FS(-p) kapcsolatot, ahol $c_{ts} = 0$. Ezt az így kibővített élhalmazt jelöljük $\mathcal{A}^*$ szimbólummal. Ez a többlet él a primál feladatot nem változtatja meg.

Ezek után a következő duál feladatot rendelhetjük a primálhoz.

Keresendő az adott hálózaton olyan φ_{ij} folyam, melynek értéke θ és minimalizálja a következő kifejezést.

$$(2^*) \qquad \min \left\{ p\theta - \sum_A z_{ij}\varphi_{ij} + \sum_{\substack{c>F \\ A^*}} [c_i - F_i]b_i - \sum_{\substack{c<F \\ A^*}} [F_i - c_i]a_i \right\}$$

$$\text{ahol } F_i := \sum_j \varphi^F_{ij} - \sum_j \varphi^F_{ji} \quad i \in N \text{ és } (i,j) \in \mathcal{A}^*.$$

Az F_i tehát az i-edik csomópontból kiinduló befejezés típusú folyamok összege csökkentve az i csomópontba menő befejezés típusú folyamok összegével.

A két feladat közti szoros kapcsolatra mutat rá az alábbi lemma.

LEMMA. *Ha létezik primált kielégítő π és τ politika valamint φ_{ij} folyam, akkor* $(1^*) \le (2^*)$, *azaz*

$$\sum_N c_i\tau_i \le p\theta - \sum_A z_{ij}\varphi_{ij} + \sum_{\substack{c>F \\ A^*}} [c_i - F_i]b_i - \sum_{\substack{c<F \\ A^*}} [F_i - c_i]a_i$$

és egyenlőség akkor és csak akkor van, ha

$$(1^\circ) \qquad \pi_j - \pi_i + \tau^F_{j,ij} - \tau^F_{i,ij} > z_{ij} \text{ esetén } \varphi_{ij} = 0$$

$$(2^\circ) \qquad b_i > \tau_i \text{ esetén } c_i \le F_i$$

$$(3^\circ) \qquad a_i < \tau_i \text{ esetén } c_i \ge F_i.$$

Bizonyítás.

$$\sum_N c_i\tau_i \le p\theta - \sum_A z_{ij}\varphi_{ij} + \sum_{\substack{c>F \\ A^*}} [c_i - F_i]b_i - \sum_{\substack{c<F \\ A^*}} [F_i - c_i]a_i.$$

Ha a b_i és az a_i helyére τ_i-t helyettesítünk, akkor a duál célfüggvény értéke biztosan csökken. Ha erre is be tudjuk bizonyítani, hogy a primál célfüggvénynél nagyobb

vagy egyenlő akkor az eredeti állítás is bizonyítva van.

$$\sum_N c_i\tau_i \le p\theta - \sum_A z_{ij}\varphi_{ij} + \sum_{\substack{c>F\\A^*}}[c_i - F_i]\tau_i - \sum_{\substack{c<F\\A^*}}[F_i - c_i]\tau_i =$$

$$= p\theta - \sum_A z_{ij}\varphi_{ij} + \sum_A [c_i - F_i]\tau_i =$$

$$= p\theta - \sum_A z_{ij}\varphi_{ij} + \sum_N c_i\tau_i - \sum_i \left[\sum_{\substack{j\\A^*}} \varphi_{ij}^F\tau_i - \sum_{\substack{j\\A^*}} \varphi_{ji}^F\tau_i\right].$$

Kivonva a $\sum c_i^*\tau_i$ értéket mindkét oldalról, és átrendezve

$$\theta^* p \ge \sum_A z_{ij}{}^*\varphi_{ij} + \sum_i \left[\sum_{\substack{j\\A^*}} \varphi_{ij}^F\tau_i - \sum_{\substack{j\\A^*}} \varphi_{ji}^F\tau_i\right]$$

$$\theta^* p \ge \sum_A z_{ij}{}^*\varphi_{ij} + \sum_A \varphi_{ij}^F\tau_i - \sum_A \varphi_{ji}^F\tau_i + \theta^*\tau_t.$$

A z_{ij} értékek helyébe a nála nagyobb vagy egyenlő $\pi_j - \pi_i + \tau_{j_{ij}}^F - \tau_{i_{ij}}^F$ mennyiséget helyettesítve,

$$\theta p \ge \sum_A (\pi_j - \pi_i)\varphi_{ij} + \sum_A (\tau_{j_{ij}}^F - \tau_{i_{ij}}^F)\varphi_{ij} + \sum_A \varphi_{ij}^F\tau_i - \sum_A \varphi_{ji}^F\tau_i + \theta\tau_t.$$

A definíciók alapján $\tau_{j_{ij}}^f\varphi_{ij} = \varphi_{ji}^F\tau_i$ és $\tau_{i_{ij}}^F\varphi_{ij} = \varphi_{ij}^F\tau_i$, ezért

$$\theta p \ge \sum_A (\pi_j - \pi_i)\varphi_{ij} + \theta\tau_t = \sum_A \pi_t\varphi_{it} + \sum_A \pi_s\varphi_{sj} + \theta\tau_t = (p - \tau_t)\theta + \theta\tau_t$$

$$\theta p \ge p\theta.$$

Ezzel a lemma egyenlőtlensége bizonyítva van, és egyenlőség akkor és csak akkor van ha a b_i, a_i, z_{ij} behelyettesítéseknél

$$\begin{aligned} c_i > F_i \quad &\text{esetén} \quad b_i = \tau_i \\ c_i < F_i \quad &\text{esetén} \quad a_i = \tau_i \\ \varphi_{ij} > 0 \quad &\text{esetén} \quad z_{ij} = \pi_j - \pi_i + \tau_{j_{ij}}^F - \tau_{i_{ij}}^F. \end{aligned}$$

Megfordítva

$$(1^\circ) \qquad \pi_j - \pi_i + \tau_{j_{ij}}^F - \tau_{i_{ij}}^F > z_{ij} \quad \text{esetén} \quad \varphi_{ij} = 0$$

$$(2^\circ) \qquad b_i > \tau_i \quad \text{esetén} \quad c_i \le F_i$$

$$(3^\circ) \qquad a_i < \tau_i \quad \text{esetén} \quad c_i \ge F_i.$$

Ezek éppen a lemmában közölt egyensúlyi feltételek. Az alábbiakban e tétel egy fontos következményét ismertetjük.

TÉTEL (gyenge equilibrium). *Ha létezik a primál feladatot, tetszőleges p átfutási idő mellett, kielégítő π_i és τ_i politika, és φ_{ij} a tervütemhálón, valamint a primál feladat célfüggvény értéke* (1*) *egyenlő a duál feladat célfüggvényével* (2*), *azaz* (1*) = (2*), *akkor mindkét feladat megoldás optimális.*

Bizonyítás (indirekt úton). Jelöljük a primál célfüggvény értékét (1*)-ot P-vel, a duál értékét D-vel.

Tegyük fel, hogy $P = D$, de létezik egy olyan P^* megoldás, hogy $P^* > P$. Ekkor $P^* > P = D$ de ez a lemma szerint ellentmondás.

Tegyük fel, hogy $P = D$, de létezik egy olyan jobb duál megoldás (D^*), hogy $D^* < D$. Ekkor $D^* < D = P$ de ez a lemma szerint ellentmondás, így a tételt bizonyítottuk.

TÉTEL (dualitási tétel). *Tetszőleges p átfutási időhöz ($p_{\min} \leq p \leq p_b$) megadható olyan π, τ és φ rendszer, melyeknél a célfüggvények értékei egyenlőek, azaz optimálisak. (A p_b a normálidőkkel számított átfutási idő.)*

A tétel bizonyítása konstruktív, azaz megadja az alogitmust is

Bizonyítás. Kiinduló triviális megoldásként a π_i rendszer legyen a $\tau_i = b_i$ értékekből előállított időpolitika, és a $\varphi_{ij} = 0$ minden élen. Ez egy optimális megoldás, hiszen a primál és a duál célfüggvény értéke azonos, $\sum_A c_i b_i$. A $\tau_i = b_i$ tevékenységidőhöz tartozó teljes átfutási idő $p = \pi_t + \tau_t$. Ezt az átfutási időt, mivel a normál időkből számítottuk ki, jelöljük p_b-vel. Ha ismerünk egy bármilyen p-hez tartozó π, τ és φ optimális rendszert, akkor megadható olyan p^*, π^*, τ^* és φ^*, amely a lemmát teljesíti v_i, azaz szintén optimális, és $p^* < p$.

Az állítás tehát azt mondja ki, hogyha van egy p átfutási időhöz tartozó optimális megoldás, akkor áttérhetünk egy kisebb átfutási időhöz tartozó optimális megoldásra. Mivel a p_b átfutási időhöz tartozó optimális megoldást ismerjük, ezért az összes p átfutási időhöz tartozó optimális megoldást meg lehet adni, ahol $p < p_b$.

Ezt kétlépéses konstrukcióval látjuk be.

Az első lépésben úgy növeljük a φ folyamot φ^*-ra, hogy a lemmában megfogalmazott dualitási feltételek érvényesek maradjanak, azaz a megoldás továbbra is optimális legyen.

A második lépésben csökkentjük a p átfutási időt úgy, hogy a dualitási feltételek továbbra is teljesüljenek, azaz a megoldás továbbra is optimális legyen. Az átfutási idő csökkentése úgy történik, hogy az elő lépés maximális folyamfeladatánál a vágásban levő élekhez tartozó potenciálok kerülnek változtatásra a primál és a lemmában közölt feltételek betartása mellett. Az első és a második lépés befejeztével egy p-hez tartozó optimális megoldásról egy $p^* < p$ optimális megoldásra térünk át, hiszen a lépések alatt végig teljesülnek a primál- duál- és a lemmában közölt egyensúlyi feltételek.

Első lépés:

Vizsgáljuk meg, hogy egy élen, ill. egy csomóponton mely dualitási feltételek milyen kombinációban teljesülnek. A '+' jelzi azt, hogy az adott dualitási feltétel

teljesül, a '–' jelzi, hogy nem.

Attól függően, hogy mely feltételek premisszái teljesülnek egy élen, ill. egy csomóponton, az ott átmenő φ_{ij}, ill. F_i értékekről bizonyos információkat kaphatunk. Ezek megmutatják, hogy a feltételek premisszáinak teljesülése esetén a φ_{ij} ill. az F_i értékeknek milyennek kell lennie. Ezeket az információkat a negyedik oszlopban tüntettük fel. Az ötödik oszlopban az éleket, ill a csomópontokat soroltuk be csoportokba ($\mathcal{A}1 - \mathcal{A}2$), ($N1 - N4$) a rajtuk áthaladó φ_{ij} ill. F_i értékeket figyelembe véve.

Egyensúlyi feltételek

premisszái 1°	2°	3°	Folyam információk	Élek besorolása	κ_{ij}	κ_{ji}
+	az éleken		$\varphi_{ij} = 0$	$\mathcal{A}1$	0	0
–	nem értelmezzük		$\varphi_{ij} \geq 0$	$\mathcal{A}2$	∞	φ_{ij}

1°	2°	3°	F információk	Cs. pont besor.	$\kappa_{i_S i_F}$	$\kappa_{i_F i_S}$
Csak	+	–	$F_i \geq c_i$	$N1$	∞	$F_i - c_i$
éleken	–	+	$F_i \leq c_i$	$N2$	$c_i - F_i$	∞
értel-	+	+	$F_i = c_i$	$N3$	0	0
mezzük	–	–	F_i bármi	$N4$	∞	∞

Az első lépésben a folyamot kell úgy növelni, hogy a megoldás továbbra is optimális legyen, azaz a dualitási feltételek minden élre teljesüljenek. Ha a folyamokat úgy növeljük, hogy az élek továbbra is ugyanabban az él osztályban ill. a tevékenységek ugyanabban a csomópont osztályban maradnak, akkor a megoldás továbbra is optimális lesz.

A folyam információkból nyert adatok segítségével meg lehet mondani, hogy egy adott élen mennyivel növelhető, ill. csökkenthető a folyam. Az ilyen okoskodással előállított él kapacitásokat mutatja a hatodik és hetedik oszlop.

A csomópontokon az F_i értékeket kell úgy változtatni, hogy a lemma egyensúlyi feltételei érvényesek maradjanak. Az F_i értékek megfelelő módosítását az alábbi technikával lehet biztosítani.

Minden csomópontot ketté vágunk, éllé alakítunk. Az egyik új csomópont reprezentálja a tevékenység kezdetét, a másik a tevékenység végét. Ezt a két csomópontot mindkét irányban éllel kötjük össze. Az (i_S, i_F) él mutasson a tevékenység kezdetéből a végébe, (i_F, i_S) él visszafelé. Az i-re nézve befejezés típusú kapcsolatok az i_F csomópontba, a többi kapcsolat a i_S pontba fusson.

Ekkor az F_i nem más, mint az i_F csomópontból kifutó folyamok összege, csökkentve az i_F-be befutó folyamok összegével, az (i_F, i_S) és (i_S, i_F) éleken áthaladó folyamokat nem számítva.

Az így transzformált hálózaton az (i_S, i_F) és (i_F, i_S) élek kapacitásainak helyes megválasztásával lehet biztosítani, hogy a lemmában az F_i értékekre vonatkozó

feltételek megmaradjanak. Az (i_F, i_S) és (i_S, i_F) éleken a helyes kapacitásokat a fenti táblázat hatodik és hetedik oszlopa mutatja.

A fenti kapacitásokkal megadott transzformált hálózaton maximális folyamot keresve, ψ_{ij} folyamot kapunk. Ezt hozzáadva az eredeti φ_{ij} folyamhoz, visszatérve az eredeti hálóra az új folyamérték az éleken

$$\varphi^*_{ij} = \varphi_{ij} + \psi_{ij} \quad \forall (i,j) \in \mathcal{A}.$$

Ezen folyamszámítás alatt a lemmában közölt egyensúlyi feltételek továbbra is teljesülnek, hiszen a kapacitásokat úgy választottuk meg, hogy az élek és a tevékenységek is besorolásuk szerinti osztályban maradjanak.

Ezzel az első lépést, melynek célja az, hogy a folyamot növeljük a dualitási feltételek betartásával, befejeztük.

Második lépés:

A növelt transzformált hálózaton keresett maximális φ^*_{ij} folyamhoz keresünk egy olyan π^*_i, τ^*_{ij}, $p^* < p$ rendszert, hogy a dualitási feltételek teljesüljenek, azaz a megoldás optimális maradjon. (A transzformált hálózaton π_{iS} a tevékenység korai kezdetét, a π_{iF} a tevékenység korai befejeztét jelöli, azaz $\pi_{iF} - \pi_{iS} = \tau_i$.)

Az első lépésben maximális folyamot keresünk. Ekkor létezik egy olyan (S,T) vágás, a módosított hálózaton, melynek élei telítettek. A vágásban $\mathcal{A}1$ típusú élek, $N2$, és $N3$ típusú tevékenységek lehetnek. A vágásban visszefelé $\mathcal{A}1$, $\mathcal{A}2$ típusú élek és $N1$, $N3$ típusú csomópontok lehetnek.

Az új potenciálok meghatározására képezzük a δ értéket.

A δ értékek meghatározása az alábbi módon történik.

$$\delta := \min\{\delta_{A1}, \delta_{N2}, \delta_{N3}, \delta_{A1^\bullet}, \delta_{A2^\bullet}, \delta_{N1^\bullet}, \delta_{N3^\bullet}\}, \quad \text{ahol}$$

A vágásban levő éleknél

$$\begin{aligned}
\delta_{A1} &:= \min\{\pi_j - \pi_i + \tau_j - \tau_i - z_{ij} : \ (i,j) \in \mathcal{A}1 \quad (i,j) \in (S,T)\} \\
\delta_{N2} &:= \min\{\tau_i - a_i : (i) \in N2 \quad (i_S, i_F) \in (S,T)\} \\
\delta_{N3} &:= \min\{\tau_i - a_i : (i) \in N3 \quad (i_S, i_F) \in (S,T)\}.
\end{aligned}$$

A vágásban visszafelé menő éleknél:

$$\begin{aligned}
\delta_{A1^\bullet} &> 0 \text{ bármekkora értéket felvehet } (i,j) \in \mathcal{A}1 \text{ és } (i,j) \in (T,S) \\
\delta_{A2^\bullet} &> 0 \text{ bármekkora értéket felvehet } (i,j) \in \mathcal{A}2 \text{ és } (i,j) \in (T,S) \\
\delta_{N1^\bullet} &= \min\{b_i - \tau_i : (i_S, i_F) \in N1 \ (i_S, i_F) \in (T,S)\} \\
\delta_{N3^\bullet} &= \min\{b_i - \tau_i : (i_S, i_F) \in N3 \ (i_S, i_F) \in (T,S)\}.
\end{aligned}$$

Az így előállított δ érték határozottan nagyobb mint zérus. Legyen az új π^*_i potenciálrendszer az alábbiak szerint meghatározva:

$$\pi^*_i := \begin{cases} \pi_i & \text{ha } i \in S, \\ \pi_i - \delta & \text{ha } i \in T. \end{cases}$$

Így $p^* = p - \delta$ lesz, egy i tevékenységidő pedig az alábbi képletből határozható meg:

$$\tau_i^* = \pi_{i_F}^* - \pi_{i_S}^* \quad \forall i \in N.$$

A δ érték úgy lett megkonstruálva, hogy a dualitási feltételek az élekre továbbra is teljesüljenek. Az olyan élek ill. éllé alakított tevékenységek, ahol az él mindkét csomópontja vagy az S vagy a T ponthalmazban van, az élek ugyanabban az osztályban maradnak.

A vágásban levő kapcsolatoknál és tevékenységeknél az alábbi a folyamnak megfelelő változás történik.

A δ_{A1} értékkel az $\mathcal{A}1$ él $\mathcal{A}2$ típusúvá válik.

A δ_{N2} értékkel az $N2$ típusú él $N1$ típusú, ennél kisebb értékkel $N3$ típusú éllé válik.

A δ_{N3} értékkel az $N3$ típusú él $N1$ típusúvá válik, ennél kisebb értéknél a saját osztályában marad.

A vágásban visszafelé menő kapcsolatoknál és tevékenységeknél, az alábbi, a folyamnak megfelelő változás történik.

Az $\mathcal{A}1$ típusú él bármely $\delta_{A1^*} > 0$ értékkel változtatva $\mathcal{A}1$ típusú él marad.

Az $\mathcal{A}2$ típusú él bármely $\delta_{A2^*} > 0$ értékkel változtatva $\mathcal{A}1$ típusú éllé válik. Ez azt jelenti, hogy a tevékenységre nézve eddig kritikus kapcsolat nem lesz kritikus. Így δ_{A2^*} csak akkor lehet nagyobb zérusnál, ha a j tevékenységhez továbbra is vezet kritikus út. Mivel csak a kritikus utakon folyik folyam, és a (j,i) kapcsolat úgy került a vágásba, hogy visszafelé $F_i - c_i$ érték folyt rajta, ez feltételezi, hogy a j csomópontba nemcsak i csomóponton keresztül visz kritikus út. Ez az automatizmus biztosítja, hogy a δ értéket tetszőlegesre felvéve, a kapcsolat ill. a tevékenység továbbra is a dualitási feltételeknek megfelelő maradjon.

Az $N1$ típusú csomópontnál a π_{iS} értéket δ_{N1^*} értékkel csökkentve a tevékenység $N2$, ennél kisebb értéknél csökkentve $N3$ típusúvá válik.

Az $N3$ típusú csomópontnál a π_{iS} értéket δ_{N3^*} értékkel csökkentve a tevékenység $N2$, ennél kisebb értékkel csökkentve saját osztályában marad.

Ezek a típus változások mindig megfelelnek a folyamnak, és ezek közül a legkisebbet választva biztosítjuk azt, hogy az összes változás megfelel a dualitási feltételeknek.

Ezzel a második lépést befejeztük.

Ezután az első lépésre visszatérve a folyam ismét növelhető, majd a második lépésben a p átfutási idő csökkenthető. Ezeket a lépéseket addig kell felváltva ismételni, amíg a kívánt átfutási időt el nem érjük, vagy a folyam végtelen nagy nem lesz. Ez esetben ugyanis van olyan $s \to t$ út amely mentén az összes kapacitás végtelen. Ez azt jelenti, hogy az ezen az úton levő élek határozzák meg a tevékenységek kezdetét és végét, ez tehát a leghosszabb, a kritikus út. Ezen az úton a csomópontok az $N1$, az $N4$, vagy az $N2$ osztályban lehetnek. Az $N1$ osztályban levő tevékenység ideje a_i, tovább nem csökkenthető. Az $N4$ osztályban levő tevékenységnél $a_i = b_i$ tehát ez a tevékenység sem gyorsítható. Az $N2$ osztályban

levő csomóponton a tevékenységidő b_i tehát csökkenthető, de ezzel az *átfutási idő növekedne*, hiszen a folyam a tevékenységen visszafelé folyik, tehát az $i_S \in T$. Tehát a kritikus út hossza nem csökkenthető tovább végtelen nagy folyam esetén. Ennek az észrevételnek, hogy az algoritmusnak vége van, ha a folyam végtelen nagy, azért nagy a jelentősége, mert az MPM hálóban a $p_{\min}$ nem számítható ki a rohamidőkből, azaz $p_{\min} \neq p_a$. Előfordul, hogy a két érték megegyezik, de az is hogy nem. Ez utóbbi akkor, ha a végtelen nagyságú folyamnál $N2$ típusú tevékenység van, azaz a tevékenység gyorsítása megnövelné a kritikus út hosszát. Ezt a jelenséget lassítási paradoxonnak nevezik. Ez az algoritmus tehát megadja az MPM hálóban elérhető legrövidebb átfutási időt is, ami eddig szintén megoldatlan probléma volt.

Ezzel a tételt konstruktívan bizonyítottuk.

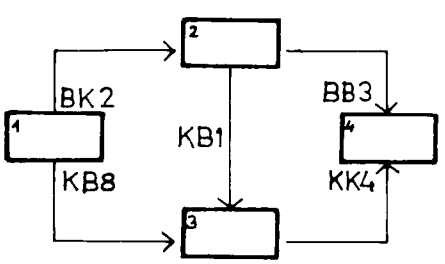
Az alábbiakban a tétel egy fontos következményét ismertetjük.

TÉTEL (erős equilibrium). *Ha létezik optimális primált kielégítő (P) megoldás és optimális duál (D) megoldás, akkor azok értéke egyenlő.*

Bizonyítás. A dualitási tétel értelmében van optimális primál megoldás (P^*) és minimális duál megoldás (D^*) amelyek optimálisak, azaz $(P^*) = (D^*)$. Mivel (P) és (D) is optimálisak, ezért $(P^*) = (P)$ és $(D^*) = (D)$, de ekkor $(P) = (D)$.

3. Mintafeladat az algoritmus illusztrálására

Adott az ábrán látható tervütem háló, a táblázatban megadott normál- és rohamidő adatokkal. Adottak még a tevékenységekhez rendelt költségtényezők (c_i) i. A tevékenységek közti kapcsolatok az éleken vannak feltüntetve. A csomópontokba kisbetűvel írt számok a tevékenység kódját jelölik.



kód	1	2	3	4
b_i	4	5	6	10
a_i	1	3	2	4
c_i	3	3	2	1

Keressük a $p_{\min}$ értékhez tartozó optimális megoldást.

0. lépés A normálidőhöz tartozó átfutási idő számítása (p_b). A feladatot CPM feladattá konvertáljuk, úgy hogy a kapcsolatokat KK kapcsolattá alakítjuk. A π_i értékek a tevékenységek legkorábbi kezdését jelentik. A legkorábbi befejezés $\pi_{iF} = \pi_i + \tau_i$

csomóp.	1	2	3	4
π_i	0	6	2	6

$$p_b = \pi_{tF} = \pi_t + \tau_t = 6 + 10 = 16.$$

Vegyük kiindulásképpen a $\tau_i = b_i$ időértékeket, valamint az ezekből számított π politikát, és a $\varphi_{ij} = 0$ folyamot. Ez a p_b átfutási értékhez tartozó optimális megoldás.

1. lépés (folyam növelés)

tevékenység kód	1	2	3	4
tev. oszt. $N1$–$N4$	2	2	2	2
F_i a kapacitásokhoz	0	0	0	0
kapacitások $\kappa_{i_S i_F}$	3	3	2	1
$\kappa_{i_F i_S}$	∞	∞	∞	∞
F_i a folyamnövelés után	0	0	−1	1

honnan (i)	1	1	2	2	3
élek hova (j)	2	3	3	4	4
élek oszt.	2	2	1	2	1
régi folyam	0	0	0	0	0
kapacitás φ_{ij}	∞	∞	0	∞	0
φ_{ji}	0	0	0	0	0
max. folyam	0	1	0	0	1
növelt folyam	0	1	0	0	1

2. lépés (π_i, τ_i, p csökkentése)

élek a honnan	2	4_S
vágásban hova	4	4_F
δ értékek	2	6
δ min	2	

tevékenység	1	2	3	4
korai kezdés	0	6	2	6
korai befejezés	4	11	8	14
tev. idő $\pi_{i_F} - \pi_{i_S}$	4	5	6	8

Az új átfutási idő $p = 14$.

A költségek az előző átfutási időhöz képest 2 egységgel növekedtek, azaz időegységenként 1 pénzegységgel.

1* lépés (folyam növelés)

tevékenység kód	1	2	3	4
tev. oszt. $N1$–$N4$	2	2	2	3
F_i a kapacitásokhoz	0	0	−1	1
kapacitások $\kappa_{i_S i_F}$	3	3	3	0
$\kappa_{i_F i_S}$	∞	∞	∞	0
F_i a folyamnövelés után	3	3	−1	1

honnan (i)	1	1	2	2	3
élek hova (j)	2	3	3	4	4
élek oszt.	2	2	1	2	2
régi folyam	0	1	0	0	1
kapacitás φ_{ij}	∞	∞	0	∞	∞
φ_{ji}	0	1	0	0	1
max. folyam	3	0	0	3	0
növelt folyam	3	1	0	3	1

2* lépés (π_i, τ_i, p csökkentése)

élek a honnan	1_S 4_S
vágásban hova	1_F 4_F
δ értékek	3 4
δ min	3

tevékenység	1	2	3	4
korai kezdés	0	3	2	6
korai befejezés	1	8	8	11
tev. idő $\pi_{i_F} - \pi_{i_S}$	1	5	6	5

Az új átfutási idő $p = 11$.

A költségek az előző átfutási időhöz képest 12 egységgel növekedtek, azaz időegységenként 4 költségegységgel.

1** lépés (folyam növelés)

tevékenység kód	1	2	3	4	honnan (i) élek hova (j)	1 2	1 3	2 3	2 4	3 4
tev. oszt. $N1$–$N4$	1	2	2	3	élek oszt.	2	2	1	2	2
F_i a kapacitásokhoz	3	3	−1	1	régi folyam	3	1	0	3	1
kapacitások $\kappa_{i_S i_F}$	∞	0	3	0	kapacitás φ_{ij}	∞	∞	0	∞	0
$\kappa_{i_F i_S}$	0	∞	∞	0	φ_{ji}	3	1	0	3	1
					max. folyam	0	0	0	0	0
F_i a folyamnövelés után	3	3	−1	1	növelt folyam	3	1	0	3	1

2** lépés (π_i, τ_i, p csökkentése)

élek a honnan	2_S 4_S
vágásban hova	2_F 4_F
δ értékek	2 1
δ min	1

tevékenység	1	2	3	4
korai kezdés	0	3	2	6
korai befejezés	1	7	8	10
tev. idő $\pi_{i_F} - \pi_{i_S}$	1	4	6	4

Az új átfutási idő $p = 10$.

A költségek az előző átfutási időhöz képest 4 egységgel növekedtek, azaz időegységenként 4 költségegységgel.

1*** lépés (folyam növelés)

tevékenység kód	1	2	3	4	honnan (i) élek hova (j)	1 2	1 3	2 3	2 4	3 4
tev. oszt. $N1$–$N4$	1	3	2	1	élek oszt.	2	2	1	2	2
F_i a kapacitásokhoz	3	3	−1	1	régi folyam	3	1	0	3	1
kapacitások $\kappa_{i_S i_F}$	∞	0	3	∞	kapacitás φ_{ij}	∞	∞	0	∞	0
$\kappa_{i_F i_S}$	0	0	∞	0	φ_{ji}	3	1	0	3	1
					max. folyam	0	∞	0	0	∞
F_i a folyamnövelés után					növelt folyam	3	∞	0	3	∞

Létezik olyan $P(s_S \to t_F)$ a start csomópont kezdetéből a vég csomópont befejezésébe futó út, amely mentén a folyamot végtelen nagy értékkel tudjuk növelni. Ez azt jelenti, hogy az adott tervütem hálón a $p = 10$ időegység átfutási időnél kisebbet nem lehet elérni. Ez az átfutási idő egyébként kisebb mint a rohamidőkből számított átfutási idő, aminek értéke $p_a = 14$ időegység. Ennek ellenőrzését az olvasóra bízzuk. Ezzel a feladatot megoldottuk.

Befejezésképpen megjegyezzük, hogy a hálón elérhető maximális átfutási idő sem egyenlő a normál időkkel számított átfutási idővel. Jelen feladatnál ha a harmadik tevékenységet a rohamidejével, az összes többit a normálidejével vesszük figyelembe, akkor kapjuk a maximális átfutási időt. Ennek értéke $p_{\max} = 20$. Mivel az algoritmus lényege az, hogy egy triviális optimális megoldásból tér át, egy nála kisebb optimális megoldásra, és csak a p_b-hez tartozó optimális megoldást ismerjük, ezért az ennél nagyobb átfutási időhöz tartozó optimális megoldásokat, és a $p_{\max}$ átfutási időt megadni nem tudjuk. Ennek egyébként csak elméleti jelentősége van, hiszen a legkisebb költségű megoldás a normál időkkel számolt időpolitikához tartozik.

IRODALOM

[1] IBM, *Programmbeschreibung für das IBM 1440 Project Control System* (1964).

[2] FONDAHL, J.W., *A Non-Computer Approach to the Critical Path Method for the Construction Industry*, 1st Edition 1961, 2nd Edition 1962 Department of Civil Engineering, Stantford University.

[3] FULKERSON, D.R., „A Network Flow Computation for Project Cost Curves", *Management Sci.* **7** (1961), 167–178.

[4] KELLEY, J.E., „Critical Path Planning and Scheduling: Mathematical Basis", *Op. Res.* 9 (1961), 296–320.

[5] KELLEY, J.E. and WALKER, M.R., „Critical Path Planning and Scheduling", *Proc. of Eastern Joint Computer Conference* (Boston, 1959).

[6] ROY, B., „Théorie des graphes", *Contribution de la théorie des graphes á l'étude de certains problémes linéaires*, Comptes Rendus des Séances de l'Académie des Sciences, séance du Avril 1959, 2437–2439.

(Beérkezett: 1992. október 20.)

HAJDU MIKLÓS
BUDAPESTI MŰSZAKI EGYETEM
ÉPÍTÉSKIVITELEZÉSI TANSZÉK
1111 BUDAPEST, MŰEGYETEM RKP. 3. KII/17.

AN ALGORITHM TO SOLVE THE TIME-COST TRADE-OFFS PROBLEM IN PRECEDENCE DIAGRAMMING

M. HAJDU

In this paper we give a fast, new algorithm for the Time-Cost Trade-Offs problem in precedence diagramming. We allow the next precedence relationships between the activities. Start-to-Start, End-to-Start, End-to-End, End-to-Start. Starting from an optimal solution according p_b project duration (p_b is the project duration calculated from the normal duration of activites) we can reach an optimal solution according to a smaller project duration. The algorithm based on network flows theory.

Alkalmazott Matematikai Lapok **17** (1993) 241–257

ALGORITMUS AZ $F2 \mid$ OVERLAP $\mid C_{\max}$ MEGOLDÁSÁRA

VATTAI ZOLTÁN ANDRÁS

Budapest

A cikk az $F_2 \mid$ overlap $\mid C_{\max}$ feladat megoldására adható algoritmust tárgyal. Az eljárás bemutatásának előkészítése során sor kerül a kvázi „O"-alakú és a szigorúan véve is „O"-alakú ütemterv definíciójára. Újszerű bizonyítást nyer Johnson algoritmusa az $F2||C_{\max}$ feladatra, illetve tételek kimondása és igazolása során mutatjuk meg, hogy a közölt algoritmus egyaránt helyes megoldást ad az $F2||C_{\max}$, $F2 \mid$ idle $\mid C_{\max}$ és $F2 \mid$ overlap, idle $\mid C_{\max}$ feladatokra is.

Kulcsszavak: *termelésirányítás, ütemezés (schheduling), egyutas ütemezési feladat (flow-shop)*

1. Bevezetés

A feladat a következőképpen foglalható össze:

Adva van $m \geq 2$ munkadarab, melynek előállításában ugyanaz az $n = 2$ db gép működik közre, munkadarabonként azonosan adott, rögzített sorrendben (*kétgépes, egyutas ütemezési feladat*). Minden munkadarab megmunkálásában mindkét gép részt vesz, munkadarabonként és gépenként ismert ($T_{i,\ell} > 0$) megmunkálási idővel. A két gép a munkadarabokat azonos sorrendben veszi munka alá (*előzés nem megengedett*). Egy gép egy munkadarab megmunkálását folyamatosan, megszakítás nélkül végzi (*folyamatmegszakítás nem megengedett*). Egy munkadarabon a két gép munkavégzése között időbeli átfedés lehetséges (*átlapolásos feladat*), melynek maximális mértékét munkadarabonként a hálótechnikából kölcsönzött $CR_i \geq 0$ kritikus megközelítéssel [3] adjuk meg. (Az átlapolásos ütemezési feladatok tipikusak az építőiparban, illetve nagy munkaigényű termékek vagy termékszériák előállításánál.) Feladatunk a munkadarabok sorrendjének meghatározása úgy, hogy az előállításukhoz szükséges össz kivitelezési idő ($T = C_{\max}$) a lehető legkisebb legyen. (Rendelkezésre állási-, átállítási- és határidők, valamint költségek, vagy sorrendi megkötések figyelembevétele nem része a feladatnak.) A Graham, Lenstra és társaik által 1981-ben javasolt szimbólumrendszert [1] alkalmazva a feladatot $F2 \mid$ overlap $\mid C_{\max}$ kóddal jelöljük. A tárgyalás szemléletesebbé tételéhez az eredeti feltételrendszerhez még egy megkötést adunk, nevezetesen, hogy a gépek ütemezésében nem lehetnek állásidők. Ez utóbbi alatt azt értjük, hogy ha egy gép bármelyik munkadarabbal megkezdte a munkadarabok megmunkálását, akkor megszakítás nélkül valamennyi munkadarab megmunkálását elvégzi. (Az [1] dolgozat ilyen típusú megkötése nem ad jelölési javaslatot.) A cikk utolsó részében megmutatjuk, hogy ezen megkötésnek két gép esetén semmi hatása nincs a feladatra.

Alkalmazott jelölések:

$T_{i,1}$ és $T_{i,2}$: A gépek tevékenységideje az i munkadarabon (input adat).

CR_i: A két gép időbeli ütemezésénél az átfedés maximalizálásához megadott minimális követési érték az i munkadarabon (input adat). Jelentése: az adott munkadarabon a két gép ütemezése semmilyen készültségi foknál nem kerülhet időben közelebb egymáshoz, mint a megadott érték.

K_i és B_i: A két gép ütemezése közötti minimális idő az i munkadarabon a megmunkálások kezdésekor (0%-os készültségi állapot), illetve befejezésekor (100%-os készültségi állapot). „Kezdési-, illetve befejezési követési idő" (számított segédmennyiségek).

D_e: A két gép ütemezése közötti idő a vizsgált- és az azt megelőző ütemtervrészlet csatlakozásánál.

D_u: A két gép ütemezése közötti idő a vizsgált- és az azt követő ütemtervrészlet csatlakozásánál.

δ_i: Követési idő növekménye az i munkadarabon.

$T = C_{\max}$: Teljes megvalósítási idő (keresett érték).

A megoldás szemléltetésére az ütemezési feladatoknál megszokottól eltérően nem sávos ütemtervet, hanem kétdimenziós ábrákat, ún. ciklogramot (egyfajta progresszió-görbe-rendszert, vagy „készültségi állapot görbe" rendszert [4]) alkalmazunk. Az eljárás során fokozott figyelmet szentelünk a kezdési-, illetve befejezési követési időknek, melyek meghatározására egyszerű algoritmust adunk.

(Lásd: 1. ábra)

$$B_i = \max\{CR_i,\ CR_i + T_{i,2} - T_{i,1}\}$$
$$K_i = \max\{CR_i,\ CR_i + T_{i,1} - T_{i,2}\}$$

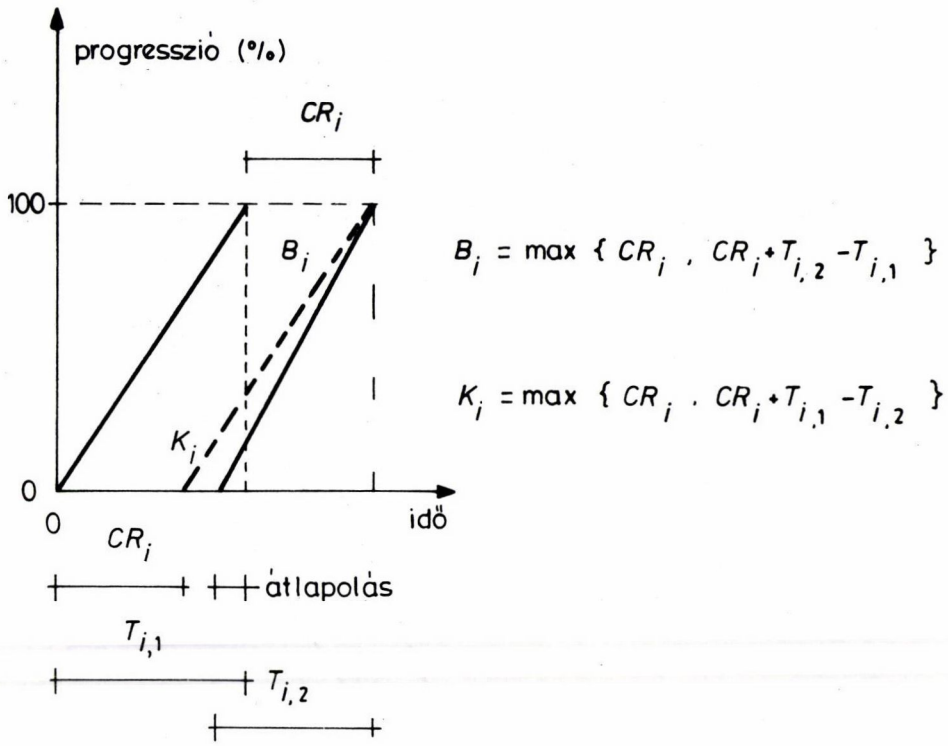
2. Kvázi „O" alakú ütemterv

2.1 Definíció. Kvázi „O"-alakú az ütemterv, ha nem található benne olyan munkadarabpár, melyre

$$B_i < K_i \quad \text{és} \quad B_j > K_j \qquad i < j \tag{2.1}$$

2.1 LEMMA. *Létezik olyan optimális ütemterv, mely kvázi „O"-alakú!*

Bizonyítás. Mivel a gépek számára folyamatos munkavégzést írtunk elő, és mivel az egyes gépeken az összes munkadarab megmunkálásához szükséges idő — bármely sorrendben történjék a munkadarabok megmunkálása — konstans, a teljes kivitelezési idő — mint célfüggvény — helyett elegendő a két gép ütemezése közötti időt vizsgálnunk a legelső munkadarab megmunkálásainak kezdetekor, illetve



1. ábra
Ciklogram a követési idők értelmezésével és meghatározásával

a legutolsó munkadarab megmunkálásainak befejezésekor (lásd még 3.1 Tétel- és Bizonyítás, valamint 10. ábra). Ugyancsak a megmunkálások folyamatosságából adódóan, ha egy kialakított ütemterv belsejében egy ütemtervrészt módosítunk, akkor a módosításnak az esetleges célértéket növelő hatásai a kapcsolódó megelőző- és követő ütemtervrészekre (azok valamennyi rész-ütemtervére — így a legelsőnek, illetve a legutólsónak választott munkadarabok megmunkálására is) teljes terjedelmükkel tovább adódnak. (Ugyanez az esetleges célértéket csökkentő hatásokról már nem állítható ilyen bizonyosan, hiszen például egy már optimális megoldást jelentő sorrend esetében a módosítás lokálisan hiába mutatna célértéket csökkentő hatást, a célérték az optimalitásból adódóan tovább már nem csökkenthető.)

Fentiek alapján a bizonyítás során feltételezünk egy létező optimális ütemtervet (a célfüggvény értékének ismerete nem szükséges), melyet alkalmasan megválasztott ütemterv-részletek módosításával (felcserélésével) megpróbálunk „elrontani", miközben figyeljük a kapcsolódó ütemtervrészek változásait ...

Tételezzük fel, hogy találtunk egy optimális ütemtervet, mely nem kvázi „O"-

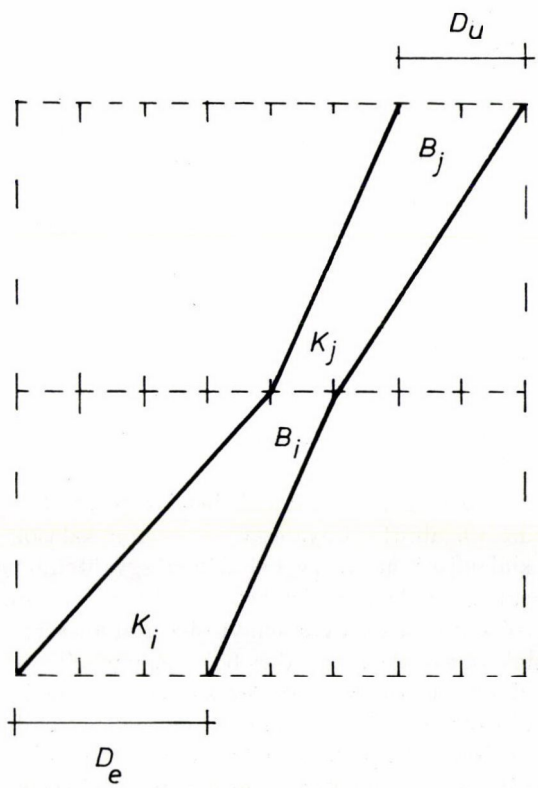
alakú. Ekkor bizonyosan található benne legalább egy olyan i, j munkadarabpár, melyre 2.1 teljesül. Több ilyen esetén tekintsük azt, melyben a munkadarabok legközelebb esnek egymáshoz ($j - i = \min$).

Ebben az esetben i és j között csak olyan munkadarab lehet, melynél $B_\ell = K_\ell$ ($i < \ell < j$), aminek hatásától, — mint azt a bizonyítás logikájából látni fogjuk — eltekinthetünk. Így praktikusan feltételezzük, hogy $j = i + 1$.

A bizonyítás során két alapesetet kell megkülönböztetnünk annak megfelelően, hogy a vizsgált munkadarabok ütemezését megelőző, illetve követő ütemtervrészek csatlakozásánál a követő idők (D_e és D_u) hogyan viszonyulnak egymáshoz:

I. A követő csatlakozó ütemtervrésznél a követési idő kisebb vagy egyenlő a megelőző ütemtervrész csatlakozásánál lévő követési idővel ($D_u \leq D_e$) :

(Lásd: 2. ábra)



2. ábra

i és j munkadarab ütemezése egy nem kvázi „O"-alakú ütemtervben (I. eset)

Felírható összefüggések:

$$B_i < K_i$$

$$B_j > K_j$$

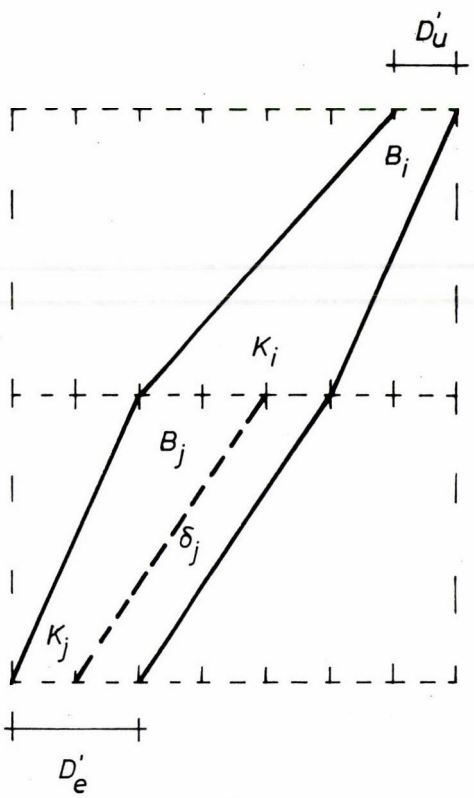
$$B_i = K_j$$

$$B_j \leq K_i$$

$$D_u = B_j$$

$$D_e = K_i$$

Változtassuk meg az adott optimális ütemtervet a két munkadarab felcserélésével! (Lásd: 3. ábra)



3. ábra

Kvázi „O"-alakú ütemterv előállítása (I. eset)

Felírható összefüggések:

$$B_i < K_i$$

$$B_j > K_j$$

$$B_i = K_j$$

$$\delta_j = K_i - B_j$$

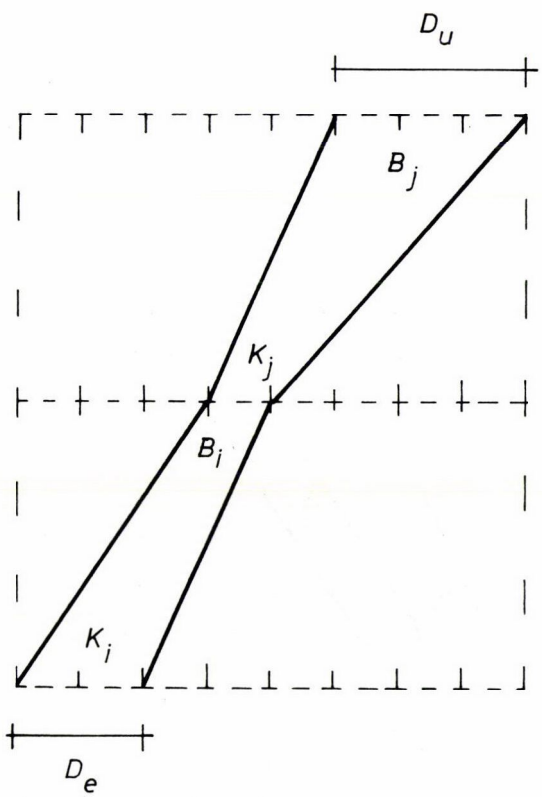
Ezekből következik:

$$D'_u = B_i = K_j < B_j = D_u \quad \text{illetve}$$

$$D'_e = K_j + \delta_j = K_j + K_i - B_j < K_i = D_e$$

II. A következő csatlakozó ütemtervrésznél a követési idő nagyobb a megelőző ütemtervrész csatlakozásánál lévő követési időnél ($D_u > D_e$) :

(Lásd: 4. ábra)



4. ábra

i és *j* munkadarab ütemezése egy nem kvázi „*O*"-alakú ütemtervben (II. eset)

Felírható összefüggések:

$$B_i < K_i$$

$$B_j > K_j$$

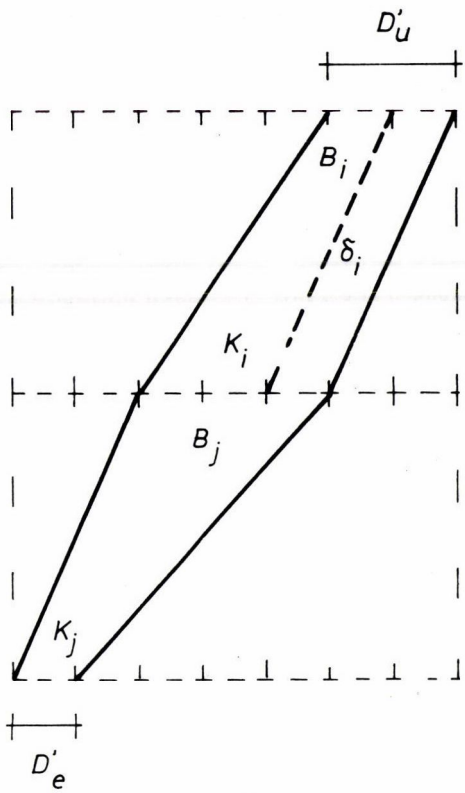
$$B_i = K_j$$

$$B_j > K_i$$

$$D_u = B_j$$

$$D_e = K_i$$

Változtassuk meg az adott optimális ütemtervet a két munkadarab felcserélésével! (Lásd: 5. ábra)



5. ábra
Kvázi „O"-alakú ütemterv előállítása (II. eset)

Felírható összefüggések:

$$B_i < K_i$$

$$B_j > K_j$$

$$B_i = K_j$$

$$\delta_j = B_j - K_i$$

Ezekből következik:

$$D'_e = K_j = B_i < K_i = D_e \quad \text{illetve}$$
$$D'_u = B_i + \delta_i = B_i + B_j - K_i < B_j = D_u$$

Látható tehát, hogy a cserével a célfüggvény értéke egyik esetben sem nőtt. Az adott módon bármilyen ütemtervből kiindulva elő tudunk állítani kvázi „O"-alakú ütemtervet.

A bizonyítás során nem volt szükség annak vizsgálatára, hogy az adott párban szereplő munkadarabok ütemtervei eredeti állapotukhoz képest szét voltak-e húzva, vagy sem.

□

3. Szigorúan véve is „O"-alakú ütemterv

3.1 Definíció. Legyen g egy kvázi „O"-alakú optimális ütemtervben az utolsó olyan munkadarab indexe, melyre $K_i < B_i$, valamint legyen h ugyanebben az ütemtervben az első olyan munkadarab inedexe, melyre $K_i > B_i$. (A kvázi „O"-alakú ütemterv definíciójából adóadóan $g < h$.)

Szigorúan véve is „O"-alakú a kvázi „O"-alakú ütemterv, ha nem található benne olyan munkadarabpár, melyre

$$K_i > K_j \quad | \quad i < j \leq g \tag{3.1}$$

illetve

$$B_i < B_j \quad | \quad h \leq i < j \tag{3.2}$$

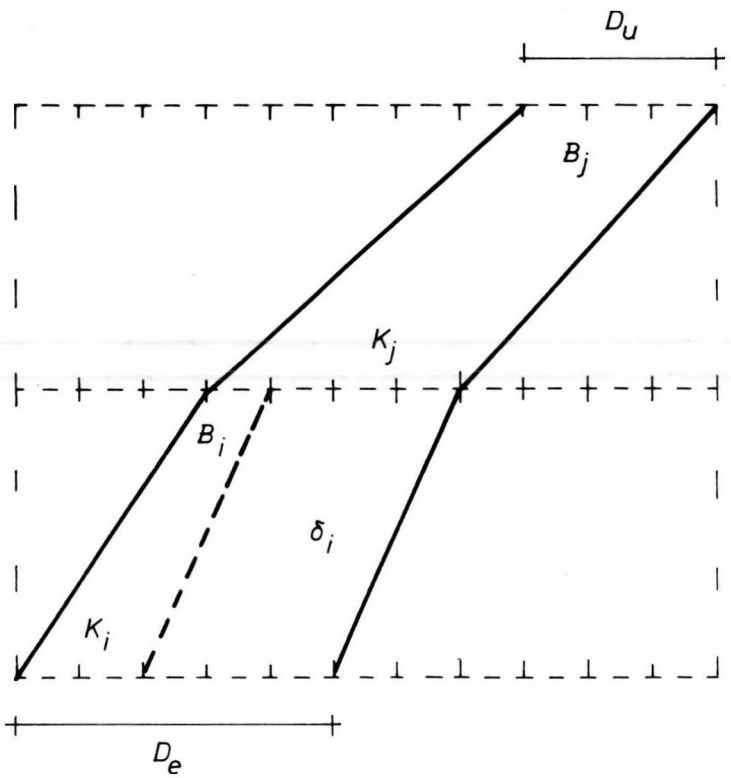
3.1 LEMMA. *Létezik olyan optimális ütemterv, mely szigorúan véve is „O"-alakú.*

Bizonyítás. Tételezzük fel, hogy találtunk olyan kvázi „O"-alakú optimális ütemtervet, mely szigorúan véve nem „O"-alakú. Ekkor bizonyosan található benne olyan szomszédos munkadarabpár, melyre vagy 3.1, vagy 3.2 teljesül. Tételezzük fel az utóbbit. Több ilyen pár esetén tekintsük azt, amelyben a munkadarabok legközelebb esnek egymáshoz. ($j - i = \min$).

Ebben az esetben i és j között csak olyan munkadarab lehet, melynél $B_\ell = K_\ell$ $(i<\ell<j)$, aminek hatásától — mint azt a bizonyítás logikájából látni fogjuk — eltekinthetünk. Így praktikusan feltételezzük, hogy $j = i + 1$.

A bizonyítás során két esetet kell megvizsgálnunk aszerint, hogy a csatlakozó ütemtervrészeknél a minimális követési idők (B_j és K_i) hogyan viszonyulnak egymáshoz:

I. A követő csatlakozó ütemtervrésznél a minimális követési idő nagyobb, vagy egyenlő a megelőző csatlakozó ütemtervrésznél lévő minimális követési idővel a j, illetve az i munkadarab ütemezésénél ($B_j \geq K_i$): (Lásd: 6. ábra)



6. ábra

i és j munkadarab ütemezése egy kvázi „O"-alakú, de szigorúan véve nem „O"-alakú ütemtervben (I. eset)

Felírható összefüggések:

$$B_i \leq K_i$$

$$B_j \leq K_j$$

$$B_i < B_j$$

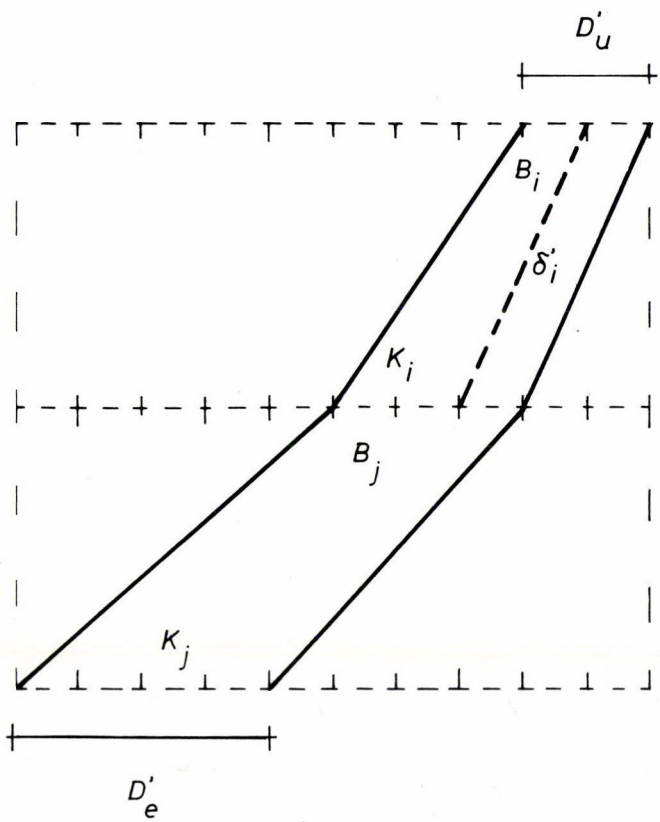
$$B_j \geq K_i$$

$$\delta_i = K_j - B_i$$

$$D_u = B_j$$

$$D_e = K_i + \delta_i$$

Változtassuk meg az adott optimális ütemtervet a két munkadarab felcserélésével! (Lásd: 7. ábra)



7. ábra

Szigorúan véve is „O"-alakú ütemterv előállítása (I. eset)

Felírható összefüggések:

$$B_i \leq K_i$$

$$B_j \leq K_j$$

$$B_i < B_j$$

$$B_j \geq K_i$$

$$\delta_i' = B_j - K_i$$

Fentiekből következően:

$$D_u' = B_i + \delta_i' = B_i + B_j - K_i \leq B_j = D_u \quad \text{illetve}$$
$$D_e' = K_j = B_i + \delta_i = B_i + K_j - B_i \leq K_i + K_j - B_i = D_e$$

II. A követő csatlakozó ütemtervrésznél a minimális követési idő kisebb a megelőző csatlakozó ütemtervrésznél lévő minimális követési időnél a j, illetve az i munkadarab ütemezésénél ($B_j < K_i$): (Lásd: 8. ábra)

Felírható összefüggések:

$$B_i \leq K_i$$

$$B_j \leq K_j$$

$$B_i < B_j$$

$$B_j < K_i$$

$$\delta_i = K_j - B_i$$

$$D_u = B_j$$

$$D_e = K_i + \delta_i$$

Majd csere után ... (Lásd: 9. ábra)

Felírható összefüggések:

$$B_i \leq K_i$$

$$B_j \leq K_j$$

$$B_i < B_j$$

$$B_j < K_i$$

$$\delta_j = K_i - B_j$$

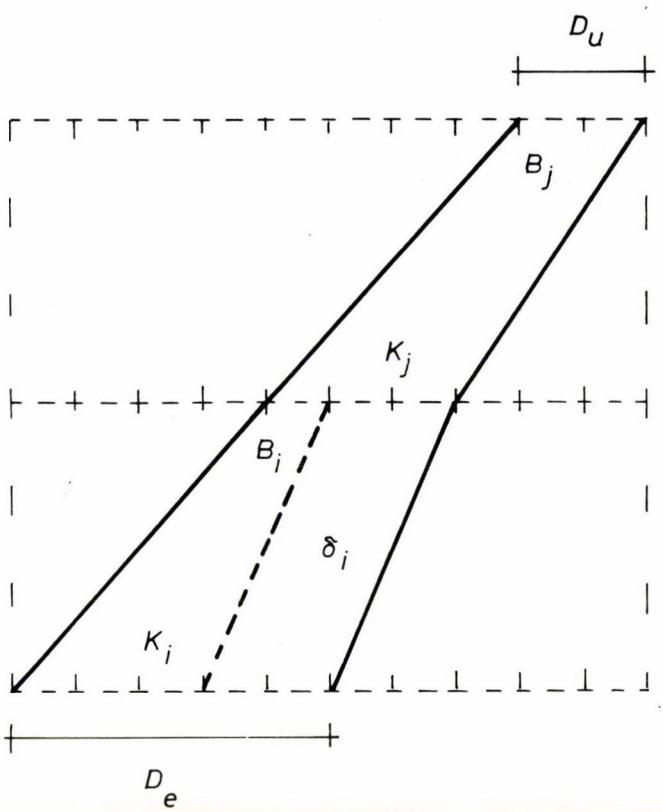
Fentiekből következően:

$$D_u' = B_i < B_j = D_u \quad \text{illetve}$$
$$D_e' = K_j + \delta_j = K_j + K_i - B_j < K_i + K_j - B_i = D_e$$

Látható tehát, hogy a cserével a célfüggvény értéke egyik esetben sem nőtt.

A bizonyítás során nem volt szükség annak vizsgálatára, hogy j munkadarab ütemterve eredeti állapotához képest szét volt-e húzva, vagy sem.

Hasonló módszer alkalmazható bizonyításra 3.1 teljesülése esetén is. Az adott módon bármilyen kvázi „O"-alakú ütemtervből kiindulva, szigorúan véve is „O"-alakú ütemterv állítható elő. □



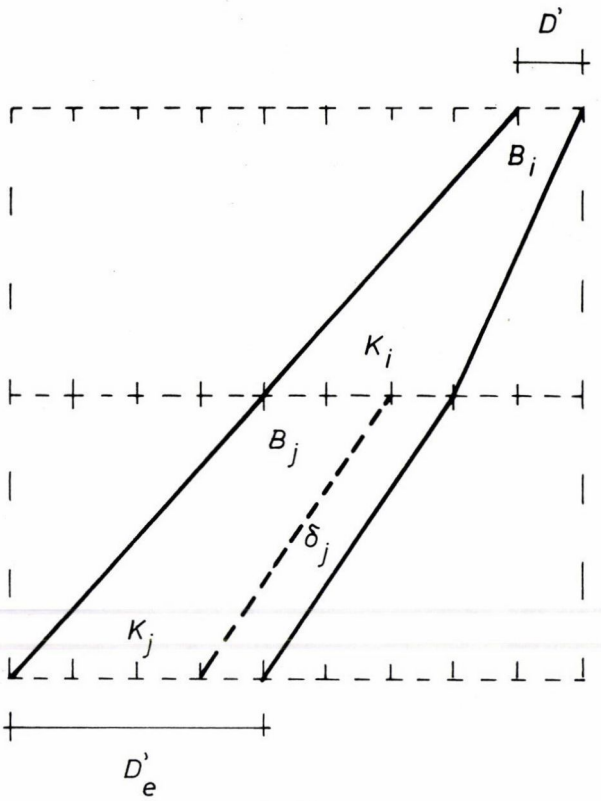
8. ábra
i és *j* munkadarab ütemezése egy kvázi „*O*"-alakú,
de szigorúan véve nem „*O*"-alakú ütemtervben (II. eset)

3.1 TÉTEL. *Ha egy ütemtervről belátható, hogy szigorúan véve is „O"-alakú, akkor bizonyos, hogy optimális is.*

Bizonyítás. Bármely ütemterv teljes átfutási ideje (T) két jól elkülöníthető részre bontható fel:

- A megmunkálási idők összege az első gépen (T'_u);
- Követési idő a két gép munkavégzése között az utolsó munkadarabon a megmunkálások befejezésekor (T''_u),

vagy



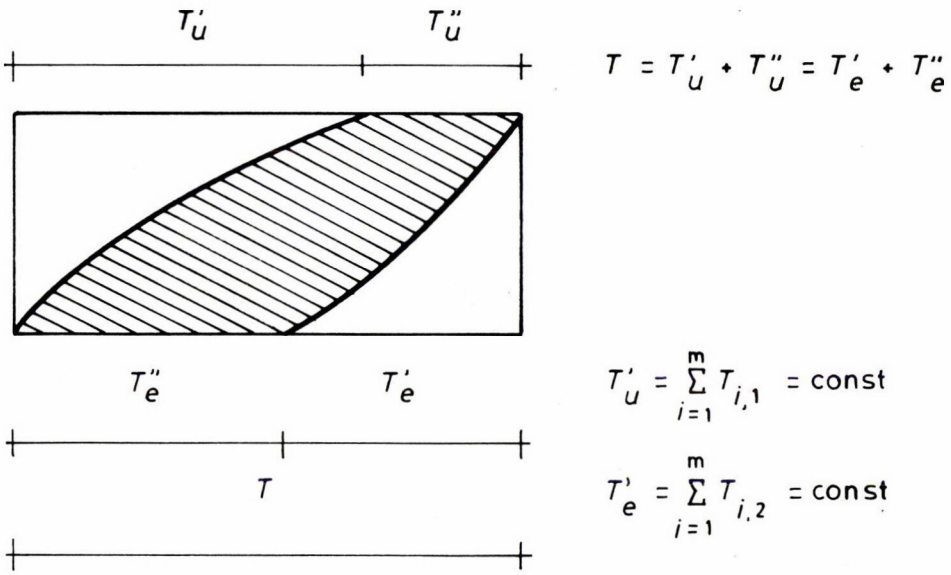
9. ábra

Szigorúan véve is „O"-alakú ütemterv előállítása (II. eset)

- A megmunkálási idők összege a második gépen (T'_e);
- Követési idő a két gép munkavégzése között az első munkadarabon a megmunkálások kezdetekor (T''_e).

Tekintve, hogy $\sum_{i=1}^{m} T_{i,1} = T'_u$ és $\sum_{i=1}^{m} T_{i,2} = T'_e$ sorrendtől független konstansok, $\min(C_{\max}) = \min(T)$ akkor van, ha $T''_e = T''_{e\min}$, vagy $T''_u = T''_{u\min}$. Márpedig ez szigorúan véve is „O"-alakú ütemtervnél — definícióból adódóan is — teljesül.

(Lásd: 10. ábra) □



10. ábra
A célfüggvény változó része a követési idő

4. Megoldó algoritmus

4.1 Algoritmus. Szigorúan véve is „O"-alakú ütemterv előállítására szolgáljon az alábbi algoritmus:

> CR_i-k növekvő sorrendjében döntsünk afelől, hogy adott munka a leendő sorrendben az első-, illetve utolsó pozíciótól befelé, a sorrend elejére (a már sorolt elsők mögé), vagy végére (a már sorolt utolsók elé) kerüljön, annak megfelelően, hogy $CR_i = K_i$, vagy $CR_i = B_i$. Amennyiben $CR_i = K_i = B_i$, úgy tetszőlegesen dönthetünk.

4.1 TÉTEL. *Fenti algoritmus optimumot szolgáltat nemcsak* $F2|\text{overlap}|C_{\max}$, *de* $F_2 \,||\, C_{max}$ *feladatnál is. Bizonyítást lásd a következő tételnél!*

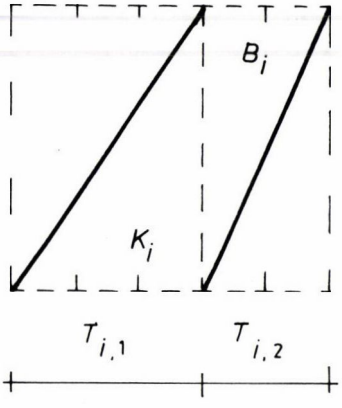
4.2 Algoritmus. Az $F2 \,||\, C_{\max}$ feladat megoldására már 1945-ben S. M. Johnson közölt algoritmust [2], miszerint:

> A megmunkálási idők növekvő sorrendjében döntsünk afelől, hogy egy adott munka a leendő sorrendben az első-, illetve utolsó pozíciótól befelé

a sor elejére (a már sorolt elsők mögé), vagy végére (a már sorolt utolsók elé) kerüljön, annak megfelelően, hogy az adott megmunkálási idő az első, vagy a második gépen jelentkezik-e.
Amennyiben az adott munkán mindkét gép azonos ideig dolgozik, úgy tetszőlegesen dönthetünk.

4.2 TÉTEL. *4.2 algoritmus 4.1 speciális esete.*

Bizonyítás. $F2 \mid\mid C_{\max}$ feladatnál is értelmezhetők — és a megmunkálási idők alapján egyszerűen meghatározhatók — a követési idők. Ezek ismeretében a feladat azonos $F2 \mid$ overlap $\mid C_{\max}$-szal. — A tevékenységidők pedig mint követési idők jelennek meg. (Lásd: 11. ábra)



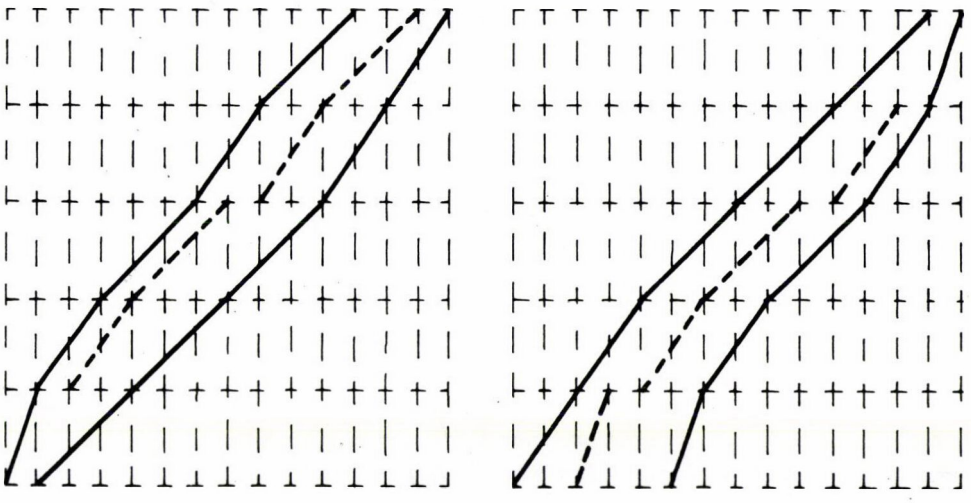
11. ábra
A követési idők az átlapolás nélküli feladatban is értelmezhetők

$$CR_i = \min\{T_{i,1}, T_{i,2}\}$$
$$B_i = \max\{CR_i, CR_i + T_{i,2} - T_{i,1}\} = T_{i,2}$$
$$K_i = \max\{CR_i, CR_i + T_{i,1} - T_{i,2}\} = T_{i,1}$$

Vegyük észre, hogy ezennel Johnson algoritmusának helyességét is bizonyítottuk, hiszen az általa szolgáltatott sorrend szintén szigorúan véve is „O"-alakú ütemtervet ad. □

4.3 TÉTEL. *$F2 \mid\mid C_{\max}$ feladat optimuma $F2 \mid$ idle $\mid C_{\max}$ feladatnak is, illetve $F2 \mid$ overlap $\mid C_{\max}$ feladat optimuma egyben optimuma $F2 \mid$ overlap, idle $\mid C_{\max}$ feladatnak is.*

Bizonyítás. $F2 \mid$ idle $\mid C_{\max}$, illetve $F2 \mid$ overlap, idle $\mid C_{\max}$ feladatok optimális megoldásaiban elképzelhető, hogy az első, vagy a második gép ütemezésében szakadás van. Ekkor az első gépen a kezdés, a második gépen pedig a befejezés irányában elmozdítva a megmunkálásokat az állásidők a célérték zavarása nélkül felszámolhatók. (Lásd: 12. ábra) □



12. ábra
A várakozási idők megengedése nincs hatással a feladatra

Megjegyzés. Az egyes gépek ütemezésének megszakítására a gyakorlatban pl. a két gép közötti korlátozott raktárkapacitás, illetve „sérülékeny állapotok" miatt kerülhet sor.

IRODALOM

[1] „Deterministic and stochastic scheduling", *Proceedings of the NATO Advanced Study and Research Institute on Theoretical approaches to scheduling problems held in Durham* (M. A. H. Dempster, J. K. Lenstra, A. H. G. Rinnooy Kan, eds.) (England, 1981).

[2] JOHNSON, S. M., „Optimal Two- and Three-Stage Production Schedules with Setup Times Included", *Naval Research Logistics Quaterly* 1 (1954), 61–68.
[3] MODER, J. J., PHILLIPS, C. R. and DAVIS, E. W., *Project Management with CPM PERT and Precedence Diagramming* (Van Norstrand Reinhold, New York, 1983).
[4] PILCHER, ROY, *Principles of Construction Management* (McGraw-Hill, Berkshire, England, 1976).

(Beérkezett: 1993. április 15.)

VATTAI ZOLTÁN ANDRÁS
BUDAPESTI MÜSZAKI EGYETEM
ÉPÍTÉSKIVITELEZÉSI TANSZÉK
1111 BP., MÜEGYETEM RKP 3.
TEL/FAX: 1813-377

ALGORITHM FOR SOLVING $F2 \mid$ OVERLAP $\mid C_{\max}$ PROBLEM

Z. A. VATTAI

A new approach algorithm for solving $F2 \mid$ overlap $\mid C_{\max}$ problem is discussed. While preparing for description of the method definition of "quasi O shaped schedule" and that of "absolute O shaped schedule" are settled. A demostrative proof for Johnson's algorithm for $F2||C_{\max}$ problem is also discussed. Via theorems and their proofs it got be shown that the new algorithm gives solutions for problems $F2||C_{\max}$, $F2 \mid$ idle $\mid C_{\max}$ and $F2 \mid$ overlap, idle $\mid C_{\max}$ too.

Alkalmazott Matematikai Lapok 17 (1993) 259–266

SZAKASZONKÉNT FOLYTONOSAN DIFFERENCIÁLHATÓ ENERGIA FÜGGVÉNNYEL JELLEMEZHETŐ TARTÓSZERKEZETEK STABILITÁSVIZSGÁLATA

CSÉBFALVI A.

Pécs

A dolgozatban egy olyan útkövető módszert mutatunk be, amelyet véges szabadságfokú, konzervatív erőrendszerrel terhelt szerkezetek stabilitásvizsgálatára dolgoztunk ki. Megmutatjuk, hogy az egydimenziós erő-elmozdulás poligonokkal jellemzett modellek töréspontjai a folytonos erő-elmozdulás függvényű modellek elágazási pontjaival analóg módon kezelhetők. A vizsgált módszer útkövető jellege miatt elkerülhető a matematikai eszköztár kiterjesztése, a szubgradiens fogalmának bevezetése. Direkt módszerek esetében a szubgradiens fogalmának bevezetése, a feladat átdefiniálása elkerülhetetlen.

1. Bevezetés

A mérnöki szerkezetek stabilitásvesztésének két alapvető oka lehet: (i) az anyagtulajdonságok megváltozásából, vagy (ii) a szerkezeti geometria megváltozásából fakadó tönkremenetel. Az esetek zömében a két jelenség nem független egymástól, leírásuk csak nemlineáris összefüggésekkel lehetséges. Az anyagi, illetve geometriai nemlinearitás egyaránt vezethet töréspontokat tartalmazó energia függvényre.

A stabilitásvizsgálati módszerek egyik irányzatát az útkövető módszerek alkotják, amelyek elsődleges célja a szerkezet állapotváltozási görbéjének meghatározása. A stabilitásvizsgálatok másik nagy csoportját a direkt módszerek képviselik, melyek alapvetően az állapotváltozási görbe kritikus pontjainak meghatározására szolgálnak. Az útkövető módszerek nem képesek pontos információt szolgáltatni a kritikus pontokkal kapcsolatban, mivel alapgondolatukat tekintve növekményi típusúak. A direkt módszerek viszont erősen függnek a kezdeti megoldás megválasztásától. Nemsima energia fügvénnyel jellemzett tartószerkezetek vizsgálatakor, reverzibilis esetben, az egyensúly feltételét a szubdifferenciál, az egyensúly stabilitását pedig a második szubdifferenciál minősíti. Mivel a potenciális energia függvény tartalmazza a poligonális anyagtörvényt, az energia függvény Jacobi-, illetve Hesse-mátrixa intervallum mátrix lesz.

Nem véletlen tehát, hogy a gyakorlati stabilitásvizsgálatok jelentős része vegyes eljárás, amelyeknek az a célja, hogy semlegesítse a két irányzat kedvezőtlen hatásait (KAMAT, WATSON, és VENKAYYA (1983), valamint KAMAT és WATSON (1984)).

Jelenleg a leggyakrabban alkalmazott nemrugalmas stabilitásvizsgálati módszerek az ún. kvázi-rugalmas növekményi módszerek. Energetikai elven alapuló

végeselemes módszert alkalmaz többek között KONDOH és ATLURI (1985) síkbeli rácsos tartók vizsgálatára. WRIGGERS, WAGNER és MIEHE (1988) egy ún. kibővített egyenletrendszeren alapuló kvadratikusan konvergens eljárást alkalmaz az elágazási pontok meghatározására, de nem vizsgálja a rúdelemek kihajlásának hatását. Módszerük a direkt és útkövető módszerek összekapcsolásával született ún. hibrid módszer, amely magán viseli a két módszer sajátosságaiból adódó hibákat és előnyöket. A módszertani váltások miatt az eljárás részben íveket, részben pontokat határoz meg. Összetett feladatok esetén, amikor szűk intervallumon belül több elágazási pont megjelenésével kell számolnunk, a módszer alkalmazása információvesztést eredményezhet.

2. A stabilitási feladat megfogalmazása

A vizsgált szerkezeti modell egy olyan térbeli rácsos tartó, amely ideális csuklókkal összekapcsolt — nemlineáris erő-elmozdulás függvényekkel jellemzett — egydimenziós rúdelemekkel modellezhető. Az összefüggéseket a megváltozott tartóalakra írtuk fel, tetszőlegesen nagy elmozdulásokat feltételezve. A térbeli rácsos tartót alkotó rúdelemek — geometriailag tökéletes rúdelemet feltételezve — explicit erő-elmozdulás poligonokkal jellemezhetők.

Feltételezzük, hogy a teljes szerkezetre felírható egy szakaszonként folytonos, szakaszonként legalább kétszer differenciálható potenciális energia függvény. Az energia függvény meghatározásakor az alábbi megkötésekkel éltünk:

- Vizsgálatainkat véges szabadságfokú, konzervatív erőrendszerrel terhelt rácsos szerkezetekre korlátozzuk;
- Az erő jellegű terhek statikusak és a rácsostartó csomópontjain hatnak;
- Feltételezzük továbbá, hogy a csomópontokra felírt általánosított tehervektor egy P_i^0 állandó alapteher és egy λ paraméter szorzataként, $P_i = P_i^0 \lambda$ formában adható meg;
- A terheletlen szerkezet ($\lambda = 0$) stabil egyensúlyi állapotban van.

Az energia függvények meghatározásakor feltételezzük, hogy a rúdelemek:

- anyaga homogén, izotróp, lineárisan rugalmas,
- legalább egy szimmetriatengellyel rendelkező állandó keresztmetszetűek;
- a keresztmetszetek alaktartóak;
- a sík keresztmetszetek síkok maradnak és a deformált rúdon is merőlegesek a rúdtengelyre, valamint
- a rúd tengelyére merőleges feszültségek hatása elhanyagolható;
- a kihajlott rúdelem alakváltozásai kicsinyek;
- az adott irányú fix megtámasztásokat „végtelen nagy" normálmerevségű rudakkal, a rugalmas megtámasztásokat pedig a rugóállandónak megfelelő normálmerevségű rudakkal helyettesítjük.

Független változónak a csomóponti u_i eltolódásokat tekintjük. A szerkezet

anyagjellemzői és geometriai adatai konstansok. A szerkezet teljes potenciális energia függvénye ílymódon végesszámú változó függvényeként írható fel:

$$V = V(u_i, \lambda), \quad i = 1, 2, \ldots 3n, \tag{2.1}$$

ahol V a teljes potenciális energia függvényt, u_i a csomóponti elmozdulást, λ pedig a teherparamétert jelöli. Az elmozduló, belső csomópontok száma n. A térbeli szerkezet szabadságfoka $3n$.

Mivel a szerkezetet a csomópontokon átadódó konzervatív erőrendszer terheli, a teljes potenciális energia függvény felírható az alábbi formában:

$$V(u_i, \lambda) \equiv U(u_i) + \lambda \sum_i P_i^0 u_i, \tag{2.2}$$

ahol $U(u_i)$ a szerkezetet alkotó rúdelemek alakváltozási energiája.

A potenciális energia stacionaritási elvét alkalmazva, a (2.1), illetve (2.2) alapján meghatározom a háromdimenziós szerkezet pre-, és posztkritikus mozgásállapotát leíró nem növekményi típusú, geometriailag nemlineáris modell egyensúlyi egyenletrendszerét:

$$V_{,i} \equiv \frac{\partial}{\partial u_i} V(u_i, \lambda) = 0, \tag{2.3}$$

illetve

$$V_{,i} \equiv \frac{\partial U}{\partial u_i} + \lambda P_i^0 = 0. \tag{2.4}$$

A szerkezet állapotváltozási egyenletrendszerét a rúdelemek alakváltozási energia függvénye alapján az alábbi formában írhatjuk:

$$G_{ij} S_j + \lambda P_i^0 = 0, \quad i = 1, 2, \ldots, n; \; j = 1, 2, \ldots, m, \tag{2.5}$$

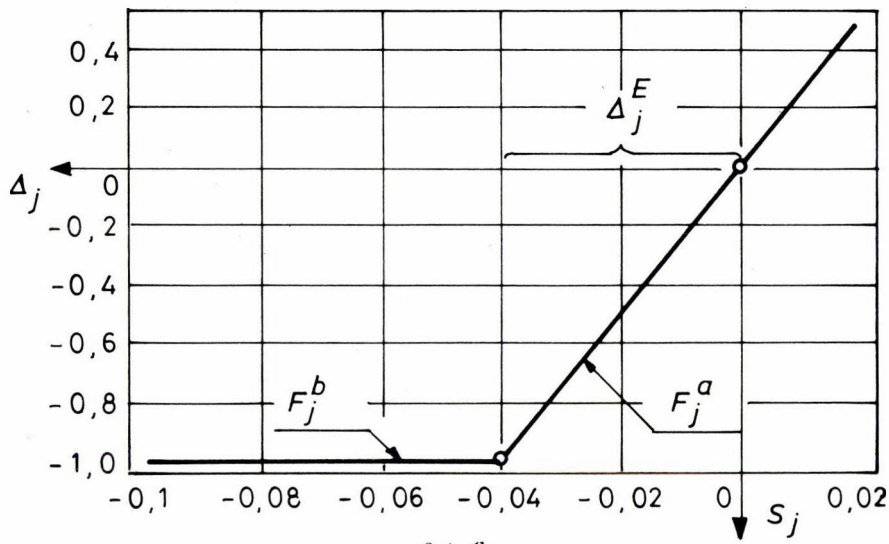
ahol G_{ij} a rácsos tartó geometriai mátrixa, amelyet a csomóponti eltolódások függvényeként a megváltozott tartóalakra írunk fel. S_j a j–edik rúdelemre vonatkozó rúderő, amely a rúdelem összenyomódása, illetve megnyúlása függvényében az alábbi összefüggésekkel adhatunk meg:

$$S_j = \begin{cases} F_j^a(\Delta_j(u_i)), & \text{ha} \quad \Delta(u_i) \geq \Delta_j^E \\ F_j^b(\Delta_j(u_i)), & \text{ha} \quad \Delta_j(u_i) \leq \Delta_j^E \end{cases}, \tag{2.6}$$

ahol

$$F_j^a(\Delta_j(u_i)) = k_j^a \Delta_j(u_i), \tag{2.7}$$

$$F_j^b(\Delta_j(u_i)) = S_j^E \left[1 - \frac{1}{2}\left(\Delta_j(u_i) - \Delta_j^E\right)\right]. \tag{2.8}$$



2.1 ábra

A (2.6)-(2.8) kifejezésekben k_j^a a rúdelem nyúlási merevsége, S_j^E a rúdelemre vonatkozó Euler-erő, Δ_j^E pedig az Euler-erő okozta összenyomódás a j-edik rúdelemben, amelyek a stabilitási feladatban konstansnak tekinthetők. A (2.7) és (2.8) kifejezések ílymódon csupán Δ_j függvényei lesznek, melyet egy adott rúdelemre a 2.1 ábrán szemléltetünk. A Δ_j alakváltozás a csomóponti eltolódások explicit függvénye, ezáltal a (2.5) állapotváltozási egyenlet felírható a csomóponti eltolódások függvényeként.

Mivel a (2.5) állapotváltozási egyenlet tartalmazza a törésponttal rendelkező (2.6) kifejezést, az állapotváltozási egyenlet nem lesz folytonosan differenciálható.

3. Az állapotváltozási görbe meghatározása

Az állapotváltozási görbét az egyensúlyi egyenletek alapján határozzuk meg, amelyet a (2.1) potenciális energia függvény stacionaritási elve alapján a (2.3), (2.4), illetve (2.5) kifejezésekkel írhatunk fel.

Feltételezzük, hogy tehermentes, deformálatlan $(u_i, \lambda) = (0, 0)$ állapotban a szerkezet stabil egyensúlyi állapotban van, illetve, hogy a λ teherintenzitási paraméter növelésével kezdetben stabil egyensúlyi úton halad.

Jelölje y_k, $k = 1, 2, \ldots, n+1$ a csomóponti eltolódások és a teherintenzitási paraméter összekapcsolásával adódó vektort, ahol $y_i = u_i$, $i = 1, 2, \ldots n$; és $y_{n+1} = \lambda$.

Induljunk ki a rendszer egy ismert y_k^a stabil egyensúlyi pontjából, például az $y_k^a = (0, 0)$ pontból. Mivel y_k^a egyensúlyi pont, ezért kielégíti a $V_{,i}\big|^a = 0$ egyensúlyi

egyenletet. Ebben az esetben $V_{,ij}$ $(i = j = 1, 2, \ldots, n)$ Jacobi-mátrix pozitív definit, invertálható, így y_k^a pont környezetében a megoldás egyértelmű.

Írjuk fel az egyensúlyi utat az y_k^a pont környezetében egy alkalmasan megválasztott s paraméter függvényében. Feltevésünknek megfelelően y_k^a elegendően kicsiny környezetében az egyensúlyi út egy folytonos görbe, amely a következő alakban állítható elő:

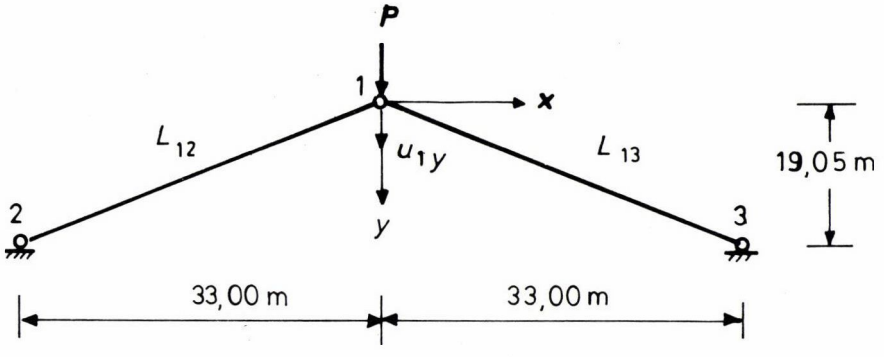
$$y_k(s) = y_k^a + y_k^{a(1)} s + \frac{1}{2!} y_k^{a(2)} s^2 + \frac{1}{3!} y_k^{a(3)} s^3 + \ldots , \tag{3.1}$$

ahol $y_k^{a(1)}, y_k^{a(2)}, \ldots$, pedig az $y_k(s)$ függvény s szerinti deriváltjait jelöli az y_k^a pontban.

Mivel az útkövető módszer egyes lépései azonosak a CSÉBFALVI (1993 a), illetve (1993 b) cikkekben ismertetett lépésekkel, ezért ezek ismertetésétől eltekintünk. Az alábbiakban csak azokat az elemeket emeljük ki, amelyek egyértelműen a töréspontok kezelésének kérdéséhez kapcsolódnak.

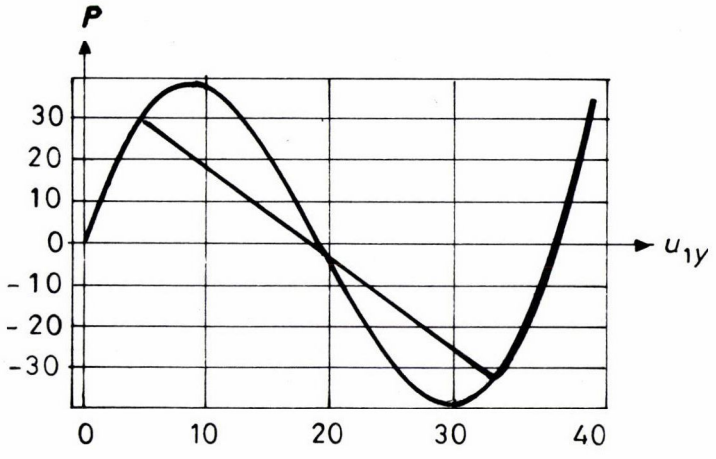
A töréspontok kezelése viszont értelemszerű módosításokkal megegyezik a folytonosan differenciálható energia függvény kritikus pontjainak meghatározásakor alkalmazott módszerrel. Így a töréspont jelzése a görbeszakasz végpontjaiban végzett ellenőrzésen alapul. A töréspont pontos helyét a (2.6)–(2.8) feltételek alapján meghatározható interpolációs polinomból képzett egyenlet megoldása szolgáltatja. A módszer kiemelendő előnye, hogy ha az adott íven több töréspont fordul elő, akkor a legelső egyszerű algebrai eszközökkel kiválasztható.

Az útkövető módszer alkalmazására a 3.1 ábrán látható egyszerű modellt vizsgáltuk.

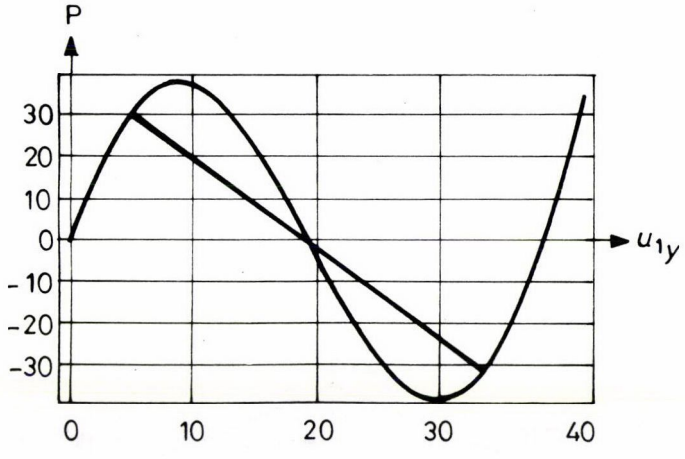


3.1 ábra

A rúdelemek kihajlásának figyelembevételével kapott eredményeket a 3.2 ábra, illetve 3.3 ábra szemlélteti. Az ábrákon viszonyítási alapként feltüntettem a rúdelemek kihajlásának figyelembevétele nélkül kapott görbét. A 3.2 ábrán jól látható,



3.2 ábra



3.3 ábra

hogy a Newton-típusú javító fázis a módszerben központi szerepet játszik, kihagyása esetén a globális és lokális hibák halmozódása miatt a számított görbe „elcsúszik". Az 3.3 ábra a Newton iterációval kapott eredményeket szemlélteti. A rúdelemek kihajlásával, illetve kihajlása nélkül számolt görbék a közös kezdeti, illetve végső szakaszokban megegyeznek. A töréspontok pontos helyét egy az ívhosszra vonatkozó algebrai egyenlet megoldása szolgáltatta. A töréspontokban meghatározott, Newton iterációval javított eredmények $P_1^t = 29,3 \times 10^5 KN$, illetve $P_2^t = -29,3 \times 10^5 KN$, amelyek megegyeznek KONDOH és ATLURI (1985) által számított értékekkel.

IRODALOM

[1] ABBOTT, J. P., „An efficient algorithm for the determination of certain bifurcation points", *J. Comput. Appl. Math.* **4** (1978), 19–27.

[2] ABBOTT, J. P., „Computing solution arcs of nonlinear equation with a parameter", *The Computer Journal* **23** (1980), 85–89.

[3] CSÉBFALVI, A. and CSÉBFALVI, GY., „Geometrically Non-Linear Analysis of Space Trusses Using Sparse Quasi-Newton Algorithms", *Proceedings, 11th International Conference on Mathematical Programming*, Mátrafüred, Hungary, March 21–26, 1992, 2–3.

[4] CSÉBFALVI, A. and VÁSÁRHELYI, A., „Some computational aspects of nonlinear space trusses", *Acta Techn. Acad. Sci. Hung.* (1993), (megjelenés alatt).

[5] CSÉBFALVI, A. and CSÉBFALVI, GY., „Post-buckling analysis of frames by a hybrid path-following method", *Generalized Convexity* (S. Komlósi, T. Rapcsák, S. Schaible, eds.) (Springer Verlag, 1993), 311–321.

[6] CSÉBFALVI, A., „Nemlineáris útkövető módszer tartószerkezetek stabilitásvizsgálatára I. Reguláris pontok", *Alkalmazott Matematikai Lapok* (1993a), (megjelenés alatt).

[7] CSÉBFALVI, A., „Nemlineáris útkövető módszer tartószerkezetek stabilitásvizsgálatára II. Elágazási és határpontok", *Alkalmazott Matematikai Lapok* (1993b), (megjelenés alatt).

[8] CRANDALL, M. G. and RABINOWITZ, P. H., „Bifurcation from simple eigenvalues", *J. Functional Analysis* **8** (1971), 321–340.

[9] FLORES, F. G. and GODOY, L. A., „Elastic postbuckling analysis via finite element and perturbation techniques, Part 1: Formulation", *I. J. Num. Meth. Eng* **33** (1992), 1775–1794.

[10] HALL, G. and WATT, J. M., *Modern Numerical Methods for Ordinary Differential Equations* (Oxford, 1978).

[11] HOUWEN, P. J., *Construction of integration formulas for initial value problems* (North-Holland Publishing Company, Amsterdam, 1977).

[12] KAMAT, M. P., WATSON, L. T. and VENKAYYA, V. B., „A quasi-Newton versus a homotopy method for structural analysis", *Comput. & Struct.* **17** (1983), 579–585.

[13] KAMAT, M. P., WATSON, L. T., „Determination of multiple equilibrium solutions through a deflation technique with tunelling", in C. Taylor, E. Hinton, D. R. Owen: Numerical Methods for Non-linear Problems, vol. 2 (Pineridge Press Swansea, 1984).

[14] KOITER, W. T., *Over de Stabiliteit van het elastisch Evenwicht*, Dissertation (Delft Technical University, H. J. Paris, Amsterdam, Holland, 1945).

[15] KONDOH, K., ATLURI, S. N., „Influence of local buckling on global instability simplified, large deformation, post-buckling analysis of plane trusses", *Comput & Struct.* **21** (1985), 613–627.

[16] KUBICEK, M., „Dependence of solution of nonlinear systems on a parameter, Algorithm 502", *CACM-Toms* **2** (1976), 98–107.

[17] LI, T. Y. and YORK, J. A., „A Simple Reliable Numerical Algorithm for Following Homotopy Paths", in Robinson, S. M.: Analysis and Computation of Fixed Point (Academic Press, New York, 1980).

[18] NOBLE, B., *Applied linear algebra* (Prentice-Hall, New Jersey, 1969).

[19] PAPADRAKAKIS, M., „Solution of the partial eigenproblem by iterative methods", *I. J. Num. Meth. Eng.* **20** (1984), 2283–2301.

[20] SHAMPE, L. F. and GORDON, M. K., *Computer Solution of Ordinary Differential Equations, The Initial Value Problem* (Freeman & Company, San Francisco, 1975).

[21] THOMPSON, J. M. T. and HUNT, G. W., *A general Theory of Elastic Stability* (Wiley, New York, 1973).

[22] WATSON, L. T., KAMAT, M. P and REASER, M. H., „A robust hybrid algorithm for computing multiple equilibrium solutions", *Eng. Comput.* **2** (March. 1985), 30–34.

[23] WATSON, L. T., „Numerical linear algebra aspects of globally convergent homotopy methods", *SIAM Rev.* **28** (1986), 529–545.

[24] WEINITSCHKE, H. J., „On the calculation of limit and bifurcation points in stability problems of elastic shells", *Int. J. Solids Structures* **21** (1985), 79–95.

[25] WERNER, B. and SPENCE, A., „The computation of symmetry-breaking bifurcation points", *SIAM J. Num. Anal.* **21** (1984), 388–399.

[26] WRIGGERS, P., WAGNER, W. and MIEHE, C., „A quadratically convergent procedure for the calculation of stability points in finite element analysis", *Computer Meth. Appl. Mech. and Eng.* **70** (1988), 329–347.

[27] WRIGGERS, P. and SIMO, J. C., „A general procedure for the direct computation of turning and bifurcation points", *I. J. Num. Meth. Eng.* **30** (1990), 155–176.

(Beérkezett: 1993. október 28.)

CSÉBFALVI ANIKÓ
POLLACK MIHÁLY MŰSZAKI FŐISKOLA
7625 PÉCS, BOSZORKÁNY ÚT 2.

STABILITY ANALYSIS OF SYSTEMS CHARACTERISED BY NONSMOOTH ENERGY FUNCTION

A. CSÉBFALVI

In this paper we present a path-following method for stability analysis of conservative mechanical systems with finite degree of freedom. Investigating singular points of a system characterised by one-dimensional force-deflection polygonal diagram, one finds that it is similar to the investigation of bifurcation points of the continual systems.

Alkalmazott Matematikai Lapok 17 (1993) 267–277

EGY STABILITÁSI FELTÉTEL A KVÁZI-IZOMETRIKUS KONJUGÁLT PÁROK MÓDSZERÉRE*

BODÓCS LÁSZLÓ

Budapest

A dolgozat speciális — izometrikus, illetve kvázi-izometrikus — vektorokkal indított, tetszőleges valós mátrixú lineáris egyenletrendszerek megoldására szolgáló konjugált irány algoritmusokat tárgyal, amelyekre teljesíthető Hestenes és Stiefel stabilitási feltétele.

1. Bevezetés

A dolgozat célja valós mátrixú lineáris egyenletrendszerek megoldására szolgáló általános konjugált irány algoritmus leírása, amely alkalmazásánál bizonyos numerikus stabilitási feltételek teljesíthetőek és megfelelő indító vektorok esetén a módszer további előnyös tulajdonságokkal rendelkezik. Az első konjugált irány algoritmusokat HESTENES és STIEFEL [9], valamint LÁNCZOS [10] készítették pozitív definit szimmetrikus mátrixokra. Ezeknek a módszereknek az alkalmazása nagy és ritka mátrixok esetén előnyös. Ilyen feladatok sokszor fordulnak elő a tudományos és mérnöki munkák során, például differenciál-egyenletek numerikus megoldása, geodéziai és űrkutatási feladatok, tomográfia, képfeldolgozás stb.

A 70-es évek közepétől intenzív kutatások folynak olyan konjugált irány módszerek kidolgozására, amelyek tetszőleges mátrixra alkalmazhatóak. A kutatások másik fő iránya különböző prekondícionálási technikák kidolgozása, amelyek az adott lineáris rendszer kondíciószámát csökkentik, s ezáltal a konvergencia sebesség növelhető. A konjugált irány módszerek különféle változatairól jó áttekinthető képet kaphatunk STOER [11], illetve GOLUB és O'LEARY [2] összefoglaló cikkéből, valamint HESTENES [8], illetve GOLUB és VAN LOAN [3] könyveiből. Külön a prekondícionált konjugált gradiens módszerekről találhatók tanulmányok a [12] és [13] kiadványokban.

Nyilvánvaló, hogy az általános — indefinit, nem szimmetrikus — mátrixú lineáris egyenletrendszer megoldása megkapható, ha a klasszikus konjugált gradiens módszert az $A^TAx = A^Tb$ normálegyenletre alkalmazzuk az eredeti $Ax = b$ egyenlet helyett. Ekkor azonban a rendszer kondíciószáma négyzetelődik és nagy mátrix esetén az elvégzendő numerikus munka is megduplázódik minden egyes iterációs lépésben. A hagyományos módszer általánosításai az úgynevezett bikonjugált módszerekhez vezettek, amelyek az eredeti feladatra alkalmazhatóak. Ezeknek az új eljárásoknak hátránya, hogy a mátrix indefinitsége miatt a rekurziós formulákban

*A dolgozat a T4012 számú OTKA téma támogatásával készült.

a nevezők nullává, vagy ellenőrízhetetlenül közel nullává válhatnak, s így a módszer vagy idő előtt megáll — azaz nem gyárt elég jó közelítést az egyenletrendszer megoldására —, vagy véges pontosságú aritmetikában numerikus instabilitások lépnek fel. A dolgozat olyan konjugált irány módszert tárgyal, amely mindkét problémára közelítő megoldást kínál. A dolgozat felépítése a következő:

A 2. fejezetben a konjugált párok módszerének rövid ismertetése után speciális — izometrikus, illetve kvázi-izometrikus — kezdővektorok bevezetésével a kvázi-izometrikus konjugált párok módszerének leírása következik.

A 3. fejezet a Hestenes–Stiefel féle stabilitási feltétel levezetését tartalmazza a kvázi-izometrikus konjugált párok módszerére. Megmutatjuk, hogy a rekurzió paraméterei választhatók úgy, hogy ez a stabilitási feltétel teljesüljön.

A mátrixokat nagybetűk, a vektorokat kisbetűk jelölik. Az A mátrix transzponáltját A^T, képterét $R(A)$, a nullterét pedig $N(A)$ jelöli. Az n-dimenziós egységmátrix I_n.

2. Az izometrikus és a kvázi-izometrikus konjugált párok módszere

Ez a fejezet a [6] és [7] dolgozatokban tárgyalt konjugált párok módszerének általánosítását ismerteti. Az általánosítás a rekurzióval gyártott vektorok skálázásán alapul. A fejezet első paragrafusa röviden bemutatja a konjugált párok módszerét, majd a módszer tulajdonságai kerülnek ismertetésre speciális — izometrikus — indító vektorok esetén. A kvázi-izometrikus konjugált párok módszerének leírását a fejezet harmadik része tartalmazza.

2.1 A konjugált párok módszere

A konjugált párok módszerének alábbi ismertetése a [6], [7] dolgozatokban publikált eredményeken alapszik.

2.1. Definíció. Legyen $A \in \mathbb{R}^{m,n}$, $\nu_j \in \mathbb{R}^m$, $u_k \in \mathbb{R}^n$, $1 \le j,\ k \le i$. Ekkor a $\{\nu_j, u_j\}_j$ rendszer konjugált párok (vagy A-konjugált párok) rendszere, ha

$$\nu_j^T A u_k = \alpha_j \delta_{jk}, \quad \alpha_j \neq 0,\ j,k = 1,2,\ldots,i, \tag{2.1}$$

ahol δ_{jk} a Kronecker szimbólum.

Ha a $\{\nu_j, u_j\}_{j=1}^i$ rendszer konjugált párokat tartalmaz, akkor definiálható az alábbi két projektor mátrix

$$P_i^b = I_m - \sum_{j=1}^{i} \frac{A u_j \nu_j^T}{\nu_j^T A u_j}, \quad P_i^j = I_n - \sum_{j=1}^{i} \frac{u_j \nu_j^T A}{\nu_j^T A u_j}. \tag{2.2}$$

E két projektor segítségével felírható a konjugált párok módszere [6]:

2.2. Algoritmus. Ha $A \in \mathbb{R}^{m,n}$, P_i^b és P_i^j a (2.2) által definiált projektorok, r_0 és q_0 nemzérus kezdővektorok, akkor a

$$r_{i+1} = P_i^b r_i, \qquad q_{i+1}^T = q_i^T P_i^j \tag{2.3}$$

$$\nu_{i+1}^T = r_{i+1}^T P_i^b, \qquad u_{i+1} = P_i^j q_{i+1} \tag{2.4}$$

rekurziók a ν_{i+1}, u_{i+1} konjugált párokat állítják elő $i = 0, 1, 2 \ldots$-re, ahol $P_0^b = I_m$ és $P_0^j = I_n$.

Megjegyzendő, hogy a [7] dolgozatban tárgyalt algoritmus a (2.3)–(2.4) rekurzióknál általánosabb, ugyanis ott szerepelnek kondícionáló mátrixok, valamint a ν_{i+1} és az u_{i+1} vektorok hosszát skálázó skalárok is. A kondícionáló mátrixok bevezetésétől ebben a dolgozatban az egyszerűség kedvéért eltekintünk, a vektorok skálázásának pedig a későbbiekben lesz szerepe.

A 2.2. Algoritmus rekurziói az alábbi alakra egyszerűsíthetők [6]:

$$r_{i+1} = r_i - \frac{\|r_i\|_2^2}{\nu_i^T A u_i} A u_i, \quad q_{i+1}^T = q_i^T - \frac{\|q_i\|_2^2}{\nu_i^T A u_i} \nu_i^T A, \tag{2.5}$$

$$\nu_{i+1} = r_{i+1} + \frac{\|r_{i+1}\|_2^2}{\|r_i\|_2^2} \nu_i, \quad u_{i+1} = q_{i+1} + \frac{\|q_{i+1}\|_2^2}{\|q_i\|_2^2} u_i \,. \tag{2.6}$$

Ez a rekurzió szimmetrikus mátrix esetén megegyezik a jól ismert konjugált gradiens módszerrel, amennyiben a kezdővektorokra az $r_0 = \nu_0 = q_0 = u_0$ feltétel teljesül.

A konjugált párok módszerének analízise, valamint a (2.5)–(2.6) rekurziótól eltérő más módszertípusok a [4]–[7] dolgozatokban találhatóak.

A (2.5)–(2.6) rekurziós formulák alakjából látszik, hogy a nevezőkben szerepel a $\nu_i^T A u_i$ belső szorzat, amely tetszőleges mátrix esetén nullává válhat. A következő paragrafus olyan indító vektorokat mutat be, amelyek alkalmazásával ez a probléma megkerülhető.

2.2 Izometrikus vektorok

Legyen az $A \in \mathbb{R}^{m,n}$ mátrix rangja ϱ és a szinguláris érték felbontása $A = E\Sigma F^T$, ahol a Σ diagonális mátrix nemzérus elemei az A mátrix pozitív szinguláris értékei, az $E \in \mathbb{R}^{m,\varrho}$ és $F \in \mathbb{R}^{n,\varrho}$ mátrixok pedig ortogonális mátrixok, amelyek oszlopvektorai kifeszítik az A illetve az A^T mátrixok képtereit. Ekkor definiálhatók az alábbi speciális vektorok.

2.3. Definíció. (Hegedűs [5]).

Az $r \in \mathbb{R}^m$ és a $q \in \mathbb{R}^n$ vektorok izometrikus (vagy A-izometrikus) párt alkotnak, ha

$$\begin{aligned} &i) \quad E^T r = F^T q \qquad \text{és} \\ &ii) \quad \|r\|_2 = \|q\|_2 \,. \end{aligned} \tag{2.7}$$

Az izometrikus párok halmazát A_{iz} jelöli.

2.4. LEMMA. *Ha $r \in R(A)$, $q \in R(A^T)$ és $E^T r = F^T q$, akkor $\|r\|_2 = \|q\|_2$.*

Bizonyítás. Ha $r \in R(A)$, $q \in R(A^T)$, akkor léteznek y és w vektorok, hogy $r = Ay$ és $q = A^T w$. Ha $E\Sigma F^T$, akkor

$$\|E^T r\|_2^2 = r^T E E^T r = y^T A^T E E^T A y = \\ = y^T F \Sigma E^T E E^T E \Sigma F^T y = y^T A^T A y = \|Ay\|_2^2 = \|r\|_2^2$$

és

$$\|F^T q\|_2^2 = q^T F F^T q = w^T A F F^T A^T w = \\ = w^T E \Sigma F^T F F^T F \Sigma E^T w = w^T A A^T w = \|A^T w\|_2^2 = \|q\|_2^2 \,.$$

Az $E^T r = F^T q$ feltételből $\|r\|_2 = \|q\|_2$ következik. □

A továbbiakban feltételezzük, hogy az $r \in R(A)$ és a $q \in R(A^T)$ *mindig teljesül*, így a lemma értelmében elég csak az $E^T r = F^T q$ feltétellel foglalkozni az izometrikusság vizsgálatakor. A következő állítások a definíció alapján könnyen bizonyíthatóak.

2.5. LEMMA. *Legyen $B_k = A(A^T A)^k$, ahol $k \geq 0$ és $\{r, q\} \in A_{iz}$. Ekkor*

i) $\{B_k q, B_k^T r\} \in A_{iz}$,

ii) $r^T B_k q > 0$, *ha* $r \notin N(A^T)$ *és* $q \notin N(A)$,

iii) $\|B_k^T r\|_2 = \|B_k q\|_2$,

iv) *ha* $\{u, \nu\} \in A_{iz}$ *és* $\alpha, \beta \in \mathbb{R}$, *akkor* $\{\alpha r + \beta u, \alpha q + \beta \nu\} \in A_{iz}$,

v) *ha A szinguláris érték felbontása* $A = E\Sigma E^T$, *akkor* $\{r, r\} \in A_{iz}$,

vi) *ha* $A = E\Sigma_1 F^T$ *és* $B = E\Sigma_2 F^T$, *akkor* $A_{iz} = B_{iz}$.

A lemma következménye, hogy ha $\{r_0, q_0\} \in A_{iz}$ a (2.5)–(2.6), rekurziókban, akkor a gyártott r_i, q_i és ν_i, u_i vektorpárok izometrikusak minden i-re és a módszer úgy viselkedik mint egy szimmetrikus nemnegatív definit mátrix esetén, amelynek sajátértékei $\sigma_1, \sigma_2, \ldots, \sigma_p$, hiszen a rekurziós formulák az alábbi alakokra írhatók át:

$$r_{i+1} = r_i - \frac{\|r_i\|_2^2}{\nu_i^T E \Sigma E^T \nu_i} E \Sigma E^T \nu_i, \qquad q_{i+1} = q_i - \frac{\|q_i\|_2^2}{u_i^T F \Sigma F^T u_i} F \Sigma F^T u_i$$

$$\nu_{i+1} = r_{i+1} + \frac{\|r_{i+1}\|_2^2}{\|r_i\|_2^2} \nu_i, \qquad u_{i+1} = q_{i+1} + \frac{\|q_{i+1}\|_2^2}{\|q_i\|_2^2} u_i \,.$$

Látható, hogy a nevezőkben szereplő belső szorzatok pozitívak, ha a bennük lévő vektorok nem elemei a megfelelő nulltereknek. Megjegyzendő, hogy ha a vektorokat az A mátrix szinguláris vektorai által kifeszített terekbe transzformáljuk, akkor a szinguláris értékekkel definiált diagonális mátrixra felírt szimmetrikus konjugált irány algoritmushoz jutunk.

2.3 A kvázi-izometrikus konjugált párok módszere

Az előzőekben leírtakból következik, hogy ha a konjugált párok módszerét izometrikus vektorokkal indítjuk, akkor a kondíciószám nem négyzetelődik, s a rekurzió mindaddig folytatható, amíg a generált vektorok nulltérbe nem esnek, amely esetben a lineáris egyenletrendszer megoldható. A gyakorlatban azonban nem állnak rendelkezésre izometrikus vektorok, hanem csak azok valamilyen közelítései az úgynevezett *kvázi-izometrikus* vektorok. A közelítés módjára a későbbiekben történik utalás. A következő definíció mértéket ad a közelítés jóságára.

2.6. Definíció. Legyen $A = E\Sigma F^T$. Ha az r és q vektorokra

$$\|E^T r - F^T q\|_2 \leq \varepsilon \tag{2.8}$$

és

$$\big|\, \|r\|_2 - \|q\|_2 \big| \leq \varepsilon \tag{2.9}$$

teljesül valamely ε pozitív számra, akkor az $\{r, q\}$ vektorok *kvázi-izometrikusak* (vagy *kvázi-A-izometrikusak*). A kvázi-izometrikus vektorok halmazát A^{ε}_{kiz} jelöli.

Amennyiben $\|E^T r\|_2 = \|r\|_2$ és $\|F^T q\|_2 = \|q\|_2$, akkor (2.8)-ból már következik (2.9) az ismert norma-egyenlőtlenség miatt. *A továbbiakban feltesszük, hogy ez a feltétel mindig teljesül.*

A következő lemma az $\{r, q\}$ izometrikus párhoz rögzített r vektor esetén kvázi-izometrikus párt rendel a q vektor tetszőleges approximációját felhasználva.

2.7. LEMMA. *Ha* $\{r, q\} \in A_{iz}$, q_a *a* q *vektor valamilyen közelítése és* $\varepsilon = \|q - q_a\|_2$, *akkor* $\{r, q_a\} \in A^{\varepsilon}_{kiz}$.

Bizonyítás.

$$\left\|E^T r - F^T q_a\right\|_2 = \left\|E^T r - F^T q + F^T(q - q_a)\right\|_2 = \left\|F^T(q - q_a)\right\|_2 \leq \varepsilon\,. \qquad \square$$

A konjugált párok módszerét lehetséges például az $r_0 = b - Ax_0$ és a $q_0 = A^T r_0$ kvázi-izometrikus vektorokkal indítani. Ekkor $\varepsilon \leq \|r_0\|_2$, ha $\|I - \Sigma\|_2 \leq 1$, ami az általánosság megszorítása nélkül feltehető. A kezdővektorokra vonatkozó további vizsgálatok találhatóak az [1] dolgozatban.

A kvázi-izometrikus konjugált párok módszere a következő alakban írható fel:

2.8. Algoritmus. Legyen $A \in \mathbb{R}^{m,n}$ és $\Lambda = \langle \lambda_i \rangle$, $M = \langle \mu_i \rangle$, $\Omega = \langle \omega_i \rangle$, $\Delta = \langle \delta_i \rangle$ valós diagonális mátrixok, amelyeknek diagonális elemei nullától különbözőek. Ekkor az $r_0 \neq 0$ és $q_0 \neq 0$ vektorokra az

$$r_{i+1} = \delta_{i+1} P_i^{\ell} r_i, \quad q_{i+1}^T = \omega_{i+1} q_i^T P_i^r, \tag{2.10}$$

$$\nu_{i+1}^T = \mu_{i+1} r_{i+1}^T P_i^{\ell}, \quad u_{i+1} = \lambda_{i+1} P_i^r q_{i+1}, \tag{2.11}$$

rekurziók konjugált párokat gyártanak.

Látható, hogy ez a módszer csak annyiban különbözik az eredeti algoritmustól, hogy az előállított vektorok hosszát változtatjuk.

A 2.8. Algoritmus az alábbi alakra egyszerűsíthető:

$$r_{i+1} = \delta_{i+1}(r_i - \alpha_i A u_i), \quad q_{i+1} = \omega_{i+1}(q_i - \beta_i A^T \nu_i) \tag{2.12}$$

$$\begin{aligned} \nu_{i+1} &= \mu_{i+1} \left(r_{i+1} + \frac{\|r_{i+1}\|_2^2}{\delta_{i+1}\mu_i \|r_i\|_2^2} \nu_i \right), \\ u_{i+1} &= \lambda_{i+1} \left(q_{i+1} + \frac{\|q_{i+1}\|_2^2}{\omega_{i+1}\lambda_i \|q_i\|_2^2} u_i \right), \end{aligned} \tag{2.13}$$

ahol

$$\alpha_i = \frac{\mu_i \|r_i\|_2^2}{\nu_i^T A u_i} \qquad \text{és} \qquad \beta_i = \frac{\lambda_i \|q_i\|_2^2}{u_i^T A^T \nu_i}.$$

Az egyenletrendszer megoldása az $x_{i+1} = x_i + \dfrac{\mu_i \|r_i\|_2^2}{\delta_i \delta_{i-1} \dots \delta_0 \nu_i^T A u_i} u_i$ iterációval számolható.

A következő tétel megmutatja, hogy a fenti rekurziók alkalmas paraméter választással egyre „jobb" kvázi-izometrikus vektorokat gyártanak.

2.9. TÉTEL. *Legyen az $A \in \mathbb{R}^{m,n}$ mátrix szinguláris érték felbontása $A = E\Sigma F^T$, ahol $\|\Sigma\|_2 \leq 1$, $0 < \varepsilon_i$, valamint $\|E^T r_i - F^T q_i\|_2 \leq \varepsilon_i$ és $\|E^T \nu_i - F^T u_i\|_2 \leq \varepsilon_i$. Ha $\omega_{i+1} = \delta_{i+1}$ és $|\delta_{i+1}| < \dfrac{\varepsilon_i}{(1 + |\alpha_i|)\varepsilon_i + |\gamma_i| \, \|\nu_i\|_2}$, ahol $\gamma_i = \beta_i - \alpha_i$, akkor*

$$\|E^T r_{i+1} - F^T q_{i+1}\|_2 \leq \varepsilon_{i+1}. \tag{2.14}$$

Ha $\lambda_{i+1} = \mu_{i+1}$ és $|\lambda_{i+1}| < \dfrac{\varepsilon_i}{\varepsilon_i + \left| \dfrac{1}{\delta_{i+1}\lambda_i} \right| \left(\dfrac{\|q_{i+1}\|_2^2}{\|q_i\|_2^2} \varepsilon_i + |\eta_i| \|\nu_i\|_2 \right)}$,

ahol $\eta_i = \dfrac{\lambda_i \|r_{i+1}\|_2^2}{\mu_i \|r_i\|_2^2} - \dfrac{\|q_{i+1}\|_2^2}{\|q_i\|_2^2}$, akkor

$$\|E^T \nu_{i+1} - F^T u_{i+1}\|_2 \leq \varepsilon_{i+1}, \tag{2.15}$$

ahol $\varepsilon_{i+1} < \varepsilon_i$.

Bizonyítás. Legyen $E^T r_i - F^T q_i = d_i^{(1)}$ és $E^T \nu_i - F^T u_i = d_i^{(2)}$. Ekkor

$$E^T r_{i+1} - F^T q_{i+1} = \delta_{i+1}\Big((E^T r_i - \alpha_i \Sigma F^T u_i) - (F^T q_i - \beta_i \Sigma E^T \nu_i)\Big) =$$
$$= \delta_{i+1}\left(d_i^{(1)} + \beta_i \Sigma E^T \nu_i - \alpha_i \Sigma F^T u_i\right) = \delta_{i+1}\left(d_i^{(1)} + \alpha_i \Sigma d_i^{(2)} + \gamma_i \Sigma E^T \nu_i\right).$$

Normára áttérve az $\|E^T \nu_i\|_2 \le \|\nu_i\|_2$ egyenlőtlenség felhasználásával $\|E^T r_{i+1} - F^T q_{i+1}\|_2 \le |\delta_{i+1}|\big((1+|\alpha_i|)\varepsilon_i + |\gamma_i|\,\|\nu_i\|_2\big)$ adódik, így igaz a (2.14) állítás. Mivel

$$E^T \nu_{i+1} - F^T u_{i+1} =$$
$$= \lambda_{i+1}\left(E^T r_{i+1} - F^T q_{i+1} + \frac{1}{\delta_{i+1}\lambda_i}\left(\frac{\lambda_i \|r_{i+1}\|_2^2}{\mu_i \|r_i\|_2^2} E^T \nu_i - \frac{\|q_{i+1}\|_2^2}{\|q_i\|_2^2} F^T u_i\right)\right) =$$
$$= \lambda_{i+1}\left(E^T r_{i+1} - F^T q_{i+1} + \frac{1}{\delta_{i+1}\lambda_i}\left(\frac{\|q_{i+1}\|_2^2}{\|q_i\|_2^2}(E^T \nu_i - F^T u_i) + \eta_i E^T \nu_i\right)\right) =$$
$$= \lambda_{i+1}\left(d_{i+1}^{(1)} + \frac{1}{\delta_{i+1}\lambda_i}\left(\frac{\|q_{i+1}\|_2^2}{\|q_i\|_2^2} d_i^{(2)} + \eta_i E^T \nu_i\right)\right),$$

és (2.14) miatt $\|d_{i+1}^{(1)}\|_2 < \varepsilon_i$, így normabecslésre áttérve (2.15) is teljesül. □

A következő lemma a módszer gyakorlati alkalmazásával kapcsolatban fontos.

2.10. LEMMA. *Legyen* $\{r, q\} \in A_{kiz}^{\varepsilon}$. *Ha* $\|r\|_2 \ge 2$ *vagy* $\|q\|_2 \ge 2$, *akkor* $\left\{\frac{r}{\|r\|_2}, \frac{q}{\|q\|_2}\right\} \in A_{kiz}^{\varepsilon_1}$, *ahol* $\varepsilon_1 \le \varepsilon$.

Bizonyítás. Legyen $\|r\|_2 - \|q\|_2 = \delta$, ahol $|\delta| \le \varepsilon$. Ekkor

$$\zeta = \left\| E^T \frac{r}{\|r\|_2} - F^T \frac{q}{\|q\|_2} \right\|_2 = \left\| \frac{1}{\|r\|_2}\left(E^T r - F^T q - \frac{\delta}{\|q\|_2} F^T q\right)\right\|_2 \le$$
$$\le \frac{1}{\|r\|_2}\left(\left\|E^T r - F^T q\right\|_2 + \frac{|\delta|}{\|q\|_2}\left\|F^T q\right\|_2\right) \le \frac{2}{\|r\|_2}\varepsilon.$$

A $\zeta \le \frac{2}{\|q\|_2}\varepsilon$ egyenlőtlenség hasonlóan igazolható, így a lemma bizonyítást nyert. □

A lemma gyakorlati fontossága abban rejlik, hogy az előállított vektorok normálhatók, ha a lemma feltételei teljesülnek. Ebben az esetben elkerülhető, hogy a vektorok a null-vektorhoz nagyon közel kerüljenek, ugyanis mind a δ_{i+1}, mind a λ_{i+1} kisebb egynél.

3. A Hestenes–Stiefel féle stabilitási feltétel a kvázi-izometrikus párok módszerénél

Amennyiben a (2.12)–(2.13) algoritmust szimmetrikus pozitív definit mátrixra alkalmazzuk, az indító vektorokat $\nu_0 = u_0 = q_0 = r_0 = b - Ax_0$-nak és a skalárokat egynek választjuk, akkor a klasszikus konjugált gradiens módszerhez jutunk [9] ami az alábbi alakban írható fel:

$$r_{i+1} = r_i - \alpha_i A\nu_i, \qquad \alpha_i = \|r_i\|_2^2 / \nu_i^T A\nu_i \tag{3.1}$$

$$\nu_{i+1} = r_{i+1} + d_i\nu_i, \qquad d_i = \frac{\|r_{i+1}\|_2^2}{\|r_i\|_2^2}. \tag{3.2}$$

Pontos számolás esetén $i \neq k$-ra az

$$r_i^T r_k = 0, \qquad \nu_i^T A\nu_k = 0 \tag{3.3}$$

ortogonalitási feltételek teljesülnek. Kerekítési hibák fellépésekor azonban ezek az ortogonalitási tulajdonságok már egzaktul nem igazak, sőt a hiba az iterációs lépések számának növekedésével olyan nagy lehet, hogy a módszer nem konvergál az egyenletrendszer megoldásához. HESTENES és STIEFEL [9, 8.1 fejezet] hibaterjedési formulákat vezetett le, amelyek azt mutatják meg, hogy az $r_{i-1}^T r_i$ és a $\nu_{i-1}^T A\nu_i$ belső szorzatokban fellépő hiba hogyan terjed tovább a következő iterációs lépésre. Ezek az összefüggések $i \geq 1$-re a következők:

$$r_i^T r_{i+1} = d_{i-1}\alpha_i \nu_{i-1}^T A\nu_i\,, \tag{3.4}$$

$$\nu_i^T A\nu_{i+1} = \frac{1}{\alpha_i} r_i^T r_{i+1}\,, \tag{3.5}$$

$$r_i^T r_{i+1} = \frac{d_{i-1}\alpha_i}{\alpha_{i-1}} r_{i-1}^T r_i\,, \tag{3.6}$$

$$\nu_i^T A\nu_{i+1} = d_{i-1}\nu_{i-1}^T A\nu_i\,. \tag{3.7}$$

A (3.6) és (3.7) összefüggések az alábbi alakban is felírhatóak:

$$\frac{r_i^T r_{i+1}}{\|r_i\|_2^2} = \frac{\alpha_i}{\alpha_{i-1}} \frac{r_{i-1}^T r_i}{\|r_{i-1}\|_2^2}\,, \tag{3.8}$$

$$\frac{\nu_i^T A\nu_{i+1}}{\nu_i^T A\nu_i} = \frac{\alpha_i}{\alpha_{i-1}} \frac{\nu_{i-1}^T A\nu_i}{\nu_{i-1}^T A\nu_{i-1}}\,. \tag{3.9}$$

Ezeknek a formuláknak a következő jelentésük van. Ha az $r_j^T r_k$ és a $\nu_j^T A\nu_k$ elemekből elkészítjük az R és a V mátrixokat, akkor a (3.8) és a (3.9) összefüggések bal oldala az ugyanabban a sorban lévő azon elemek hányadosát reprezentálja, amelyek közül az egyik a főátlóban, a másik attól jobbra helyezkedik el. A (3.8) és a (3.9) összefüggések így ezen hányadosok változását mutatják a főátló mentén lefelé haladva. Ezen megfontolások alapján a következő stabilitási feltételt definálhatjuk szimmetrikus, pozitív definit mátrix esetén:

3.1. Definíció (Hestenes–Stiefel).

Legyen az A mátrix szimmetrikus pozitív definit. A (3.1)–(3.2) rekurzió stabil, ha $\dfrac{\alpha_i}{\alpha_{i-1}} < 1$ minden i-re.

A továbbiakban megmutatjuk, hogy a (3.4)–(3.9) összefüggések általánosíthatóak a kvázi-izometrikus konjugált párok módszerére is, és hasonló stabilitási feltétel definiálható mint a klasszikus konjugált gradiens módszer esetén. Megmutatjuk, hogy a skalár szorzók alkalmas választásával ez a stabilitási feltétel teljesíthető, sőt a kvázi-izometrikusságra kirótt kritériummal együtt egyszerre is kielégíthető.

A következő lemma a (2.12)–(2.13) általános rekurzióra felírt hibaterjedési formulákat mutatja. A lemma bizonyítását ez a dolgozat annak hosszadalmassága miatt nem tartalmazza, az összefüggések hasonló módon vezethetők le, mint a hagyományos módszer esetén [9].

3.2. LEMMA. *Legyen A tetszőleges valós mátrix. Ekkor a* (2.12)–(2.13) *általános rekurzióra igazak az alábbi összefüggések:*

$$r_i^T r_{i+1} = \frac{\delta_{i+1}}{\delta_i \mu_{i-1}} \alpha_i d_{i-1} \nu_{i-1}^T A u_i \,, \tag{3.10}$$

$$q_i^T q_{i+1} = \frac{\omega_{i+1}}{\omega_i \lambda_{i-1}} \beta_i c_{i-1} u_{i-1}^T A^T \nu_i \,, \tag{3.11}$$

$$\nu_i^T A u_{i+1} = \frac{\lambda_{i+1}}{\beta_i} q_i^T q_{i+1} \,, \tag{3.12}$$

$$\nu_{i+1}^T A u_i = \frac{\mu_{i+1}}{\alpha_i} r_{i+1}^T r_i \,, \tag{3.13}$$

$$\frac{r_{i+1}^T r_i}{\|r_i\|_2^2} = \frac{\delta_{i+1}\lambda_i}{\delta_i \lambda_{i-1}} \frac{\alpha_i}{\alpha_{i-1}} \frac{q_i^T q_{i-1}}{\|q_{i-1}\|_2^2} = \gamma_1 \frac{q_i^T q_{i-1}}{\|q_{i-1}\|_2^2} \,, \tag{3.14}$$

$$\frac{q_{i+1}^T q_i}{\|q_i\|_2^2} = \frac{\omega_{i+1}\mu_i}{\omega_i \mu_{i-1}} \frac{\beta_i}{\beta_{i-1}} \frac{r_i^T r_{i-1}}{\|r_{i-1}\|_2^2} = \gamma_2 \frac{r_i^T r_{i-1}}{\|r_{i-1}\|_2^2} \,, \tag{3.15}$$

$$\frac{\nu_{i+1}^T A u_i}{\nu_i^T A u_i} = \frac{\mu_{i+1}\delta_{i+1}}{\mu_i \delta_i} \frac{\alpha_i}{\alpha_{i-1}} \frac{u_i^T A^T \nu_{i-1}}{u_{i-1}^T A^T \nu_{i-1}} = \gamma_3 \frac{u_i^T A^T \nu_{i-1}}{u_{i-1}^T A^T \nu_{i-1}} \,, \tag{3.16}$$

$$\frac{\nu_i^T A u_{i+1}}{\nu_i^T A u_i} = \frac{\lambda_{i+1}\omega_{i+1}}{\lambda_i \omega_i} \frac{\beta_i}{\beta_{i-1}} \frac{u_{i-1}^T A^T \nu_i}{u_{i-1}^T A^T \nu_{i-1}} = \gamma_4 \frac{u_{i-1}^T A^T \nu_i}{u_{i-1}^T A^T \nu_{i-1}} \,, \tag{3.17}$$

ahol $\alpha_i = \dfrac{\mu_i \|r_i\|_2^2}{\nu_i^T A u_i}$, $\beta_i = \dfrac{\lambda_i \|q_i\|_2^2}{u_i^T A^T \nu_i}$, $c_i = \dfrac{\|q_{i+1}\|_2^2}{\|q_i\|_2^2}$, $d_i = \dfrac{\|r_{i+1}\|_2^2}{\|r_i\|_2^2}$.

A (3.14)–(3.17) összefüggésekből hasonló kapcsolatok értelmezhetők az $R = (r_j^T r_k)$, a $Q = (q_j^T q_k)$, valamint a $V_1 = (\nu_j^T A u_k)$ és a $V_2 = (u_j^T A^T \nu_k)$ mátrixok

elemeire mint a hagyományos konjugált gradiens módszernél. A különbség annyi, hogy jelen esetben az R mátrix megfelelő elemei a Q mátrix, míg a V_1 megfelelő elemei a V_2 mátrix elemeivel vannak kapcsolatban. Az általánosított Hestenes–Stiefel féle stabilitási feltétel a következőképpen fogalmazható meg:

3.3. Definíció. A (2.12)–(2.13) rekurzió stabil, ha a (3.14)–(3.17) formulákban szereplő γ_j $(1 \le j \le 4)$ együtthatókra teljesül az $|\gamma_j| < 1$ feltétel.

A (3.14)–(3.17) összfüggésekből következik az alábbi tétel.

3.4. TÉTEL. *A (2.12)–(2.13) rekurzió stabil, ha*

$$|\delta_{i+1}| < \left|\delta_i \frac{\lambda_{i-1}}{\lambda_i}\frac{\alpha_{i-1}}{\alpha_i}\right|, \quad |\omega_{i+1}| < \left|\omega_i \frac{\mu_{i-1}}{\mu_i}\frac{\beta_{i-1}}{\beta_i}\right|,$$

$$|\mu_{i+1}| < \left|\mu_i \frac{\delta_i}{\delta_{i+1}}\frac{\alpha_{i-1}}{\alpha_i}\right|, \quad |\lambda_{i+1}| < \left|\lambda_i \frac{\omega_i}{\omega_{i+1}}\frac{\beta_{i-1}}{\beta_i}\right|.$$

Legyenek $\delta_{i+1}^{(1)}$, $\lambda_{i+1}^{(1)}$ a 2.9 tétel feltételei szerint, a $\delta_{i+1}^{(2)}$, $\omega_{i+1}^{(2)}$, $\lambda_{i+1}^{(2)}$, $\mu_{i+1}^{(2)}$ pedig a 3.4 tétel feltételei szerint választott pozitív skalárok. A fenti eredmények következménye az alábbi tétel, amely megmutatja, hogy a kvázi-izometrikus konjugált párok módszerében a skalárok alkalmas választásával egyszerre biztosítható az, hogy a rekurzió a 2.6 definíció értelmében egyre jobb kvázi-izometrikus vektorokat állít elő, valamint az általánosított Hestenes–Stiefel stabilitási feltétel.

3.5. TÉTEL. *A (2.12)–(2.13) kvázi-izometrikus konjugált párok módszerére teljesül az általánosított Hestenes–Stiefel stabilitási feltétel, ha*

$$\delta_{i+1} = \omega_{i+1} = \min\left(\delta_{i+1}^{(1)}, \delta_{i+1}^{(2)}, \omega_{i+1}^{(2)}\right) \quad \text{és} \quad \lambda_{i+1} = \mu_{i+1} = \min\left(\lambda_{i+1}^{(1)}, \lambda_{i+1}^{(2)}, \mu_{i+1}^{(2)}\right)$$

IRODALOM

[1] BODÓCS, L., „Conjugate direction algorithms with special (isometric) initial vectors", *Report KFKI-1993-14/M* (Hungarian Academy of Sciences, Central Research Institute for Physics, Budapest, 1993).

[2] GOLUB, G. H. and O'LEARY, D. P., „Some history of conjugate gradient and Lanczos algorithms: 1948-1976", *SIAM Review* **31** (1989), 50.

[3] GOLUB, G. H. and VAN LOAN, C. F., *Matrix Computations*, 2nd Edition (John Hopkins Univ. P., Baltimore and London, 1989).

[4] HEGEDŰS, CS. J., „Generating conjugate directions for arbitrary matrices by matrix equations, Parts 1 and 2", *Report KFKI-1990-36/M* (Hungarian Academy of Sciences, Central Research Institute for Physics, Budapest, 1990).

[5] HEGEDŰS, CS. J., „Generation of conjugate directions for abtitrary matrices and solution of linear systems", in Contributed Papers NATO ASI Conf., Computer Algorithms for Solving Linear Algebraic Equations: the State of the Art (E. Spedicato and M. T. Vespucci, eds.) (Univ. of Bergamo, Bergamo, Italy, 1991), 26–49.

[6] HEGEDŰS, CS. J. and BODÓCS, L., „General recursions for A-conjugate vector pairs", *Research Report, KFKI 1982-56* (Hungarian Academy of Sciences, Central Research Institute for Physics, Budapest, 1982).

[7] HEGEDŰS, CS. J. and BODÓCS, L., „Konjugált irányok előállítása — a konjugált párok módszere", *Alkalmazott Mat. Lapok* **11** (1985), 297.
[8] HESTENES, M. R., „Conjugate direction methods in optimization", *Applications of Mathematics 12* (A. V. Balakrishnan, ed.) (Springer–Verlag, New York, Heidelberg, Berlin, 1980).
[9] HESTENES, M. R. and STIEFEL, E., „Methods of the conjugate gradients for solving linear systems", *J. Res. Nat. Bur. Standards* **Sect. B 49** (1952), 409.
[10] LANCZOS, C., „Solution of systems of linear equations by minimized iterations", *J. Res. Nat. Bur. Standards* **Sect. B 49** (1952), 33.
[11] STOER, J., „Solution of large linear systems of equations by conjugate gradient type methods", in Mathematical Programming — The State of the Art (A. Bachem, M. Grötschel and B. Korte, eds.) (Springer, Berlin Heidelberg New York Tokyo, 1983), 540.
[12] „Special Issue on Preconditioned Conjugate Gradient Methods", *BIT 29* **No. 4** (1989).
[13] *Preconditioned Conjugate Gradient Methods* (O. Axelsson and L. Yu. Kolotilina, eds.), Proceedings, Nijmegen 1989, Lecture Notes in Mathematics 1457 (Springer–Verlag, Berlin–Heidelberg –New York–London–Paris–Tokyo–Hong Kong–Barcelona, 1990).

(Beérkezett: 1994. április 23.)

BODÓCS LÁSZLÓ
MTA KFKI ANYAGTUDOMÁNYI KUTATÓ INTÉZET
H-1525 BUDAPEST, PF. 49.

A STABILITY CONDITION FOR THE METHOD OF QUASI-ISOMETRIC CONJUGATE PAIRS

L. BODÓCS

Conjugate direction algorithms with special initial — isometric, or quasi-isometric — vectors are described for solving linear systems with respect to an arbitarary real matrix. For these methods the stability condition of Hestenes and Stiefel can be satisfied.

Alkalmazott Matematikai Lapok **17** (1993) 279–286

KALMAN–FÉLE RANGFELTÉTELEK AZ IDŐTŐL FÜGGŐ LINEÁRIS RENDSZEREKRE

MOLNÁR SÁNDOR, SZIGETI FERENC ÉS VERA CARMEN E.

A közismert Kalman-féle algebrai rangfeltételeket általánosítjuk időtől függő lineáris rendszerekre. A többváltozós rangfeltételekben szereplő tagok a rendszer struktúramátrixai által generált Lie-algebra bázisának az elemei.

Kulcsszavak: Lie-algebra, lineáris rendszer, Kalman-féle rangfeltétel, exponenciális szorzat.

1. Bevezetés

Egy korábbi dolgozatunkban Kalman-féle rangfeltételt adtunk meg az $A(t) = a_1(t)A_1 + a_2(t)A_2 \in \mathbb{R}^{n\times n}$ struktúramátrixszal rendelkező időfüggő lineáris rendszer irányíthatóságára és megfigyelhetőségére [2]. Feltettük, hogy $[A_1, A_2] = A_2$, azaz, hogy az $A(t)$; $t \in [0,T]$ mátrixok által generált L Lie-algebra volt az egyetlen kétdimenziójú nem kommutatív Lie-algebra. Most ezt az eredményt általánosítjuk az

$$\begin{aligned} \dot{x}(t) &= A(t)x(t) + Bu(t) \\ y(t) &= Cx(t) \end{aligned} \tag{1}$$

véges dimenziós rendszerre, ahol $A : [0,T] \to \mathbb{R}^{n\times n}$ analitikus, $B \in \mathbb{R}^{n\times r}$, $C \in \mathbb{R}^{s\times n}$. Legyen L az $A(t)$; $t \in [0,T]$ mátrixok által generált Lie-algebra. A véges dimenzió miatt tekinthetjük az $A_1, \ldots, A_k$ véges bázist. Ekkor $A(t)$ a fenti bázissal kifejezhető az

$$A(t) = a_1(t)\ A_1 + \cdots + a_k(t)A_k \tag{2}$$

alakban. A következőkben rangfeltételt adunk az (1) rendszer irányíthatóságára és megfigyelhetőségére. A

$$G_c = \int_0^T \Phi(t)^{-1} B B^* \Phi(t)^{*-1} dt \text{ és a}$$

$$G_o = \int_0^T \Phi(t)^* C^* C \Phi(t) dt$$

formulával definiált irányíthatósági és megfigyelhetőségi Gram-mátrixokkal kifejezett Kalman-féle feltételtől indulunk ki, ahol Φ az (1) alapmátrixa [1]. Az (1) alapmátrixának exponenciálisok szorzataként való reprezentációját [4] alkalmazva a

$$\begin{aligned} &\operatorname{rang}(B, \ldots, A_1^{n_1} \ldots A_k^{n_k} B, \ldots, A_1^{n-1} \ldots A_k^{n-1} B) = n, \\ &\operatorname{rang}(C^*, \ldots, A_1^{*n_1} \ldots A_k^{*n_k} C^*, \ldots, A_1^{*n-1} \ldots A_k^{*n-1} C^*) = n \end{aligned} \tag{3}$$

algebrai feltételeket látjuk be rendre az (1) irányíthatóságára és megfigyelhetőségére. Ezeket a feltételeket Kalman-féle többváltozós rangfeltételeknek tekinthetjük. A [2]-ben szereplő ellenpéldák szerint a (3) alatti feltételek további feltételekkel kiegészítve az (1) irányíthatóságát és megfigyelhetőségét jellemzik időtől függő $a_1, \ldots, a_k$ együtthatók esetén. Ezek a feltételek szoros kapcsolatban állnak az L algebrai tulajdoságaival; ezek lényegében csak az

$$[A_i, A_j] = \sum_{\ell=1}^{k} \Gamma_{ij}^{\ell} A_\ell \tag{4}$$

multiplikációs táblától függnek, ahol a Lie-zárójelet az $A_iA_j - A_jA_i$ kommutátor definiálja.

Szigeti [3]-ban a

$$\begin{aligned} &\partial_t u(t,x) + \sum_{i=1}^{m} f_i(t,x) \partial_{x_i} u(t,x) = A(t,x)u(t,x) + Bv(t,x), \\ &y(t,x) = Cu(t,x) \end{aligned} \tag{5}$$

parciális differenciálegyenlet-rendszer alapmátrixának exponenciálisok szorzataként való reprezentációjával kapcsolatban egy, a [4]-ben szereplőhöz hasonló eredményt bizonyít. Így az (5) rendszer irányíthatóságára és megfigyelhetőségére ugyanazokat a rangfeltételeket kapjuk, mint az (1) véges dimenziójú rendszer esetében. Megjegyezzük, hogy mind a véges, mind a végtelen dimenziójú esetekben ugyanazok az algebrai meggondolások alkalmazhatóak (annak ellenére, hogy az (5) rendszer végtelen dimenziójú).

2. Előzmények

Tekintsük az (1) véges dimenziójú rendszert! Az $A(t)$ struktúramátrixot állítsuk elő a (2) szerinti lineáris kombinációként! Most a rendszer Φ alapmátrixának néhány tulajdonságát adjuk meg.

A [4] szerint a

$$\dot{\Phi} = A(t)\Phi(t) \qquad \Phi(0) = I$$

mátrix-differenciálegyenlet megoldását megadhatjuk exponenciálisok szorzataként a

$$\Phi(t) = e^{A_1 g_1(t)} e^{A_2 g_2(t)} \ldots e^{A_k g_k(t)}$$

alakban, ahol $g = (g_1, g_2, \ldots, g_k)$ a

$$\dot{g} = \left(\sum_{i=1}^{k} e^{\Gamma_1 g_1} e^{\Gamma_2 g_2} \ldots e^{\Gamma_{i-1} g_{i-1}} E_{ii} \right)^{-1} a, \qquad g(0) = 0 \tag{6}$$

közönséges differenciálegyenlet megoldása. E_{ii} egy $(0,1)$-mátrix, amelynek egyetlen nem nulla eleme az i. diagonális elem, a Γ_i-ket a multiplikációs tábla definiálja, vagy ami ezzel ekvivalens, P_i az az $A_i : L \to L$ lineáris leképezés mátrix reprezentációja az $A_1 \ldots A_k$ bázis szerint, végül $a = (a_1, \ldots, a_k)$ (lásd. [4]). Ez a reprezentáció általában lokális. Ha az L Lie-algebra feloldható, akkor az

$$M(g) = \sum_{i=1}^{k} e^{\Gamma_1 g_1} \ldots e^{\Gamma_{i-1} g_{i-1}} E_{ii} \tag{7}$$

mátrix invertálható, és a reprezentáció globális. Ismert, hogy az (1) rendszer irányíthatósága és megfigyelhetősége rendre ekvivalens a G_c és a G_o Gram-mátrixok invertálhatóságával. Ennélfogva az (1) rendszer akkor és csak akkor nem irányítható és nem megfigyelhető a $[0,T]$ felett, ha léteznek $\xi, \eta \in \mathbb{R}^n$, $\xi, \eta \neq 0$ vektorok, amelyek kielégítik az

$$\begin{aligned} \langle u, B^* \Phi(t)^{*-1} \xi \rangle = 0 \quad &\text{és az} \\ \langle y, C\Phi(t)\eta \rangle = 0 & \end{aligned} \tag{8}$$

egyenlőségeket rendre minden $t \in [0,T]$, $u \in \mathbb{R}^r$ és $y \in \mathbb{R}^s$ esetén. Végül megjegyezzük, hogy a

$$\dot{\Phi}(t) = \left(\sum_{i=1}^{k} a_i(t) A_i \right) \Phi(t)$$

differenciálegyenlet adjungáltja a következő formában kapható

$$\left(\Phi(t)^{*-1}\right)^{\cdot} = -\left(\sum_{i=1}^{k} a_i(t) A_i^* \right) \Phi(t)^{*-1}. \tag{9}$$

3. A fő eredmények

Ezt a szakaszt egy olyan lemma bizonyításával kezdjük, amelynek fontos szerepe lesz a főtétel bizonyításában. A második lemma tulajdonképpen egy egyszerű algebrai számítás, amely arra szolgál, hogy a főtételt a Kalman-féle rangfeltétellel analóg formában mondhassuk ki.

1. LEMMA. *Legyen* $a = (a_1, \dots, a_m, 0, \dots, 0) : [0,T] \to \mathbb{R}^k$ *egy analitikus függvény az alábbi tulajdonságokkal:*

a) a (6) *differenciálegyenletnek van egy* g *globális megoldása a* $[0,T]$ *intervallumon,*

b) *minden* $(\alpha,\beta) \subset [0,T]$ *részintervallumhoz létezik egy* $i \in \{1,2,\dots,k\}$ *és egy* $(\alpha_i,\beta_i) \subset (\alpha,\beta)$ *úgy, hogy* g_i' *nem tűnik el az* (α_i,β_i) *intervallumon,*

c) *nincs a* $g_1, g_2, \dots, g_k$ *között analitikus függőség, azaz ha* $i \neq j$*-re a* $g_i = F(g_j)$ *egyenlőség fennállna, akkor az* F *nem analitikus.*

Tegyük fel, hogy minden $t \in [0,T]$ *és valamely* $u \in \mathbb{R}^r$, $y \in \mathbb{R}^s$, $\xi, \eta \in \mathbb{R}^n$ *esetén a* (8) *egyenlőségek fennállnak.*

Ekkor minden $t \in [0,T]$ *és* $i_1, i_2, \dots, i_\ell \in \{1,2,\dots,m\}$*-re fennállnak az*

$$\begin{aligned} \langle u, B^* A_{i_1}^* \dots A_{i\ell}^* \Phi(t)^{*-1} \xi \rangle &= 0, \\ \langle y, C A_{i_1} \dots A_{i\ell} \Phi(t) \eta \rangle &= 0 \end{aligned} \tag{10}$$

egyenlőségek is.

Bizonyítás. A lemmát az ℓ szerinti indukcióval látjuk be. Az $\ell = 0$ esetén az eredeti (8) egyenlőséget kapjuk. Tegyük fel, hogy a (10) fennáll bizonyos $\ell \geq 0$ és $i_1, i_2, \dots i_\ell \in \{1,2,\dots,m\}$ esetén. Ekkor a (10) differenciálásával és a (9) és (6) egyenletek felhasználásával a

$$\begin{aligned} &\sum_{i=1}^{k} \dot{g}_i \sum_{j=1}^{m} \langle e_j, e^{g_1 \Gamma_1} \dots e^{g_{i-1}\Gamma_{i-1}} e_i \rangle \langle u, B^* A_{i_1}^* \dots A_{i\ell}^* A_j^* \Phi(t)^{*-1} \xi \rangle = 0, \\ &\sum_{i=1}^{k} \dot{g}_i \sum_{j=1}^{m} \langle e_j, e^{g_1 \Gamma_1} \dots e^{g_{i-1}\Gamma_{i-1}} e_i \rangle \langle y, C A_{i_1} \dots A_{i\ell} A_j \Phi(t) \eta \rangle = 0 \end{aligned} \tag{11}$$

egyenlőségeket kapjuk. A bizonyítást elég csak az első egyenletre elvégezni, mivel a második esetben hasonló okoskodással élhetünk.

Tegyük most fel, hogy létezik egy $\langle u, B^* A_{i_1}^* \dots A_{i\ell}^* A_j^* \Phi(t)^{*-1} \xi \rangle$ nem nulla tag! Ekkor létezik egy $(\alpha,\beta) \subset [0,T]$ intervallum úgy, hogy minden $t \in (\alpha,\beta)$ esetén nem áll fenn az egyenlőség, azaz

$$\langle u, B^* A_{i_1}^* \dots A_{i\ell}^* A_j^* \Phi(t)^{*-1} \xi \rangle \neq 0.$$

A (7)-ben szereplő $M(g)$ mátrix invertálhatóságából ((a) tulajdonság) az következik, hogy a (11)-ben legalább egy $\dot{g}_i$ koefficiens nem nulla. Tegyük fel, hogy

$$\varphi_p(g) = \sum_{j=1}^{k} \langle e_j, e^{g_1 \Gamma_1} \dots e^{g_{p-1}\Gamma_{p-1}} e_p \rangle \langle u, B^* A_{i_1}^* \dots A_{i\ell}^* A_j^* e^{-A_1^* g_1} \dots e^{-A_k^* g_k} \xi \rangle \neq 0$$

az (α, β) intervallumon. Ekkor a b) tulajdonság alapján létezik egy $(\alpha_i, \beta_i) \subset (\alpha, \beta)$ úgy, hogy $\dot{g}_i$ nem tűnik el ezen az (α_i, β_i) intervallumon.

Ha $i = p$, akkot az i. tagra $\dot{g}_i \varphi_i(g) \neq 0$ az (α_i, β_i) intervallumban. Ezért létezik legalább egy másik $\dot{g}_q \varphi_q(g)$ tag is, amely különbözik nullától egy esetleg kisebb $(\alpha', \beta') \subset (\alpha_i, \beta_i)$ intervallumon.

Ha $i \neq p$, akkor legyen $q := i$. Így a

$$\frac{dx}{dg_q} = \sum_{\substack{i=1 \\ i \neq p,q}}^{k} \frac{dg_i \cdot q_q^{-1}}{dg_q} \frac{\varphi_i(\bar{g})}{\varphi_p(\bar{g})} + \frac{\varphi_q(\bar{g})}{\varphi_p(\bar{g})}$$

x-re vonatkozó differenciál egyenlet nem triviális, analitikus a g_q független változó ismeretlen függvényére, ahol $\bar{g}$ a

$$(g_1 \circ g_q^{-1}, \ldots, \underset{p}{\check{x}}, \ldots, \underset{q}{\check{g}_q}, \ldots, q_k \circ g_q^{-1})$$

vektort jelöli.

A $t \longleftrightarrow g_q$ időtranszformáció definiálható, mivel a $\dot{g}_q(t)$ derivált nem tűnik el az (α', β') intervallumon; az a analitikus voltából pedig a differenciálegyenlet jobb oldalának analitikussága következik. Kezdeti feltételnek tetszőleges $\rho_0 = g_q(t_0)$, $t_0 \in (\alpha', \beta')$ pontot választhatunk:

$$x(t_0) = g_p(t_0).$$

Ekkor az x megoldás analitikus és $x = g_p \circ g_q^{-1}$ az unicitás miatt. Ezért $g_p = x \circ g_q$, ami azt jelenti, hogy analitikus függőség van a g_p és a g_q között, s ez ellentmond a c) feltevésnek. Ezzel a lemmát bebizonyítottuk.

2. LEMMA. *A (10) alatti egyenletek családja*
$\ell = 0, 1, \ldots i_1, i_2 \ldots i_\ell \in \{1, 2, \ldots, m\}$*-re ekvivalens az*

$$\langle u, B^* A_k^{*n_k} \ldots A_1^{*n_1} \Phi(t)^{*-1} \xi \rangle = 0,$$
$$\langle u, C A_1^{n_1} \ldots A_k^{n_k} \Phi(t) \eta \rangle = 0$$

egyenletek családjával minden $t \in [0, T]$, $0 \leq n_1, n_2, \ldots, n_k$ *esetén.*

Bizonyítás. Egészítsük ki az $A_1, A_2, \ldots A_m$-et az L $A_1, A_2, \ldots, A_k$ bázisává. Az L Lie-algebrát $A_1, \ldots, A_m$ generálja, ezért léteznek olyan
$L_1 = \left[A_{i_1^1}, \left[A_{i_2^1} \ldots \left[A_{i_{n_1-1}^1}, A_{i_{n_1}^1}\right] \ldots\right]\right] \ldots$

$$L_j = \left[A_{i_1^j}, \left[A_{i_2^j}, \ldots, \left[A_{i_{n_{j-1}}^j}, A_{i_{n_j}^j}\right] \ldots\right]\right], \ldots, \qquad j = 1, \ldots, k - m$$

Lie-elemek, hogy $A_{m+j} = L_j$ és $A_{i_k^j} \in \{A_1, \ldots, A_m\}$. Így, ha a (10) egyenletek fennállnak, akkor $i_1, \ldots, i_\ell \in \{1, 2, \ldots, m\}$-re

$$\langle u, B^* A_{i_1}^* \ldots A_{i_h}^* A_{m+j}^* A_{i_{h+1}}^* \ldots A_{i_\ell}^* \Phi(t)^{*-1} \xi \rangle = 0.$$

Valóban, az A_{m+j} kifejezhető az L_j Lie-elemmel. Ezért a fenti kifejezés felírható a (10) alakú tagok összegeként. A bizonyítást most az $A_{m+1}, \dots, A_k$ tagok teljes fokára vonatkozó indukcióval folytatjuk. Legyen

$$\langle u, B^*, A^*_{\ell_1} \dots A^*_{m+j_1} \dots A^*_{m+j_2} \dots A^*_{m+j_p} \dots A^*_{i\ell} \Phi(t)^{*-1} \xi \rangle \tag{13}$$

olyan, hogy $i_1, \dots, i_\ell \in \{1, 2, \dots, m\}$, $j_1, \dots, j_p \in \{1, 2, \dots, k-m\}$. Tegyük fel, hogy $p-1$, $p \geq 2$ esetén a (13) minden tagja egyenlő nullával! Ha ekkor az A_{m+j_1}-eket L_{j_1} Lie-elemekkel helyettesítjük, akkor a (13)-at az $A_{m+1}, \dots, A_k$ teljes fokainak összegeként kapjuk, amely $p-1$-gyel egyenlő. Így a (13)-nak egyenlőnek kell lennie nullával. A bizonyítás e részének megfordítása nyilvánvaló, mivel a (10) speciális esete a (13)-nak. A megfigyelhetőségre vonatkozó tagokra a bizonyítás hasonló módon történik. A második lemma összes egyenlete a (13) speciális esete. Így a (13)-at a (12)-ből kell belátnunk. Az állításunknak ez a része ekvivalens az $i_1 \leq i_2 \leq \dots \leq i_\ell$, $i_1, \dots, i_e \in \{1, 2, \dots, k\}$ indexek monotonicitásával. Ha $\ell = 1$, akkor az állítás triviális. Tegyük fel, hogy $\ell - 1$, $\ell \geq 2$, a (13) következik a (12)-ből és azt, hogy nem teljesül az indexek monotonitása! Ekkor léteznek olyan i_j, i_{j+1} indexek, hogy $i_j > i_{j+1}$. Az

$$A_{i_j} A_{i_{j+1}} = A_{i_{j+1}} A_{i_j} + \sum_{\ell=1}^{k} \Gamma^{\ell}_{i_j i_{j+1}} A_\ell,$$

reláció és az indukciós feltevés alapján

$$\begin{aligned} &\langle u, B^* A^*_{i_1} \dots A^*_{i_j} A^*_{i_{j+1}} \dots A^*_{i\ell} \Phi(t)^{*-1} \xi \rangle = \\ &= \langle u, B^* A^*_{i_1} \dots A^*_{i_{j+1}} A^*_{i_j} \dots A^*_{i\ell} \Phi(t)^{*-1} \xi \rangle. \end{aligned}$$

Az egymást követő tagokat felcserélve a növekvő index reprezentációt kapjuk. Ezzel a lemmát bebizonyítottuk.

Megjegyzés. Nyilvánvaló, hogy a

$$\begin{aligned} B^* A_1^{*n_1} \dots A_k^{*n_k} \xi = B^* A_k^{*n_k} \dots A_1^{*n_1} \xi = 0, \\ C A_1^{n_1} \dots A_k^{n_k} \eta = C A_k^{n_k} \dots A_1^{n_1} \eta = 0 \end{aligned} \tag{14}$$

egyenletek a (10)-ből következnek.

TÉTEL. *Legyen* $a_i : [0, T] \to \mathbb{R}$, $i = 1, 2, \dots, m$, $A_1, \dots, A_m \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times r}$, $C \in \mathbb{R}^{s \times n}$. *Tegyük fel, hogy a analitikus, és azt, hogy az* $A_1, \dots, A_m$ *által generált* L *Lie-algebra* $A_1, \dots, A_m, A_{m+1}, \dots, A_k$ *bázisa kielégíti a* (4) *multiplikációs táblát. Ha* $a = (a_1, \dots, a_m, 0, \dots, 0)$ *kielégíti az első lemma* a), b) *és* c) *feltételeit, akkor az*

$$\begin{aligned} \dot{x}(t) &= \left(\sum_{i=1}^{m} a_i(t) A_i \right) x(t) + B u(t), \\ y(t) &= C x(t) \end{aligned}$$

rendszer irányíthatósága és megfigyelhetősége rendre ekvivalens a (3) *alatti rangfeltételekkel.*

Bizonyítás. Elegendő azt bebizonyítani, hogy az állítás igaz az irányíthatóságra, mivel a megfigyelhetőséget vagy analóg módon, vagy a dualitás segítségével láthatjuk be.

A rangfeltétel szükségességét a következőképpen mutathatjuk meg. Ha a rang kisebb mint n, akkor létezik olyan $\xi \in \mathbb{R}^n$, hogy

$$0 = B^*\xi, \ldots, 0 = B^*A_1^{*n_1} \ldots A_k^{*n_k}\xi, \ldots, 0 = B^*A_1^{*n-1} \ldots A_k^{*n-1}\xi.$$

Ezért a Cayley-Hamilton tételből következik, hogy $B^*A_1^{*n_1}, \ldots, A_k^{*n_k}\xi = 0$ minden $n_1, n_2, \ldots n_k \geq 0$-ra. Ezért

$$B^*\Phi^{*-1}(t)\xi = B^*e^{-A_1^*g_1(t)} \ldots e^{-A_k^*g_k(t)}\xi =$$

$$\sum_{n_1,\ldots,n_k=0}^{\infty} \frac{(g_1(t))^{n_1}}{n_1!} \ldots \frac{(-g_k(t))^{n_k}}{n_k!} B^*A_1^{*n_1} \ldots A_k^{*n_k}\xi = 0 .$$

Minden $t \in [0, T]$, ami ekvivalens a rendszer nem-irányíthatóságával.

A tétel elégségességét az alábbi módon láthatjuk be. Elegendő bebizonyítani azt, hogy a rangfeltétel következik a (8)-ból. Az első lemma alapján, ha létezik olyan $\xi \neq 0$, hogy a (8) fennáll, akkor a (10)-ből és a második lemmából következik, hogy a (12) fennáll. Ha ez utóbbiba a $t = 0$ értéket helyettesítjük, akkor az eredményül kapott relációból az következik, hogy a rangfeltétel nem teljesül. Ezzel a tétel bizonyítását befejeztük

A következő példa azt mutatja, hogy az

$$\dot{x}(t) = (A_1a_1(t) + A_2a_2(t))x(t) + bu(t)$$

rendszernek az a_1 és a_2 koefficiensek közti speciális viszonytól függően különböző rangfeltételei lehetnek. Legyenek

$$A_1 = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} .$$

Először is megjegyezzük, hogy $[A_1, A_2] = A_2$. Valamely $b \in \mathbb{R}^4$-re a rangfeltétel akkor és csak akkor teljesül, ha $b_4 \neq 0$.

a) Ha $a = a_1 = a_2 \neq 0$, akkor ennek a speciális rendszernek az alapmátrixa:

$$\Phi(t) = e^{(A_1+A_2)\int_0^t a} .$$

Ennélfogva az irányíthatóság ekvivalens a $b_4 \neq 0$ és $b_2 + b_4 \neq 0$ feltétellel, amely nyilvánvalóan különbözik a Kalman-féle rangfeltételtől.

b) Ha $a_2 = a_1 e^{\int_0 a_1}$, akkor az alapmátrix:

$$\Phi(t) = e^{A_1 \int_0 a_1} e^{A_2 \int_0 a_1}.$$

Így a rendszer irányíthatósága ekvivalens az előző esetben látott $b_4 \neq 0$ és $b_2 + b_4 \neq 0$ feltétellel.

Megjegyezzük, hogy

$$e^{(A_1+A_2) \int_0 a_1} \neq e^{A_1 \int_0 a_1} e^{A_2 \int_0 a_1},$$

bár az irányíthatóság feltétele ugyanaz maradt.

IRODALOM

[1] KALMAN, R. E., FALB, P.L. és ARBIB, M.A., *Topics in Mathematical Systems Theory* (McGraw-Hill Book Company, New York, 1969).

[2] SZIGETI, F. és VERA, C. E., „Kalman's type rank conditions for certain time dependent systems", *Proc. Conf. MTNS'91 Kobe* (Japan MITA Press, Tokyo).

[3] SZIGETI, F., „On the product of exponentials as solutions of first order linear partial differential equations.", *PUMA*, Ser B, 1 (1990), 169–178.

[4] WEI, J., és NORMAN, E., „On global representation of the solution of linear differential equations as a product of exponentials", *Proc. Amer. Math. Soc.* **15** No. 12 (1964), 327–334.

(Beérkezett: 1992. április 25.)

MOLNÁR SÁNDOR
KÖZPONTI BÁNYÁSZATI FEJLESZTÉSI INTÉZET
RENDSZERELEMZÉSI ÉS GEOMATEMATIKAI OSZTÁLY BUDAPEST III.
MIKOVINY S. U. 2-4
H-1037
SZIGETI FERENC
EÖTVÖS LÓRÁND TUDOMÁNYEGYETEM
ALKALMAZOTT ANALÍZIS TANSZÉK
BUDAPEST, MÚZEUM KRT. 6-8 1088
VERA CARMEN E.
CENTRAL UNIVERSITY OF VENEZUELA
DEPARTMENT OF MATHEMATICS CARACAS
VENEZUELA

KALMAN'S RANK CONDITIONS FOR TIME DEPENDENT LINEAR SYSTEMS

S. MOLNÁR, F. SZIGETI and C.E. VERA

The well-known Kalman's algebraic rank conditions are generalized for time-varying linear systems. The terms of the multivariable rank conditions are the elements of the basis of the Lie algebra, generated by the structure matrices of the system.

Alkalmazott Matematikai Lapok **17** (1993) 287–292

DINAMIKUS FOLYAMATOK KONVERGENCIÁJA ÉS STABILITÁSA

MOLNÁR SÁNDOR ÉS SZIDAROVSZKY FERENC

Budapest, Tucson

A dolgozatban diszkrét és folytonos idejű algoritmikus modellek konvergenciájára vonatkozó általános feltételeket adunk meg. Eredményeink az iterációs módszerek konvergenciájára vonatkozó korábbi eredmények, valamint a differencia- és differenciálegyenletek klasszikus stabilitási tételeinek az általánosításai.

1. Bevezetés

Az elmúlt két évtizedben számos szerző foglalkozott az iterációs folyamatok általános elméletével. ZANGWILL (1969), POLAK (1971), TISHYADHIGAMA és szerzőtársai a leginkább használatos konvergenciafeltételek átfogó összegzését adják, s a feltételek összehasonlítását és az optimumszámításban való alkalmazásukat mutatják be. Az általános elmélet a megoldástéren definiált monoton függvényen alapszik, amely a differenciaegyenletek ismert Ljapunov-féle stabilitáselméletét (LA SALLE, 1986) általánosítja. A differencia- és a differenciálegyenletek megoldásainak aszimptotikus viselkedése közötti analógia közismert. Itt mi a stabilitás és az aszimptotikus stabilitás feltételei, valamint az algoritmikus modellek konvergenciája (POLAK 1971) és differenciálegyenletek kvázistabilitása, illetve ω-határpontjai közti analógiára utalunk (UZAWA 1961, HIRSCH és SMALE 1974).

Az alábbiakban először általános dinamikus folyamatokra egy új konvergenciatételt bizonyítunk be, amely a diszkrét és folytonos idejű modellekre egymástól függetlenül kapott korábbi eredmények közönséges általánosításának tekinthető. Ezután az iterációs módszerekben és a differenciálegyenletek stabilitáselméletében való néhány alkalmazást tárgyalunk. Végül összefoglaljuk a kapott eredményeket, és azok következményeit.

2. A főtétel

Tegyük fel, hogy X egy Hausdorff-féle topológikus tér, amely kielégíti az első megszámlálhatósági axiómát. Tegyük fel továbbá, hogy I a $[0, \infty)$ intervallum nem korlátos részhalmaza.

A kutatást a Magyar Amerikai Közös Alap támogatta (JF No. 224).

Definíció. Függvények egy F halmazát *dinamikus folyamatnak* nevezzük, ha minden $f \in F$ esetén $D(f) = I$ és $R(f) \subseteq X$. Az F halmaz elemeit *trajektóriának* nevezzük.

Legyen $S \subset X$ egy adott halmaz, amelyet a továbbiakban *az alkalmas pontok halmazának* fogunk nevezni.

Definíció. Az F dinamikus folyamatot *konvergensnek* nevezzük, ha $f \in F$, $t_1 < t_2 < \dots$ $(t_i \in I,\ i = 1, 2, \dots)$, $t_i \to \infty$ és $f(t_i) \to f^*$, akkor $f^* \in S$.

Tegyük most fel, hogy Y szintén egy Hausdorff-féle topológikus tér, amely kielégíti az első megszámlálhatósági axiómát és legyen $\leqq$ egy parciális rendezés az Y téren. Feltesszük továbbá, hogy a topológia és a parciális rendezés olyan kapcsolatban áll, melyet az alábbi tulajdonság ad meg:

(P) Ha $y_1 \geqq y_2 \geqq \cdots$ $(y_i \in Y,\ i = 1, 2, \dots)$ és $y_i \geqq y$ $(i = 1, 2, \dots)$ valamely $y \in Y$ elemre, akkor az $\{y_i\}$ sorozat egy $y^* \in Y$ elemhez konvergál úgy, hogy $y_i \geqq y^*$ minden i esetén.

Definíció. A $V : I \times X \to Y$ függvényt *általánosított Ljapunov-függvénynek* nevezzük, ha teljesülnek az alábbi feltételek:

(i) Nagy t értékekre a V függvények alulról *egyenletesen lokálisan korlátosak* az $X \setminus S$ halmazon, azaz létezik egy Q_1 nem negatív szám úgy, hogy minden $z \in X \setminus S$ elemnek létezik egy U környezete és egy $y \in Y$ elem (amely függhet z-től), hogy minden $t \geqq Q_1$ $(t \in I)$ és $z' \in U$ esetén

$$V(t, z') \geqq y;$$

(ii) Ha $f \in F$ és $Q_1 \leqq t < t'$ $\quad (t, t' \in I)$, akkor

$$V(t', f(t')) \leqq V(t, f(t));$$

(iii) Minden $z^* \in X \setminus S$ elem esetén, ha a $\{z_i\} \subset X$ olyan, hogy $z_i \to z^*$, és $\{t_i\} \subset I$ egy szigorúan növekedő sorozat, úgy, hogy $t_i \to \infty$, továbbá $V(t_i, z_i) \to y^*$, akkor minden $f \in F$ trajektóriához, amelyre $f(t_i) = z_i$ $(i = 1, 2, \dots)$, létezik egy $t \in I$ úgy, hogy $t \geqq Q_1$ és $V(t, f(t)) < y^*$.

TÉTEL. *Ha az F dinamikus folyamatnak van általánosított Ljapunov-függvénye, akkor az F konvergens.*

Bizonyítás. Tegyük fel, hogy létezik egy $f \in F$ úgy, hogy $f(t_i) \to f^* \in X \setminus S$ valamely szigorúan növekvő $\{t_i\} \subset I$, $t_i \to \infty$ sorozattal. A (ii) feltételből tudjuk, hogy a $\{V(t_i, f(t_i))\}$ sorozat nagy i értékekre csökkenő, és az (i) feltevésből az következik, hogy a $V(t_i, f(t_i))$ alulról korlátos, ha i elég nagy. A rendezésre vonatkozó (P) feltételből ezért az következik, hogy a $\{V(t_i, f(t_i))\}$ sorozat egy $y^* \in Y$ határértékhez konvergál, továbbá a (ii) feltétel ismételt felhasználásával látható, hogy

$$V(t, f(t)) \geqq y^* \tag{1}$$

minden $t \geqq Q_1 \quad (t \in I)$.

Ha most a (iii) feltételbe $\{z_i\}$ helyett az $\{f(t_i)\}$ sorozatot írjuk, akkor

$$V(t, f(t)) < y^*$$

valamely $t \geqq Q_1$ értékre, s ez ellentmond (1)-nek. Ezzel a tétel bizonyítását befejeztük.

Következmény. Tegyük fel, hogy bármely trajektória egy X-beli kompakt halmazban van, továbbá azt, hogy S csak egy x^* pontból áll. Ekkor a tétel feltételei mellett minden $f \in F$ trajektóriára $f(t_i) \to x^*$ ha $t_i \in I$ $(i = 1, 2, \dots)$ és $t_i \to \infty$, azaz az F dinamikus folyamat globálisan aszimptotikusan stabil.

3. Alkalmazások

1. Tekintsük először azt az esetet, amikor $I = \{0, 1, 2, \dots\}$, és ahol minden trajektóriát az

$$f(i+1) \in A_i(f(i)) \quad (i \geqq 0) \tag{2}$$

algoritmikus modell generál, ahol $A_i : X \to 2^X$ egy pont–halmaz leképezés minden $i = 0, 1, 2, \dots$ esetén, és $f(0) \in X$ a kezdőpont. Tegyük fel továbbá, hogy fennáll az (i) és az alábbi két feltétel:

(ii_1) Ha $i \geqq Q_1$ és $x' \in A_i(x)$ $(x \in X)$, akkor

$$V(i+1, x') \leqq V(i, x).$$

(iii_1) Minden $z^* \in X \setminus S$ elem esetén, ha a $\{z_i\} \subset X$ olyan, hogy $z_i \to z^*$ és $\{k_i\} \subset I$ egy szigorúan növekedő sorozat úgy, hogy $V(k_i, z_i) \to y^*$, akkor létezik egy j egész szám, hogy $k_j \geqq Q_1 - 1$ és minden $y \in A_{k_j}(z_j)$ elemre

$$V(k_j + 1, y) < y^*.$$

Könnyű belátni, hogy a (ii_1) és (iii_1) feltételekből következik a (ii) és a (iii). Következésképpen az (i), (ii_1) és (iii_1) elegendőek a (2) algoritmikus modell konvergenciájához. Megjegyezzük, hogy ez az eredmény Tishyadhigama és szerzőtársai (1979) 4.2. tételét, illetve Higle és Sen (1989) legutóbbi eredményeit általánosítja.

Tekintsük most azt a speciális esetet, amikor a V nem függ az i-től. Ekkor az alábbi feltételek lesznek elegendőek a (2)-beli algoritmikus modellek konvergenciájához:

(i_2) V alulról lokálisan korlátos az $X \setminus S$ halmazon.

(ii_2) Létezik egy Q_1 egész szám úgy, hogy $i \geqq Q_1$ és

$$x' \in A_i(x) \ (x \in X) \text{ esetén } V(x') \leqq V(x).$$

(iii$_2$) Minden $z^* \in X \setminus S$ elem esetén, ha a $\{z_i\} \subset X$ olyan, hogy $z_i \to z^*$ és $V(z_i) \to y^*$, akkor létezik egy $j \geqq Q_1$ egész szám úgy, hogy $V(y) < y^*$ minden $y \in A_i(x_j)$ és $i \geqq j$ esetén.

Egy további speciális esetben, amikor a (2) stacionárius, a (ii$_2$) és (iii$_2$) feltételek az alábbiaknak megfelelően módosíthatók:

(ii$_3$) Minden $x \in X$ és $x' \in A(x)$ esetén $V(x') \leqq V(x)$;

(iii$_3$) Minden $z^* \in X \setminus S$ elem esetén ha a $\{z_i\} \subset X$ olyan, hogy $z_i \to z^*$ és $V(z_i) \to y^*$, akkor létezik egy $j \geqq 0$ egész szám úgy, hogy $V(y) < y^*$ minden $y \in A(z_j)$ esetén. A fenti konvergenciafeltételeket például az algebrai egyenletek (Szidarovszky és Yakowitz 1978) és az optimumszámítási problémák (Zangwill 1969, Polak 1971) megoldására szolgáló iterációs folyamatok konvergenciaanalízisében is lehet alkalmazni. Az az általános feltevés, hogy a V nem szükségképpen valós értékű, lehetővé teszi a többcélú programozásban való alkalmazást is, ahol a különböző döntési alternatívákat a következménytéren definiált valamilyen preferenciarendezés segítségével hasonlítjuk össze. Ebben az esetben az Y teret következménytérnek választhatjuk. Ez egy igen alkalmas választás azokban az esetekben, amikor a preferenciarendezést nem lehet valamely valós értékű függvénnyel jellemezni (Fishburn 1970).

2. Tekintsük most azt az esetet, amikor $I = [0, \infty)$, $X \subset \mathbb{R}^n$ és a trajektóriákat az alábbi differenciálegyenlet generálja:

$$\frac{d}{dt} f(t) = A(t, f(t)) \tag{3}$$

ahol $A : I \times X \to X$ valamilyen függvény. Tegyük fel, hogy minden $x_0 \in X$ és $t_0 \in I$ esetén a (3) egyenletnek van legalább egy megoldása úgy, hogy $f(t_0) = x_0$, továbbá ezek a megoldások minden $t \in I$ esetén definiálva vannak, és az X halmazban maradnak. Jelölje ezeknek a megoldásoknak a halmazát $F(t_0, x_0)$. Legyen továbbá

$$F = \bigcup_{\substack{t_0 \in I \\ x_0 \in X}} F(t_0, x_0).$$

Könnyű belátni, hogy a (iii) feltételt most az alábbiak szerint módosíthatjuk:

(iii$_4$) Minden $z^* \in X \setminus S$ elem esetén, ha a $\{z_i\} \subset X$ olyan, hogy $z_i \to z^*$ és $\{t_i\} \subset I$ egy szigorúan növekedő sorozat úgy, hogy $t_i \to \infty$, továbbá $V(t_i, z_i) \to y^*$, akkor minden $f \in \bigcap_i F(t_i, z_i)$ trajektóriához létezik egy $t \geqq Q_1$ úgy, hogy $V(t, f(t)) < y^*$. Tegyük fel, hogy az A és a V függvények nem függnek explicit módon a t-től, és minden $x_0 \in X$ kezdőpontra az $F(0, x_0)$ csak az $f(t, x_0)$ trajektóriát tartalmazza. Ebben az esetben a (iii$_4$)-et a következő feltétel implikálja:

(iii$_5$) Minden $z^* \in X \setminus S$ elem esetén, ha a $\{z_i\} \subset X$ olyan, hogy $z_i \to z^*$ és $V(z_i) \to y^*$, akkor létezik egy $j > 0$ úgy, hogy valamely $t > 0$ esetén $V(f(t, z_j)) < y^*$.

A következőkben a fenti feltételeknek Uzawa híres stabilitási eredményeivel (1961) való kapcsolatát elemezzük.

Tegyük fel ismét, hogy A és V nem függ explicit módon a t-től, továbbá azt, hogy minden $x_0 \in X$ kezdőpontra az $F(0, x_0)$ pontosan egy, az $f(t, x_0)$ trajektóriát tartalmazza. Tegyük fel továbbá, hogy

(a) V csökkenő az S, és szigorúan csökkenő az $X \setminus S$ halmazon,

(b) V folytonos az X halmazon,

(c) $f(t, x_0)$ folytonosan függ az x_0 kezdőponttól minden $t \geqq 0$ esetén.

Most belátjuk, hogy az (a)–(c) feltételek az (i), (ii) és (iii_5)-öt implikálják, azaz hogy a fenti eredményeket Uzawa klasszikus tételének (1961) közvetlen kiterjesztésének tekinthetjük. V folytonosságából következik (i), míg a (ii) az (a)-ból következik.

Tegyük most fel, hogy $y^* \in X \setminus S$, akkor (a)-ból az következik, hogy minden $t \geqq 0$ esetén $V(f(t, y^*)) < V(y^*)$. Ekkor V és f folytonosságából adódik, hogy $V(f(t, z_j)) < V(y^*)$, ha z_j elég közel van y^*-hoz, így (iii_5) fennáll.

4. Következtetések

A fentiekben a konvergenciára és a stabilitásra vonatkozó néhány klasszikus eredményt általánosítottunk úgy, hogy azok a legtöbb diszkrét és folytonos idejű dinamikus folyamatot magukba foglalják. Elemzésünket az általánosított Ljapunov függvényekre alapoztuk, amelyekről nem kell feltétlenül feltennünk a folytonosságot, csak a teljes megoldástéren való monotonitást. A folytonosságot a Ljapunov függvények egyenletes lokális korlátosságával, a szigorú monotonitást pedig az általunk megadott gyengébb (iii) feltétellel helyettesítjük.

Eredményeinket az algebrai egyenletek és az egy- és többcélú programozási problémák megoldásában, továbbá a diszkrét és folyotonos idejű dinamikus rendszerek stabilitásának vizsgálatában lehet alkalmazni.

IRODALOM

[1] FISHBURN, P. C., *Utility Theory for Decision Making* (J. Wiley & Sons, New York/London, 1970).

[2] HIGLE J. and SEN, S., „On the Convergence of Algorithms with Applications to Stochastic and Nondifferentiable Optimization", *Working Paper* 89–027 (SIE Dept, University of Arizona, Tucson, Arizona, 1989).

[3] HIRSCH, M. W. and SMALE, S., *Differential Equations, Dinamical Systems and Linear Algebra* (Academic Press, New York, 1974).

[4] LA SALLE, J. P., *The Stability and Control of Discrete Processes* (Springer-Verlag, New York, Berlin, 1986).

[5] POLAK, E., *Computational Methods in Optimization. A Unified Approach* (Academic Press, New York, 1971).

[6] SZIDAROVSZKY, F. and YAKOWITZ, S., *Principles and Procedures of Numerical Analysis* (Plenum, New York/London., 1978).

[7] TISHYADHIGAMA, S., POLAK, E. and KLESSING, R., „A Comparative Study of Several General Convergence Conditions for Algorithms Modeled by Point-to-Set Maps", *Math. Programming Study* 10 (1979), 172–190.

[8] UZAWA, H, „The stability of Dynamic Processes", *Econometrica* **29** (1961), 617–631.
[9] ZANGWILL, W. I., *Nonlinear Programming: A Unified Approach* (Prentice-Hall, Englewood Cliffs, N.J., 1969).

(Beérkezett: 1992. április 25.)

MOLNÁR SÁNDOR
KÖZPONTI BÁNYÁSZATI FEJLESZTÉSI INTÉZET
RENDSZERELEMZÉSI ÉS GEOMATEMATIKAI OSZTÁLY
BUDAPEST III. MIKOVINY S. U. 2–4
H–1037
SZIDAROVSZKY FERENC
DEPARTMENT OF SYSTEMS AND INDUSTRIAL ENGINEERING
ARIZONAI EGYETEM, TUCSON
ARIZONA, USA–85721

CONVERGENCE AND STABILITY OF DYNAMIC PROCESSES

S. MOLNÁR and F. SZIDAROVSZKY

General conditions are presented for the convergence of algorithmic models with discrete and continuous time scales. These results generalize earlier results for the convergence of iteration methods as well as classical stability theorems for difference and differential equations.

Alkalmazott Matematikai Lapok 17 (1993) 293–320

A SZERKEZETANALÍZIS EGY MATEMATIKAI MODELLJE LOKÁLIS EGYENSÚLYI FOLYAMATOK ESETÉN

VÁSÁRHELYINÉ SZABÓ ANNA

Budapest

Szerkezetek állapotváltozásának vizsgálatára a műszaki irodalomban az úgynevezett „path following" eljárásokat használják. Most a probléma egy másfajta megközelítését mutatjuk be paraméteres optimalizálás felhasználásával.

A feladatot függvényterek direkt szorzatán értelmezett matematikai programozási problémára vezetjük vissza. A szélsőérték függvény létezésének szükséges feltételét vizsgáljuk és módszert adunk a megoldásra a feladat ℓ^2 térbe történő transzformálásával.

1. A probléma vázlata

A matematikai fizikában az összetett folyamatok leírásánál egyensúlyi, lokális egyensúlyi és nemegyensúlyi állapotokat különböztetnek meg [12]. A továbbiakban a lokális egyensúlyi állapotban történő állapotváltozások számítására adunk matematikai modellt és megoldási algoritmust.

A lokális egyensúlyi állapotok esetén az állapotfüggvények terében felírt, a folyamatot jellemző függvénynek (pl. különböző energiafüggvények vagy entrópia függvény), képezik az idő szerinti első deriváltját egy adott időpontban. Az időpont megválasztása nem tetszőleges, hanem azokban az időpontokban lehet lokális egyensúlyi állapotot feltételezni, amikor a nemegyensúlyi állapotokat jellemző disszipatív hatásokat leíró úgynevezett belső változók értékét nullának lehet tekinteni [12]. Feltételezik, hogy a lokális egyensúlyi állapotban az állapothatározók sebességeire felírhatók az egyensúlyi és a kompatibilitási egyenletek, illetve az energia-sebességekre vonatkozó szélsőérték létezésének szükséges és elégséges feltételei [16]. A lokális egyensúlyi állapotokkal a nemegyensúlyi állapotot közelítik az egyensúlytól nem távoli állapotokban; így — mivel a lokális egyensúlyi állapotban a folyamat egyensúlya nem áll be — az állapothatározókra felírt egyenlőtlenségi feltételek közül legalább egy egyenlőséggel kell teljesüljön.

A legközismertebb feladat ebben a témakörben a tartószerkezetek képlékeny viselkedésének leírása [11]. A képlékeny állapotváltozás a legegyszerűbb esetben (csak alakváltozási és potenciális energiák figyelembevétele esetén) a következő feladatpárral adható meg:

a. Az egyensúlyi egyenleteket kielégítő feszültség-sebesség állapotok közül az a feszültség-sebesség állapot fog létrejönni, amely mellett a kiegészítő alakváltozási energia-sebesség minimális és a feszültségek kielégítik a képlékenységi feltételeket.

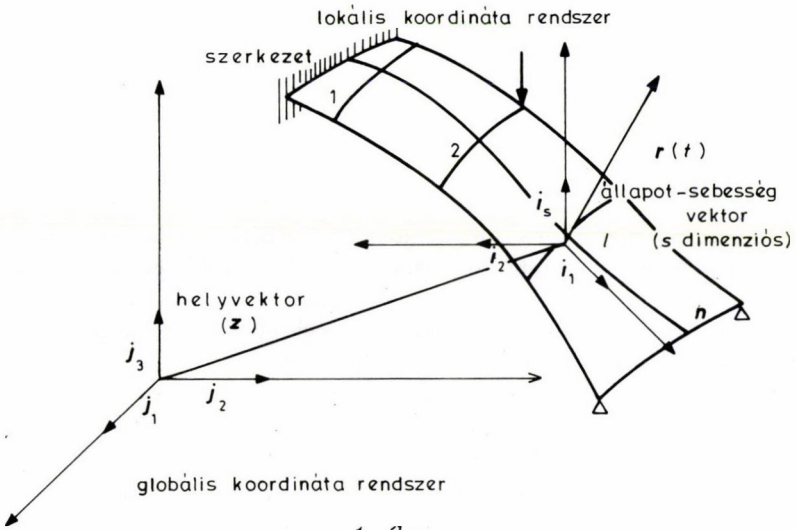
b. A kompatibilitási egyenletek kielégítő alakváltozás-sebesség állapotok közül az az alakváltozás-sebesség állapot realizálódik, amelynél az alakváltozási energia-sebesség minimális és a képlékeny sebességszorzók pozitívak.

Ezen feladatpárt csak azon feltételezés mellett oldották meg, hogy a lokális egyensúlyi állapotban az állapothatározók sebességei már nem függenek az időtől [4].

Ebben a megközelítésben erre a megszorításra nincs szükség, továbbá módszert adunk arra, hogy az állapothatározók sebességeinek az időtől való függését hogyan lehet figyelembe venni.

A lokális egyensúlyi állapot esetén a szerkezetek állapotjellemzőinek sebességeit (például feszültség-sebesség, alakváltozás-sebesség, stb.) vektor térben vektor-skalár függvényként lehet leírni. Ha a szerkezetet diszkretizáljuk, akkor a vektortér véges dimenziójú. A szerkezethez rendelt globális koordinátarendszerben minden egyes, a diszkretizálás során kialakított pontra — a csomópontra — mutató helyvektorhoz hozzárendelünk egy állapotjellemző-sebesség (feszültség-sebesség, alakváltozás-sebesség, stb.) vektort, melyet a csomópontban definiált lokális koordináta-rendszerben adunk meg. A vektortér dimenziója a csomópontszám (n) és a szabadságfok (s) szorzata. (Például ha egy csomópontban s független feszültség- ill. alakváltozás-sebesség komponenst tételezünk fel és a csomópontok száma n, akkor a vektortér dimenziója ns.)

Lokális egyensúlyban a szerkezet állapotváltozását az egyensúlyi állapot leírásánál használt lokális és globális koordináta-rendszerekben, függvényként lehet megadni.



1. ábra

A globális koordináta-rendszerben értelmezett helyvektorokhoz rendelt állapotjellemző-sebességek tehát időtől függő vektor–skalár függvények, amelyeket a lokális koordináta-rendszerekben határozunk meg.

A továbbiakban csak kis elmozdulásokkal kívánunk foglalkozni, ezért feltételezzük, hogy a szerkezet Euler és Lagrange leírási módja megegyezik, vagyis a helyvektor időtől független. Ez azt jelenti, hogy csak a lokális koordináta-rendszerben leírt állapotjellemző-sebességek időfüggőek.

A szerkezet lokális egyensúlybeli állapotváltozásának időbeli lefolyását az úgynevezett stacionárius görbe írja le. Ennek a görbének, vagy diszkrét pontjainak a megahatározása a cél. A feladat nehézségét mutatja, hogy a differenciáltopológia eszközeinek használatával kimutatták, hogy az egy paramétertől függő stacionárius görbék nem folytonosak; ezért jelenleg csak egy-egy folytonos darab – komponens meghatározására van lehetőség [13].

Az optimalizáláselméletben jelenleg az egyik legfontosabb kérdés az egy vagy több paramétertől függő feladatok Karush–Kuhn–Tucker stacionárius görbéinek struktúrális vizsgálata, amely lehetővé teszi az optimalizálási feladatok érzékenység- és/vagy stabilitásvizsgálatát. Az egy paramétertől függő feladatok esetén az egyik lehetséges eszköz az ilyen típusú kérdések megválaszolására a differenciáltopológia [7], [8], [13], [14]. Az egy paraméteres stacionárius görbék meghatározására jelenleg az úgynevezett útkövető („path-following") algoritmusokat használják, melyeknek az egyik újabb alkalmazási területe az úgynevezett belső pontos algoritmusok. Erről a témáról az utóbbi néhány évben több, mint 2000 dolgozat született, ezért csak a Math. Programming egyik összefoglaló számára hivatkozunk [17].

Ebben a dolgozatban az egy paramétertől függő Karush–Kuhn–Tucker stacionárius görbék egy folytonos komponens-darabjának meghatározására javasolunk új módszert.

A 2. pontban megadjuk a tér struktúrájának matematikai leírását. Az így kialakított térben általános esetben írjuk fel a megoldandó feladatot a 3. pontban. A szélsőérték tételek a ℓ^2 térben bizonyítottak, ezért a 4. pontban a L^2 és ℓ^2 terek közötti átmenettel foglalkozunk. Az 5. pontban a Fritz–John feltételekkel foglalkozunk az előzőekben vizsgált terekben értelmezett matematikai programozási problémák esetén. A nemlineáris függvényeket is tartalmazó esettel a 6. pont foglalkozik. A 7. pontban a Fritz–John optimalitási feltételeket transzformáljuk a ℓ^2 térbe és ennek segítségével eljárást adunk a probléma megoldására. A funkcionális deriváltak használatát mutatjuk be a 8. pontban. Végül a 9. pontban két mintafeladaton mutatjuk be a módszert.

2. A tér struktúrájának matematikai leírása

Tekintsünk egy tetszőleges s dimenziós vektor–skalár függvényt. Ennek a függvénynek minden komponense legyen eleme az $L^2(\Omega)$, $\Omega = [1, 0]$ térnek (négyzetesen integrálható függvények tere [18]). Mivel az $L^2(\Omega)$ Hilbert tér, ezért választhatunk olyan $P_i(t)$ $(i = 1, \ldots, \infty)$ bázist, mely az $\Omega = [0, 1]$ intervallumon ortonormált és amelyben a Hilbert tér bármely eleme a következőképpen írható fel:

$$x(t) = \sum_{i=1}^{\infty} \alpha_i P_i(t), \quad P_i(t) \in L^2(\Omega), \ \alpha_i \in \mathbb{R}, \ i = 1, \ldots, \infty, \ t \in [0, 1], \tag{1}$$

ahol $\mathbb{R}$ az 1 dim. Euklideszi tér, vagyis a valós számok halmaza.

A továbbiakban a $P_i(t)$ bázisok ortonormált polinomrendszerek lesznek a $[0, 1]$ intervallumon értelmezve.

Az s dimenziós függvény teret, amely egy csomópont állapotjellemzőinek leírására szolgál, a megfelelő Hilbert terek direkt szorzataként definiáljuk:

$$L_1^2 \times L_2^2 \times \cdots \times L_j^2 \times \cdots \times L_s^2.$$

E tér egy elemét a következő alakban írhatjuk fel:

$$\boldsymbol{r}(t) = \sum_{j=1}^{s} x_j(t) \ \boldsymbol{i}_j = \sum_{j=1}^{s} \left(\sum_{i=1}^{\infty} \alpha_{ij} P_i^j(t) \right) \boldsymbol{i}_j,$$

$$\alpha_{ij} \in \mathbb{R}, \ P_i^j(t) \in L_j^2([0, 1]), \ t \in [0, 1], \ j = 1, \ldots, s, \ i = 1, \ldots, \infty, \tag{2}$$

ahol $\boldsymbol{i}_j$ $(j = 1, \ldots, s)$ a lokális koordináta-rendszer s dimenziós terének az egységvektorai, $P_i^j(t)$ a lokális koordináta-rendszer j-edik tengelyéhez rendelt i-edik bázis komponens.

A szerkezet minden egyes kitüntetett pontjához — a csomópontokhoz — tartozó helyvektorhoz az alábbiakban definiált tereket rendeljük hozzá. Az ℓ-edik csomóponthoz tartozó tér:

$$F^{\ell} = (\mathbb{R}^3 \times L_1^2 \times L_2^2 \times \cdots \times L_s^2), \quad \ell = 1, \ldots, n, \tag{3}$$

amelynek egy eleme:

$$\boldsymbol{y}^{\ell}(t) = (z_1^{\ell}, z_2^{\ell}, z_3^{\ell}, x_1^{\ell}(t), x_2^{\ell}(t), \ldots, x_s^{\ell}(t)), \quad t \in [0, 1], \tag{4}$$

ahol z_1^{ℓ} az ℓ-edik csomópont $(\ell = 1, \ldots, n)$ i-edik helyvektorának $(i = 1, 2, 3)$ koordinátái. Tehát egy szerkezet egyensúlyi állapotváltozására vonatkozó feladatokat az

$$\mathcal{F} = (\mathbb{R}^3 \times L_1^2 \times L_2^2 \times \cdots \times L_s^2)^n \tag{5}$$

térben írjuk le.

A meghatározandó elemeket egy mátrixban foglaljuk össze:

$$\boldsymbol{Y}(t) = \begin{bmatrix} z_1^1 & z_2^1 & z_3^1 & x_1^1(t) & x_2^1(t) & \dots & x_s^1(t) \\ z_1^2 & z_2^2 & z_3^2 & x_1^2(t) & x_2^2(t) & \dots & x_s^2(t) \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ z_1^n & z_2^n & z_3^n & x_1^n(t) & x_2^n(t) & \dots & x_s^n(t) \end{bmatrix}, \quad t \in [0,1].$$

Az egyes függvények együtthatóit tekintve ismeretlennek az $\boldsymbol{Y}(t)$ mátrix felírható a következő módon:

$$[\boldsymbol{Y}(t)] = [\boldsymbol{A}\hat{\boldsymbol{\alpha}}] \cdot \begin{bmatrix} \boldsymbol{E} & \boldsymbol{0} \\ \boldsymbol{0} & \hat{\boldsymbol{B}}(t) \end{bmatrix}, \tag{6}$$

ahol

$$[\boldsymbol{A}] = [z_i^\ell,\ \ell = 1,\dots,n,\ i = 1,2,3],$$

$$[\hat{\boldsymbol{\alpha}}] = \begin{bmatrix} \alpha_{11}^1 \cdots \alpha_{\infty 1}^1 & \alpha_{12}^1 \cdots \alpha_{\infty 2}^1 & \alpha_{1s}^1 \cdots \alpha_{\infty s}^1 \\ \alpha_{11}^n \cdots \alpha_{\infty 1}^n & \alpha_{12}^n \cdots \alpha_{\infty 2}^n & \alpha_{1s}^n \cdots \alpha_{\infty s}^n \end{bmatrix},$$

$$\alpha_{ij} \in \mathbb{R}, \quad j = 1,\dots,s, \quad i = 1,\dots,\infty,$$

$[\boldsymbol{E}] = [j_1, j_2, j_3]$ a globális koordináta-rendszer egységvektorait tartalmazó egységmátrix,

$[\hat{\boldsymbol{B}}(t)] = [\hat{\boldsymbol{B}}_j(t),\ j = 1,\dots,s]$ hiperdiagonál mátrix, melynek egy blokkja a következő oszlopvektor:

$$\hat{\boldsymbol{B}}_j(t) = [P_i^j(t),\ i = 1,\dots,\infty], \tag{7}$$

és a $P_i^j(t)$, $i = 1,\dots,\infty$, $j = 1,\dots,s$ lineárisan független, ortonormált polinom rendszert alkot.

A feladatot az $\mathcal{F} = (\mathbb{R}^3 \times L_1^2 \times L_2^2 \times \cdots \times L_s^2)^n$ térben fogalmazzuk meg, melynek egy csomóponthoz tartozó bázisát az $\begin{bmatrix} \boldsymbol{E} & \boldsymbol{0} \\ \boldsymbol{0} & \hat{\boldsymbol{B}}(t) \end{bmatrix}$ hiperdiagonál mátrix oszlopai jelentik. A teljes tér bázisát a csomópontok számának megfelelően az $\begin{bmatrix} \boldsymbol{E} & \boldsymbol{0} \\ \boldsymbol{0} & \hat{\boldsymbol{B}}(t) \end{bmatrix}$ mátrix segítségével képzett, n blokkból álló hiperdiagonál mátrix oszlopai adják.

A továbbiakban, a számítások egyszerűbb leírása érdekében a következő jelölésrendszert használjuk: az ismeretlen együtthatókat az $[\hat{\boldsymbol{\alpha}}]$ mátrix tartalmazza. Képezzünk az $[\hat{\boldsymbol{\alpha}}]$ mátrixból egy $\boldsymbol{\alpha}$ hipervektort a sorok egymásután írásával. Az egyes szabadságfokokhoz, illetve csomópontokhoz rendelés a vektor particionálásával történik. A bázis $\hat{\boldsymbol{B}}(t)_j$ blokkjaival hiperdiagonál mátrixot képezünk a csomópontszámnak megfelelően. Így

$$\underset{[1,\infty\cdot s\cdot n]}{\boldsymbol{\alpha}^*} \begin{bmatrix} \boldsymbol{B}(t) \\ [\infty\cdot s\cdot n, s\cdot n] \end{bmatrix} = [\boldsymbol{r}^1(t), \boldsymbol{r}^2(t), \dots, \boldsymbol{r}^n(t)]^* = \boldsymbol{x}(t)^* =$$

$$= [x_1^1(t), x_2^1(t), \dots, x_s^1(t), x_1^2(t), x_2^2(t), \dots, x_s^2(t), \dots, x_1^n(t), x_2^n(t), \dots, x_s^n(t)]^*,$$

ahol az α hipervektor és $B(t)$ hiperdiagonál mátrix particionálását a jelük alatt adtuk meg.

A mátrixjelölés használatánál a vektorok alapértelmezése oszlopvektor, a transzponálást *-gal jelöltjük.

3. Az általános feladat felírása

A szerkezet állapotváltozását a következő t paramétertől függő feladat sorozat segítségével vizsgálhatjuk:

$$
\begin{gathered}
\min f(\boldsymbol{y}) \\
g_i(\boldsymbol{y}) \leq 0, \quad i = 1, \ldots, m, \\
h_j(\boldsymbol{y}) = 0, \quad j = 1, \ldots, \ell, \\
\boldsymbol{y} \in \mathbb{R}^{ns}, f, g_i, h_j : \mathbb{R}^{ns} \to \mathbb{R}, \\
\boldsymbol{y} = \boldsymbol{x}(t), \quad \forall \text{ fix } t\text{-nél}, \quad \boldsymbol{x}(t) \in \mathcal{F}, \quad t \in [0,1].
\end{gathered} \tag{8}
$$

Ez a feladat sorozat úgy értendő, hogy $\forall\, t \in [0,1]$ esetén a (8) nemlineáris programozási feladatnak stacionárius pontja van, ami meghatároz egy $\boldsymbol{x}(t) \in \mathcal{F}$ görbét.

A cél a fenti optimalizálási feladat elméleti kezelhetőségének és numerikus megoldhatóságának biztosítása. Ilyen típusú feladatok a paraméteres optimalizálásban vetődnek fel [7], [8], [13].

Az n dimenziós Euklideszi térben értelmezett klasszikus nemlineáris programozási feladat a következő:

$$
\begin{gathered}
\min\{f(\boldsymbol{x}) \mid g_i(\boldsymbol{x}) \leq 0,\ h_j(\boldsymbol{x}) = 0,\ i = 1, \ldots, m,\ j = 1, \ldots, \ell,\ \boldsymbol{x} \in \mathbb{R}^n\}, \\
f(\boldsymbol{x}), \quad g_i(\boldsymbol{x}), \quad h_j(\boldsymbol{x}) \in C^1,
\end{gathered} \tag{9}
$$

ahol ismeretlenek az $\boldsymbol{x}$ vektor elemei.

Az optimális pontok elsőrendű jellemzésére a Kuhn-Tucker tétel szolgál [2].

A Banach térben értelmezett matematikai programozási feladatot az előző feladat mintájára az alábbi formában írhatjuk:

$$
\min\{f(\boldsymbol{x}) \mid g_i(\boldsymbol{x}) \leq 0,\ h_j(\boldsymbol{x}) = 0,\ i = 1, \ldots, m,\ j = 1, \ldots, \ell,\ \boldsymbol{x} \in B\}, \tag{10}
$$

ahol B egy lineáris normált tér (Banach tér).

Erre a feladatra is általánosítható a Kuhn-Tucker tétel [3].

A (8) feladat elméleti vizsgálatához elsősorban a szélsőérték tételeket kell bebizonyítani. Ehhez felhasználjuk a Banach térre vonatkozó eredményeket.

4. Átmenet a $(\mathbb{R}^3 \times L_1^2 \times L_2^2 \times \cdots \times L_s^2)^n$ és $(\mathbb{R}^3 \times \ell_1^2 \times \ell_2^2 \times \cdots \times \ell_s^2)^n$ terek között

A (8) feladatban szereplő ismeretleneket az $L^2(\Omega)$ terek direkt szorzatán értelmezett tér helyett az izomorfia tétel alapján [15], [18] a ℓ^2 terek direkt szorzatán tekintjük. Erre az átmenetre akkor van lehetőség, ha a g_i, h_j és f függvények a szorzattér lineáris és metrikus struktúrájához illeszkednek.

Tekintsünk egy teteszőleges $q(t) \in L^2$ függvényt, amelyet a $P_i(t)$, $i = 1, \ldots, \infty$ ortonormált bázis segítségével adunk meg. Ennek a $\boldsymbol{q} \in \ell^2$ végtelen dimenziós vektor feleltethető meg az alábbi formában a $\sum_{i=1}^{\infty} \alpha_i^2 < \infty$ feltétel mellett:

$$(11) \quad q(t) = \alpha_0 + \sum_{i=1}^{\infty} \alpha_i \, P_i(t), \; t \in [0,1], \; \alpha_i \in \mathbb{R}; \quad \boldsymbol{q} = \alpha_0 + \sum_{i=1}^{\infty} \alpha_i \boldsymbol{e}_i, \quad \alpha_i \in \mathbb{R},$$

ahol $\boldsymbol{e}_i$ az ℓ^2 tér i-edik egységvektora.

Definiálunk egy olyan $\boldsymbol{f}$ teret a ℓ^2 terek direkt szorzatán, amelynek struktúrája megegyezik az $\mathcal{F}$ tér struktúrájával azaz:

$$\boldsymbol{f} = (\mathbb{R}^3 \times \ell_1^2 \times \ell_2^2 \cdots \times \ell_s^2)^n.$$

A $\boldsymbol{f}$ térben értelmezünk egy $[\boldsymbol{b}]$ hipermátrixot, melynek struktúrája megegyezik a $[\boldsymbol{B}(t)]$ mátrixéval, a $[\boldsymbol{B}(t)]$ mátrixban lévő $P_i^j(t)$ elemeknek a $\boldsymbol{f}$ tér egységvektorai $\boldsymbol{e}_i^j$ felelnek meg bármely i és j esetén.

A továbbiakban a $\mathcal{F}$ és a $\boldsymbol{f}$ terek közötti átmenetet mutatjuk meg lineáris-, paraméter szerinti derivált- valamint integrál- és nemlineáris függvények esetén a 2. pontban definiált, k-adik csomóponthoz tartozó $\boldsymbol{x}_j^k(t)$, $k = 1, \ldots, n$, $j = 1, \ldots, s$ típusú, ismeretlen függvényekre vonatkozóan.

Az α_0 konstans a két tér közötti átmenetet nem befolyásolja, így a továbbiakban nem jelezzük, kivéve a relációra vonatkozó átmenet tárgyalását, ahol az egyenlőtlenségek vizsgálatánál van jelentősége.

a. Az $\mathcal{F}$ és a $\boldsymbol{f}$ terek között az átmenet az összeadásnál, a skalárral való szorzásnál és a skalárszorzatnál a következő:

Az $\mathcal{F}$ térben:

$$\boldsymbol{x}(t) + \boldsymbol{y}(t) = \boldsymbol{\alpha}^*[\boldsymbol{B}(t)] + \beta^*[\boldsymbol{B}(t)] = (\boldsymbol{\alpha} + \beta)^* \; [\boldsymbol{B}(t)],$$

$$c\,\boldsymbol{x}(t) = c\,\boldsymbol{\alpha}^*[\boldsymbol{B}(t)],$$

$$(12) \qquad \int_0^1 \boldsymbol{x}(t)^* \boldsymbol{y}(t) dt = \int_0^1 \boldsymbol{\alpha}^*[\boldsymbol{B}(t)][\beta^*[\boldsymbol{B}(t)]]^* dt =$$

$$= \boldsymbol{\alpha}^* \int_0^1 [\boldsymbol{B}(t)][\boldsymbol{B}(t)]^* dt \, \beta = \boldsymbol{\alpha}^* \beta,$$

ahol $x(t)$ és $y(t)$ ns méretű vektorok és c skalár. A skalárszorzatnál kihasználtuk, hogy a $P_i^j(t)$ polinomrendszerek ortonormáltak a $[0,1]$ intervallumon.

A $\boldsymbol{f}$ térben

$$\begin{aligned} \boldsymbol{x}+\boldsymbol{y} &= \boldsymbol{\alpha}^*[b]+\beta^*[b]=(\boldsymbol{\alpha}+\beta)^*[b], \\ c\,\boldsymbol{x} &= c\,\boldsymbol{\alpha}^*[b], \\ \boldsymbol{x}^*\boldsymbol{y} &= \boldsymbol{\alpha}^*[b]\left[\beta^*[b]\right]^* = \boldsymbol{\alpha}^*[b]\ [b]^*\beta = \boldsymbol{\alpha}^*\beta, \end{aligned} \tag{13}$$

ahol az $\boldsymbol{x}$ és $\boldsymbol{y}$ vektorok végtelen sok ns méretű vektorból tevődnek össze és c skalár.

b. Ha az $\mathcal{F}$ térben felírt feladatban t szerinti derivált függvények (melyeket a függvény felett lévő pont jelöl) lineáris összefüggésekben szerepelnek, akkor a derivált függvényeket fel kell írni a tér bázisában a bázis szerinti általánosított Fourier-sor segítségével:

$$\dot{P}_i(t)=\sum_{k=1}^{\infty}\gamma_{ik}P_k(t),\ \text{ahol}\ \gamma_{ik}=\int_0^1 P_k(t)\dot{P}_i(t)dt. \tag{14}$$

Az $\boldsymbol{x}(t)$ vektor j-edik elemének deriváltját a $\mathcal{F}$ térben a következőképp írjuk fel:

$$\begin{aligned} \dot{x}(t)_j &= \sum_{i=1}^{\infty}\alpha_{ij}\dot{P}_i^j(t)=\sum_{k=1}^{\infty}\sum_{i=1}^{\infty}\gamma_{ki}\alpha_{ij}P_k^j(t)=\sum_{k=1}^{\infty}\beta_{kj}P_k^j(t), \\ &\text{ahol}\ \beta_{kj}=\sum_{i=1}^{\infty}\gamma_{ki}\alpha_{ij}. \end{aligned} \tag{15}$$

Mátrix-jelöléssel a t szerinti deriváltvektor a következő:

$$\dot{\boldsymbol{x}}(t)=\beta^*[\boldsymbol{B}(t)].$$

Az eddigiek alapján át lehet térni az $\boldsymbol{f}$ térbe:

$$\dot{x}_j=\sum_{k=1}^{\infty}\sum_{i=1}^{\infty}\gamma_{ik}\alpha_{ij}e_k^j=\sum_{k=1}^{\infty}\beta_{kj}e_k^j, \tag{16}$$

vagy

$$\dot{\boldsymbol{x}}=\beta^*[b].$$

c. Hasonlóan, ha a feladatban a bázis függvények paraméter szerinti integrálja szerepel, a lineáris összefüggésekben az integrált a bázisok szerint Fourier-sorba fejtjük:

$$\int_0^t P_i(\tau)d\tau=\sum_{k=1}^{\infty}\nu_{ik}P_k(t),\quad \text{ahol}\ \nu_{ik}=\int_0^1\left(\int_0^t P_i(\tau)d\tau P_k(t)\right)dt. \tag{17}$$

Az $\boldsymbol{x}(t)$ vektor j-edik komponensének paraméter szerinti integrálja a $\mathcal{F}$ térben, a tér bázisában az alábbi módon adható meg:

(18)
$$\tilde{x}_j(t) = \int_0^t x_j(\tau)d\tau = \int_0^t \sum_{i=1}^{\infty} \alpha_{ij} P_i^j(\tau)d\tau = \sum_{k=1}^{\infty}\sum_{i=1}^{\infty} \nu_{ki}\alpha_{ij} P_k^j(t) = \sum_{k=1}^{\infty} \eta_{kj} P_k^j(t),$$

ahol $\eta_{kj} = \sum_{i=1}^{\infty} \nu_{ki}\alpha_{ij}$, vagy mátrix jelöléssel

$$\tilde{\boldsymbol{x}}(t) = \int_0^t \boldsymbol{x}(\tau)d\tau = \boldsymbol{\eta}^*[\mathbf{B}].$$

A (18) átalakítás eredménye átvihető a $\boldsymbol{f}$ térbe:

(19)
$$\tilde{x}_j = \sum_{k=1}^{\infty}\sum_{i=1}^{\infty} \nu_{ki}\alpha_{ij} e_k^j = \sum_{k=1}^{\infty} \eta_{k1} e_k^j,$$

vagy $\tilde{\boldsymbol{x}} = \boldsymbol{\eta}^*[b]$

d. Az $\mathcal{F}$ térben felírt egyenlőségi relációk $\boldsymbol{f}$ térbeni megfeleltetéséhez az egyenlőségben résztvevő tagokat ismét a $P_i^j(t)$, $i = 1, \ldots, \infty$, $j = 1, \ldots, ns$ bázis szerinti általánosított Fourier-sorba fejtjük.

Legyen $\boldsymbol{x}(t) = \mathbf{c}$, ahol $\boldsymbol{c}$ t-től független, konstansokat tartalmazó vektor.

Az $x_j(t)$ függvények és a c_j, $j = 1, \ldots, ns$ konstansok $P_i^j(t)$, $i = 1, \ldots, \infty$ függvényrendszer szerinti általánosított Fourier-sorai az alábbiak:

(20)
$$c_j = K_j + \sum_{i=1}^{\infty} \varphi_{ij} P_i^j(t), \quad \varphi_{ij} = c_j \int_0^1 P_i^j(t)dt,$$
$$x_j(t) = k_j + \sum_{i=1}^{\infty} \alpha_{ij} P_i^j(t), \quad \alpha_{ij} = \int_0^1 x_j(t) P_i^j(t)dt,$$

ahol

$$k_j = \int_0^1 x_j(t)dt, \qquad K_j = \int_0^1 c_j dt = c_j.$$

Ebben a részben figyelembe vesszük a (11) kifejezésben szereplő konstansokat is.

Az egyenlőséget nullára rendezzük és felhasználjuk az összeadásra levezetett (12) összefüggést. Az $\boldsymbol{x}(t)$ és a $\boldsymbol{c}$ vektor j-edik elemei esetén az egyenlőségi reláció azt adja, hogy

(21)
$$\sum_{i=1}^{\infty} (\alpha_{ij} - \varphi_{ij}) P_i^j(t) + k_j - K_j = 0.$$

A (21) egyenlőség csak $\alpha_{ij} = \varphi_{ij}$, $k_j = K_j$, $i = 1, \ldots, \infty$, $j = 1, \ldots, ns$ esetén állhat fenn.

A j-edik vektor i-edik eleme esetén az egyenlőségi reláció a $\boldsymbol{f}$ térben

$$\sum_{i=1}^{\infty}(\alpha_{ij} - \varphi_{ij})e_i^j + k_j - K_j = 0, \tag{22}$$

azaz

$$\alpha_{ij} = \varphi_{ij}, \quad k_j = K_j, \quad i = 1, \ldots, \infty, \quad j = 1, \ldots, ns.$$

e. Egyenlőtlenségi feltétel esetén nehézséget okoz az, hogy az L^2 és a ℓ^2 terek közötti izomorfia nem rendezéstartó, azaz nem létezik olyan ortonormált függvényrendszer az L^2 térben, hogy a $g(x) \geq 0$, $g(x) \in L^2[0,1]$ egyenlőtlenség akkor és csak akkor teljesül, ha a Fourier-együtthatók nem negatívak [6]. Azt bizonyítjuk be, hogy az L^2 és a ℓ^2 terek közötti izomorfia akkor és csak akkor rendezéstartó, ha a $\int_0^1 g\, P_n(t)dt \geq 0$ egyenlőtlenség csak $g \geq 0$ esetén áll fenn. A bizonyítás Dancs István gondolata alapján a következő:

Ha az állítás igaz, akkor $P_n(t) \geq 0$, $t \in [0,1]$ majdnem mindenütt.

Tegyük fel az állítással ellentétben, hogy létezik egy pozitív mértékű A_n halmaz, amelyben $P_n(t) < 0$. Legyen az A_n halmaz karakterisztikus függvénye $\chi_{A_n} \in L^2([0,1])$, amelyre a feltétel miatt $\int_0^1 \chi_{A_n} P_n(t)dt \leq 0$, ami ellentmond a rendezési feltételnek.

Az ortogonalitás miatt az $A_n = \{t \mid P_n(t) \neq 0 \mid P_n(t) > 0\}$ halmazok majdnem mindenütt diszjunktak.

χ_{A_n} Fourier-sora $\chi_{A_n} = P_n(t) \int_0^1 \chi_{A_n} P_n(t)dt$, amiből következik, hogy

$$P_n(t) = \frac{1}{\int_0^1 \chi_{A_n} P_n(t)dt}\chi_{A_n} = \alpha_n \chi_{A_n} \text{ és } \int_0^1 P_n(t)^2 dt = \int_0^1 \alpha_n^2 \chi_{A_n}^2 dt = \alpha_n^2 \mu(A_n) = 1,$$

azaz $P_n(t) = \frac{1}{\sqrt{\mu(A_n)}}\chi_{A_n}$.

A kérdés már csak az, hogy a $\{P_n(t)\}$ rendszer teljes-e. A Lebesque mértékre igaz, hogy pozitív mértékű A_n halmaz esetén létezik A_n' és A_n'' úgy, hogy

$$A_n = A_n' \cup A_n'', \; A_n' \cap A_n'' = 0 \text{ és } \mu(A_n') = \mu(A_n'') = \tfrac{1}{2}\mu(A_n).$$

$$\text{Legyen } f(t) = \begin{cases} 1, & \text{ha } t \in A_n' \\ -1, & \text{ha } t \in A_n'', \\ 0, & \text{ha } t \notin A_n \end{cases}$$

akkor

$$\int_0^1 f(t)P_n(t)dt = \frac{1}{\sqrt{\mu(A_n)}} \int_0^1 (\chi'_{A_n} - \chi''_{A_n})\chi_{A_n} dt = 0,$$

ami azt jelenti, hogy a $P_n(t)$ rendszer nem teljes, tehát a bizonyítandó állítás igaz.

A fenti tétel miatt az egyenlőtlenségeket csak becsülni tudjuk.

A számításokban az $\mathcal{F}$ tér bázisából csak véges sokat veszünk figyelembe. Ha az $\mathcal{F}$ tér bázisfüggvényeit lépcsős függvényekkel közelítjük, akkor az egyenlőtlenségek az egyes lépcsőknél kiértékelhetők.

A gyakorlatban tehát a felhasználó által megadott m számú időpontban (t_r, $r = 1, \ldots, m$) a $\mathcal{F}$ tér véges számú bázisának függvényértékei kiszámíthatók, így az adott t értékeknél felírt egyenlőtlenségek már nem függenek a t-től, csak Fourier-együtthatókat és konstansokat tartalmaznak.

Legyen $\boldsymbol{x}(t) \leq \boldsymbol{c}$.

Az $\boldsymbol{x}(t)$ függvényeket és a $\boldsymbol{c}$ konstansokat a $P_i(t)$, $i = 1, \ldots, \infty$ függvényrendszer szerinti általánosított Fourier-sorba fejtjük, majd az egyenlőtlenséget nullára rendezzük. A különbség vektor j-edik eleme esetén az egyenlőtlenség

$$\sum_{i=1}^{\infty} (\alpha_{ij} - P_{ij})P_i^j(t_r) + k_j - K_j \leq 0, \quad r = 1, \ldots, m, \tag{23}$$

amely becslés jósága természetesen függ a megadott időpontoktól.

5. Szélsőérték tételek bizonyítása

Tekintsük a következő feladatot:

$$\begin{aligned} &\min = F(\boldsymbol{y}) \\ &G^k(\boldsymbol{y}) \leq 0 \quad k = 1, \ldots, q \\ &H^j(\boldsymbol{y}) = 0 \quad j = 1, \ldots, m, \\ &\boldsymbol{y} \in \mathbb{R}^{ns}, \ F(\boldsymbol{y}) : \mathbb{R}^{ns} \Longrightarrow \mathbb{R}, \ G(\boldsymbol{y}), H(\boldsymbol{y}) : \mathbb{R}^{ns} \Longrightarrow \mathbb{R} \\ &\boldsymbol{y} = \boldsymbol{x}(t) \in \mathcal{F}, \quad \forall \text{fix}\, t, \quad t \in [0, 1]. \end{aligned} \tag{24}$$

Feltételezzük, hogy az $F(\boldsymbol{y})$, $H^j(\boldsymbol{y})$ és $G^k(\boldsymbol{y})$ függvények $\boldsymbol{y}$ szerint folytonosan deriválhatók.

Belátjuk, hogy az $\boldsymbol{y}$-ra vonatkozó (24)-es feladathoz tartozó Fritz–John feltétel — bizonyos feltételek mellett — megegyezik annak a feladatnak a Fritz–John feltételével az α változókra vonatkozóan, amelyet a (24) feladatból úgy származtatunk, hogy az $\boldsymbol{y}$ vektor helyébe az $\boldsymbol{x}(t)$ függvények (2)-ben megadott alakját helyettesítjük.

A t paramétert a t_0 pontban rögzítve és bevezetve az $y^0 = x(t_0)$, illetve az ℓ-edik elem esetén az $y^0_\ell = x(t_0)_\ell$ jelölést, a (24) feladathoz tartozó Fritz–John feltétel a következő:

$$\eta(t_0)\frac{\partial F(y^0)}{\partial y^0_\ell} + \sum_{j=1}^{m} \lambda_j(t_0)\frac{\partial H^j(y^0)}{\partial y^0_\ell} - \sum_{k=1}^{q} \mu_k(t_0)\frac{\partial G^k(y^0)}{\partial y^0_\ell} = 0, \quad \ell = 1,\dots,ns,$$

$$\mu_k(t_0)G^k(y^0) = 0, \quad \mu_k(t_0) \geq 0, \quad k = 1,\dots,q, \tag{25}$$

ahol $\eta(t_0)$, $\lambda_j(t_0)$, $\mu_k(t_0)$ a célfüggvényhez, a j-edik egyenlőségi ($j = 1,\dots,m$), illetve a k-adik egyenlőtlenségi ($k = 1,\dots,q$) feltételekhez tartozó Lagrange szorzók.

Az optimalizáláselméletben ismert, hogy az egy paramétertől függő Fritz–John és Karush–Kuhn–Tucker stacionárius görbék folytonos komponensekből álló, nem összefüggő halmazt alkotnak [14]. A folytonos komponensek struktúrális vizsgálatából adódnak azok a feltételek, amelyek teljesülése biztosítja a stacionárius görbe egy darabjának folytonosságát [7], [8], [14].

Ezért feltételezzük, hogy a t paramétert „futtatva" 0-tól 1-ig a $\lambda_k(t)$ és a $\mu_j(t)$ multiplikátorok t szerint folytonosak. A (24)-hez tartozó Fritz–John feltétel:

$$\eta(t)\frac{\partial F(x(t))}{\partial x(t)_\ell} + \sum_{j=1}^{m} \lambda(t)_j \frac{\partial H^j(x(t))}{\partial x(t)_\ell} - \sum_{k=1}^{q} \mu(t_k)\frac{\partial G^k(x(t))}{\partial x(t)_\ell} = 0,$$

$$\ell = 1,\dots,ns,$$

$$\mu_k(t)G^k(x(t)) = 0, \quad \mu_k(t) \geq 0, \quad k = 1,\dots,q,$$

vagyis

$$\eta(t)\,\nabla F(x(t)) + \lambda(t)\,JH(x(t)) - \mu(t)\,JG(x(t)) = 0,$$

$$\mu_k(t)G^k(x(t)) = 0, \quad \mu_k(t) \geq 0, \quad k = 1,\dots,q, \tag{26}$$

ahol a ∇ és J az — F, G, és H függvényeknek az y változó vektor komponensei szerinti deriváltjait tartalmazó — gradienst és Jacobi mátrixot jelenti $\forall \text{fix}\, t$, $t \in [0,1]$ pontban. A JH mátrix j-edik sora a j-edik feltétel gradienseinek az elemeit tartalmazza. Hasonló módon értelmezzük a JG mátrixot is. A $\lambda(t)$, illetve $\mu(t)$ a Lagrange szorzókat tartalmazó, függvény komponensekből álló vektorok.

Az ismeretlen $x(t)$ vektor elemei függvények. Ezeket általánosított Fourier-sorral

$$x(t)_\ell = \sum_{i=1}^{\infty} \alpha_{i\ell} P_i^\ell(t), \quad \ell = 1,\dots,ns,\ P_i^\ell(t) \neq 0,\ i = 1,\dots,\infty,\ \ell = 1,\dots,ns$$

megadva a (24) feladatból a következő szemi-infinit programozási feladatot kapjuk:

$$\begin{aligned} &\min F(\alpha^*[B(t)]) \\ &G^k(\alpha^*[B(t)]) \leq 0, \quad k = 1,\dots,q, \\ &H^j(\alpha^*[B(t)]) = 0, \quad j = 1,\dots,m, \\ &\alpha_{i\ell} \in \mathbb{R}, \quad i = 1,\dots,\infty, \quad \ell = 1,\dots,ns, \quad P_i^\ell(t) \in L^2_\ell(\Omega), \quad t \in [0,1], \end{aligned} \tag{27}$$

ahol a változók $\alpha_{i\ell}$, $i = 1, \ldots, \infty$, $\ell = 1, \ldots, ns$.

Az (24) és a (27) feladat közötti kapcsolatot vizsgáljuk.

A (27)-ben szereplő függvények $\alpha_{r\ell}$ szerinti deriváltjai a paraméter $t = t_0$ helyén

$$\frac{\partial F\left(\boldsymbol{\alpha}^*[\boldsymbol{B}(t_0)]\right)\partial\left(\sum\limits_{i=1}^{\infty}\alpha_{i\ell}P_i^{\ell}(t_0)\right)}{\partial\left(\sum\limits_{i=1}^{\infty}\alpha_{i\ell}P_i^{\ell}(t_0)\right)\partial\alpha_{r\ell}} = \frac{\partial F\left(\boldsymbol{\alpha}^*[\boldsymbol{B}(t_0)]\right)}{\partial\left(\sum\limits_{i=1}^{\infty}\alpha_{i\ell}P_i^{\ell}(t_0)\right)}P_r^{\ell}(t_0).$$

A (27) feladathoz tartozó Fritz–John feltétel a $t = t_0$ pontban

$$(28)\qquad \hat{\eta}(t_0)\nabla F\left(\boldsymbol{\alpha}^*[\boldsymbol{B}(t_0)]\right)P_r^{\ell}(t_0) + \sum_{j=1}^{m}\hat{\lambda}_j(t_0)\nabla H^j\left(\boldsymbol{\alpha}^*[\boldsymbol{B}(t_0)]\right)P_r^{\ell}(t_0) -$$

$$-\sum_{k=1}^{q}\hat{\mu}_k(t_0)\nabla G^k\left(\boldsymbol{\alpha}^*[\boldsymbol{B}(t_0)]\right)P_r^{\ell}(t_0) = 0, \quad \ell = 1, \ldots, ns, \quad r = 1, \ldots, \infty,$$

$$\mu_k(t_0)G^k\left(\boldsymbol{\alpha}^*[\boldsymbol{B}(t_0)]\right) = 0, \quad \mu_k(t_0) \geq 0 \quad k = 1, \ldots, q.$$

A (28) egyenleteit osztva $P_r^{\ell}(t_0) \neq 0$-val és áttérve a mátrixos írásmódra a következőket kapjuk:

$$(29)\qquad \begin{gathered}\hat{\eta}(t_0)\nabla F(\boldsymbol{\alpha}^*[\boldsymbol{B}(t_0)]) + \hat{\lambda}(t_0)\,\boldsymbol{J}H(\boldsymbol{\alpha}^*[\boldsymbol{B}(t_0)]) - \hat{\mu}(t_0)\,\boldsymbol{J}G(\boldsymbol{\alpha}^*[\boldsymbol{B}(t_0)]) = 0,\\ \hat{\mu}_k(t_0)G^k(\boldsymbol{\alpha}^*[\boldsymbol{B}(t_0)]) = 0, \quad \hat{\mu}_k(t_0) \geq 0, \quad k = 1, \ldots, q.\end{gathered}$$

A (26) egyenletek megegyeznek a (29)-es egyenletekkel, vagyis

$$\eta(t_0) = \hat{\eta}(t_0), \quad \lambda_j(t_0) = \hat{\lambda}_j(t_0), \quad \mu_k(t_0) = \hat{\mu}_k(t_0).$$

A (29) egyenletet minden t értékre képezve a (27)-es feladathoz tartozó Fritz–John feltétel

$$(30)\qquad \begin{gathered}\hat{\eta}(t)\nabla F(\boldsymbol{x}(t)) + \hat{\lambda}(t)\,\boldsymbol{J}H(\boldsymbol{x}(t)) - \hat{\mu}(t)\,\boldsymbol{J}G(\boldsymbol{x}(t)) = 0,\\ \hat{\mu}(t)^*G(\boldsymbol{x}(t)) = 0, \quad \hat{\mu}(t) \geq 0.\end{gathered}$$

A (30) egyenletek megegyeznek a (25) egyenletekkel, amiből következik, hogy

$$\eta(t) = \hat{\eta}(t), \quad \lambda_j(t) = \hat{\lambda}_j(t), \quad \mu_k(t) = \hat{\mu}_k(t),$$

vagyis a (24) és a (28) feladatok stacionárius függvényei megegyeznek.

6. Fritz–John feltétel a nemlineáris függvények Fourier-sorfejtése esetén

Belátjuk, hogy a (24) feladat F, G, és H nemlineáris függvényeinél ezen függvények Fourier-sorfejtése [18] a feladathoz tartozó Fritz–John feltételt nem befolyásolja.

Nemlineáris esetben a (27)-ben szereplő függvényeket $P_i(t)$ szerint általánosított Fourier-sorba fejtve a következő szemi-infinit programozási feladathoz jutunk:

$$
\begin{aligned}
(31) \qquad & \min \sum_{r=1}^{\infty} P_r(t) \int_0^1 F(\boldsymbol{\alpha}([\boldsymbol{B}(t)]^*) P_r(t) dt \\
& \sum_{r=1}^{\infty} P_r(t) \int_0^1 G^k(\boldsymbol{\alpha}[\boldsymbol{B}(t)]^*) P_r(t) dt \le 0, \quad k = 1, \ldots, q, \\
& \sum_{r=1}^{\infty} P_r(t) \int_0^1 H^j(\boldsymbol{\alpha}[\boldsymbol{B}(t)]^*) P_r(t) dt = 0, \quad j = 1, \ldots, m, \\
& \alpha_{i\ell} \in \mathbb{R}, \ i = 1, \ldots, \infty, \ \ell = 1, \ldots, ns, \ P_i(t) \in L^2(\Omega), \ t \in [0,1],
\end{aligned}
$$

ahol az $\alpha_{i\ell}$ együtthatók a változók.

A (31) feladat az alábbi formába írható, ahol t-től csak a bázis függvények függenek:

$$
\begin{aligned}
(32) \qquad & \min_{\alpha} \sum_{r=1}^{\infty} P_r(t) f_r(\boldsymbol{\alpha}) \\
& \sum_{r=1}^{\infty} P_r(t) g_r^k(\boldsymbol{\alpha}) \le 0, \quad k = 1, \ldots, q, \\
& \sum_{r=1}^{\infty} P_r(t) h_r^j(\boldsymbol{\alpha}) = 0, \quad j = 1, \ldots, m,
\end{aligned}
$$

ahol

$$
f_r(\boldsymbol{\alpha}) = \int_0^1 F(\boldsymbol{\alpha}^*[\boldsymbol{B}(t)]) P_r(t) dt, \quad g_r^k(\boldsymbol{\alpha}) = \int_0^1 G^k(\boldsymbol{\alpha}^*[\boldsymbol{B}(t)]) P_r(t) dt,
$$

$$
h_r^j(\boldsymbol{\alpha}) = \int_0^1 H^j(\boldsymbol{\alpha}^*[\boldsymbol{B}(t)]) P_r(t) dt.
$$

A Fourier együtthatók $\alpha_{s\ell}$ szerinti deriváltjai

$$\frac{\partial}{\partial \alpha_{s\ell}} f_r(\boldsymbol{\alpha}) = \frac{\partial}{\partial \alpha_{s\ell}} \int_0^1 F(\boldsymbol{\alpha}^*[\boldsymbol{B}(t)]) P_r(t) dt = \int_0^1 \frac{\partial}{\partial \alpha_{s\ell}} \left(F(\boldsymbol{\alpha}^*[\boldsymbol{B}(t)]) P_r(t)\right) dt =$$

$$= \int_0^1 \frac{\partial F(\boldsymbol{\alpha}^*[\boldsymbol{B}(t)])}{\partial \left(\sum_{s=1}^{\infty} \alpha_{s\ell} P_s^{\ell}(t)\right)} P_s^{\ell}(t) P_r(t) dt,$$

mivel az $\alpha_{s\ell}$ szerinti deriválás és a t szerinti integrálás sorrendje felcserélhető.

A Fritz–John feltétel a fenti deriváltakkal

$$\hat{\eta}(t) \sum_{r=1}^{\infty} P_r(t) \nabla \boldsymbol{f}_r(\boldsymbol{\alpha}) + \sum_{j=1}^{m} \hat{\lambda}_j(t) \sum_{r=1}^{\infty} P_r(t) \nabla \boldsymbol{h}_r^j(\alpha) - \tag{33}$$

$$- \sum_{k=1}^{q} \hat{\mu}_k(t) \sum_{r=1}^{\infty} P_r(t) \nabla \boldsymbol{g}_r^k(\alpha) = 0,$$

ahol $\hat{\eta}(t)$, $\hat{\lambda}_j(t)$, $j = 1, \ldots, m$, $\hat{\mu}_k(t)$, $k = 1, \ldots, q$ jelölik a (32) feladathoz tartozó Lagrange szorzókat.

Szorozzuk be a (26) egyenletet $P_s^{\ell}(t) \neq 0$-val, akkor

$$\eta(t) \nabla F(\boldsymbol{x}(t)) P_s^{\ell}(t) + \sum_{j=1}^{m} \lambda_j(t) \nabla H^j(\boldsymbol{x}(t)) P_s^{\ell}(t) - \sum_{k=1}^{q} \mu_k(t) \nabla G^k(\boldsymbol{x}(t)) P_s^{\ell}(t) = 0.$$

A deriváltaknak a polinommal való szorzatát $P_i(t)$ szerinti Fourier-sorba fejtve:

$$\eta(t) \sum_{r=1}^{\infty} P_r(t) \nabla \boldsymbol{f}_r(\boldsymbol{\alpha}) + \sum_{j=1}^{m} \lambda_j(t) \sum_{r=1}^{\infty} P_r(t) \nabla \boldsymbol{h}_r^j(\alpha) - \tag{34}$$

$$- \sum_{k=1}^{q} \mu_k(t) \sum_{r=1}^{\infty} P_r(t) \nabla \boldsymbol{g}_r^k(\boldsymbol{\alpha}) = 0.$$

A (34) egyenlet megegyezik a (33)-mal, vagyis a Fourier-sorfejtés a szélsőérték függvényt nem befolyásolja.

7. A Fritz–John optimalitási feltételek transzformálása a $\mathcal{F}$ térből a f térbe

A (24) feladathoz tartozó Fritz–John optimalitási feltételek a következők:

$$
\begin{aligned}
&G^k(\boldsymbol{x}(t)) \le 0, \quad k = 1, \dots, q,\\
&H^j(\boldsymbol{x}(t)) = 0, \quad j = 1, \dots, m,\\
&\eta(t)\,\nabla F(\boldsymbol{x}(t)) + \sum_{j=1}^{m} \lambda_j(t)\,\nabla H^j(\boldsymbol{x}(t)) - \sum_{k=1}^{q} \mu_k(t)\,\nabla G^k(\boldsymbol{x}(t)) = 0,\\
&\mu_k(t) \ge 0, \quad k = 1, \dots, q,\\
&\mu_k(t) G^k(\boldsymbol{x}(t)) = 0, \quad k = 1, \dots, q,\\
&\boldsymbol{x}(t) \in (L^2)^{sn}, \quad t \in [0,1].
\end{aligned}
\tag{35}
$$

A (35) rendszernél használjuk fel a (2) összefüggést, továbbá az $\eta(t)$, $\lambda_j(t)$ $(j = 1, \dots, m)$ és $\mu_k(t)$ $(k = 1, \dots, q)$ függvények $P_s(t)$ bázis szerinti sorfejtéseit:

$$
\begin{aligned}
&G^k(\boldsymbol{\alpha}^*[\boldsymbol{B}(t)]) \le 0, \quad k = 1, \dots, q,\\
&H^j(\boldsymbol{\alpha}^*[\boldsymbol{B}(t)]) = 0, \quad j = 1, \dots, m,\\
&\sum_{s=1}^{\infty} \kappa_s P_s(t)\,\nabla F(\boldsymbol{\alpha}^*[\boldsymbol{B}(t)]) + \sum_{j=1}^{m}\sum_{s=1}^{\infty} \zeta_{sj} P_s^j(t)\,\nabla H^j(\boldsymbol{\alpha}^*[\boldsymbol{B}(t)]) -\\
&\qquad - \sum_{k=1}^{q}\sum_{s=1}^{\infty} \theta_{sk} P_s^k(t)\,\nabla G^k(\boldsymbol{\alpha}^*[\boldsymbol{B}(t)]) = 0,\\
&\qquad \sum_{s=1}^{\infty} \theta_{sk} P_s^k(t) \ge 0, \quad k = 1, \dots, q,\\
&\sum_{s=1}^{\infty} \theta_{sk} P_s^k(t) G^k(\boldsymbol{\alpha}^*[\boldsymbol{B}(t)]) = 0, \quad k = 1, \dots, q,\\
&P_i(t) \in L^2, \quad t \in [0,1],
\end{aligned}
\tag{36}
$$

ahol $\eta(t) = \sum_{s=1}^{\infty} \kappa_s P_s(t)$, $\lambda_j(t) = \sum_{s=1}^{\infty} \zeta_{sj} P_s^j(t)$, $\mu_k(t) = \sum_{s=1}^{\infty} \theta_{sk} P_s^k(t)$. Az optimalitás szükséges feltételeiben szereplő ismeretlenek α_{il}, κ_s, ζ_{sj}, θ_{sk}, $\ell = 1, \dots, ns$, $j = 1, \dots, m$, $k = 1, \dots, q$, $i = 1, \dots, \infty$.

A következő (37) szemi-indefinit feladathoz szintén a (35) Fritz–John rendszer tartozik, így a stacionárius függvények megegyeznek:

$$
\begin{aligned}
(37)\qquad & \max F(\boldsymbol{\alpha}^*[\boldsymbol{B}(t)]) + \sum_{k=1}^{q}\sum_{s=1}^{\infty}\theta_{sk}P_s^k(t)G^k(\boldsymbol{\alpha}^*[\boldsymbol{B}(t)]) + \\
& + \sum_{j=1}^{m}\sum_{s=1}^{\infty}\zeta_{sj}P_s^j(t)H^j(\boldsymbol{\alpha}^*[B(t)]) \\
& \sum_{s=1}^{\infty}\kappa_s P_s(t)\,\nabla F(\boldsymbol{\alpha}^*[B(t)]) + \sum_{j=1}^{m}\sum_{s=1}^{\infty}\zeta_{sj}P_s^j(t)\,\nabla H^j(\boldsymbol{\alpha}^*[\boldsymbol{B}(t)]) - \\
& - \sum_{k=1}^{q}\sum_{s=1}^{\infty}\theta_{sk}P_s^k(t)\,\nabla G^k(\boldsymbol{\alpha}^*[\boldsymbol{B}(t)]) = 0, \\
& \theta_{sk}P_s^k(t) \geq 0, \quad z = 1,\ldots,\infty, \quad k = 1,\ldots,q. \\
& \alpha \in \mathbb{R}, \quad P_i(t) \in (L^2(\Omega))^{ns}, \quad t \in [0,1].
\end{aligned}
$$

Visszatérve az $x(t)_t$, $\mu_k(t)$, $\eta(t)$, $\lambda_j(t)$ változókra azt a feladatot kapjuk, hogy

$$
\begin{aligned}
(38)\qquad & \max F(\boldsymbol{y}) + \sum_{k=1}^{q}\mu_k(t)G^k(\boldsymbol{y}) + \sum_{j=1}^{m}\lambda_j(t)H^j(\boldsymbol{y}) \\
& \eta(t)\,\nabla F(\boldsymbol{y}) + \sum_{j=1}^{m}\lambda_j(t)\,\nabla H^j(\boldsymbol{y}) - \sum_{k=1}^{q}\mu_k(t)\,\nabla G^k(\boldsymbol{y}) = 0, \\
& \mu_k(t) \geq 0, \quad k = 1,\ldots,q, \\
& \boldsymbol{y} = \boldsymbol{x}(t), \quad \boldsymbol{x}(t) \in \mathcal{F}, \quad \forall \operatorname{fix} t, \quad t \in [0,1].
\end{aligned}
$$

A (24) és a (38) feladathoz is a (35) optimalitási rendszer tartozik, vagyis stacionárius pontokból álló függvényeik megegyeznek.

A továbbiakban a (24) feladatot primálnak, a (38) feladatot pedig duálnak nevezzük.

Ahhoz, hogy a (36) rendszer a $\boldsymbol{f}$ térbe átvihető legyen, a nemlineáris függvényeket Fourier-sorba kell fejteni. A (32) rendszernél bevezetett jelölések felhasználásával a (38) rendszer a következő alakba írható:

$$
\begin{aligned}
(39)\qquad & \sum_{r=1}^{\infty}P_r(t)g_r^k(\boldsymbol{\alpha}) \leq 0, \quad k = 1,\ldots,q, \\
& \sum_{r=1}^{\infty}P_r(t)h_r^j(\boldsymbol{\alpha}) = 0, \quad j = 1,\ldots,m, \\
& \sum_{s=1}^{\infty}\kappa_s P_s(t)\sum_{r=1}^{\infty}P_r(t)\int_0^1 \nabla F(\boldsymbol{x}(t))P_s^{\ell}(t)P_r(t)dt +
\end{aligned}
$$

$$+\sum_{j=1}^{m}\sum_{s=1}^{\infty}\zeta_{sj}P_s^j(t)\sum_{r=1}^{\infty}P_r(t)\int_0^1\nabla H^j(\boldsymbol{x}(t))P_s^\ell(t)P_r(t)dt-$$

$$-\sum_{k=1}^{q}\sum_{s=1}^{\infty}\theta_{sk}P_s^k(t)\sum_{r=1}^{\infty}P_r(t)\int_0^1\nabla G^k(\boldsymbol{x}(t))P_s^\ell(t)P_r(t)dt=0,$$

$$\sum_{s=1}^{\infty}\theta_{sk}P_s^k(t)\geq 0,\quad k=1,\dots,q,$$

$$\sum_{s=1}^{\infty}\theta_{sk}P_s^k(t)G^k(\boldsymbol{\alpha}[\boldsymbol{B}(t)]^*)=0,\quad k=1,\dots,q,$$

$$P_i(t)\in L^2(\Omega),\quad \forall\,\text{fix}\,t,\ t\in[0,1].$$

A (39)-es rendszer ismeretlenjei: α_{it}, κ_s, ζ_{sj}, θ_{sk}, $\ell=1,\dots,ns$, $j=1,\dots,m$, $k=1,\dots,q$, $i=1,\dots,\infty$.

Bevezetve az

$$\eta(t)=\sum_{s=1}^{\infty}\kappa_sP_s(t),\quad \lambda_j(t)=\sum_{s=1}^{\infty}\zeta_{sj}P_s^j(t),\quad \mu_k(t)=\sum_{s=1}^{\infty}\theta_{sk}P_s^k(t),$$

$$\nabla f_r(\boldsymbol{\alpha})=\int_0^1\nabla F(\boldsymbol{\alpha}^*[\boldsymbol{B}(t)])P_r(t)dt,\quad \nabla g_r^k(\boldsymbol{\alpha})=\int_0^1\nabla G^k(\boldsymbol{\alpha}^*[\boldsymbol{B}(t)])P_r(t)dt,$$

$$\nabla h_r^j(\boldsymbol{\alpha})=\int_0^1\nabla H^j(\boldsymbol{\alpha}^*[\boldsymbol{B}(t)])P_r(t)dt,\quad \tilde{g}_r^k(\boldsymbol{\alpha})=\int_0^1P_i(t)G^k(\boldsymbol{\alpha}^*[\boldsymbol{B}(t)])P_r(t)dt$$

jelöléseket a (39) rendszer tovább alakítható figyelembe véve, hogy az egyenlőtlenségeket csak adott t értékeknél vizsgáljuk:

$$\sum_{r=1}^{\infty}P_r(t_v)g_r^k(\boldsymbol{\alpha})\leq 0,\quad k=1,\dots,q,\quad v=1,\dots,w,\tag{40}$$

$$\sum_{r=1}^{\infty}P_r(t)h_r^j(\boldsymbol{\alpha})=0,\quad j=1,\dots,m,$$

$$\sum_{r=1}^{\infty}\sum_{s=1}^{\infty}P_r(t)\left(P_s(t)\kappa_s\nabla f_r(\boldsymbol{\alpha})+\sum_{j=1}^{m}\zeta_{sj}P_s^j(t)\nabla h_r^j(\boldsymbol{\alpha})-\sum_{k=1}^{q}\theta_{sk}P_s^k(t)\nabla g_r^k r(\alpha)\right)=0$$

$$\ell=1,\dots,ns,$$

$$\sum_{s=1}^{\infty}\theta_{sk}P_s^k(t_v)\geq 0,\quad k=1,\dots,q,\quad v=1,\dots,w,$$

$$\sum_{s=1}^{\infty}\theta_{sk}\sum_{r=1}^{\infty}P_r^k(t)\tilde{g}_r^k(\alpha)=0,\quad k=1,\dots,q,\quad \ell=1,\dots,ns,$$

$$P_i(t)\in L^2(\Omega),\quad t\in[0,1],$$

ahol w azon időpontok száma, ahol az egyenlőtlenségek tejlesülését vizsgáljuk.

A harmadik egyenlet-csoportnál a polinomszorzatot ismét Fourier-sorba fejtve:

$$P_r(t)P_s^j(t)=\sum_{z=1}^{\infty}P_z(t)\int_0^1 P_r(t)P_s^j(t)P_z(t)dt=\sum_{s=1}^{\infty}P_z(t)p_{rsz}^j,$$

ahol $p_{rzs}^j=\int_0^1 P_r(t)P_z^j(t)P_s(t)dt$. A v-edik időpontban a k-adik bázishoz tartozó r-edik polinom függvényértékét jelölje:

$$P_{rv}^k=P_z^k(t_v).$$

Így a (40) rendszer $\boldsymbol{f}$ térbeli alakja

$$p_{rv}^k g_r^k(\alpha)\le 0,\quad k=1,\dots,q,\quad r=1,\dots,\infty,\quad v=1,\dots,w,$$

$$h_r^j(\alpha)=0,\quad j=1,\dots,m,\quad r=1,\dots,\infty,$$

$$\sum_{r=1}^{\infty}\sum_{s=1}^{\infty}\left(p_{rzs}\nabla f_r(\alpha)+\sum_{j=1}^{m}p_{rzs}^j\eta_z^j\nabla h_r^j(\alpha)-\sum_{k=1}^{q}p_{rzs}^k\chi_z^k\nabla g_r^k(\alpha)\right)=0,$$

$$\ell=1,\dots,ns,\quad i=1,\dots,\infty,\quad z=1,\dots,\infty,$$

$$p_{rv}^k\theta_{zk}\ge 0,\quad k=1,\dots,q,\quad z=1,\dots,\infty,\quad v=1,\dots,w,$$

$$\sum_{z=1}^{\infty}\theta_{zk}\tilde{g}_r^k(\alpha)=0,\quad k=1,\dots,q,\quad r=1,\dots,\infty,$$

$$\text{ahol } \eta_z^j=\zeta_{zj}/\kappa_z,\ \chi_z^k=\theta_{zk}/\kappa_z. \tag{41}$$

Ha a (8) feladatban egyenlőtlenségi feltétel is szerepel, akkor a feladat megoldását a (41) rendszerrel meghatározott α Fourier együtthatóknak a bázis elemeivel képzett szorzatösszege adja.

Ha a (8) feladatban csak egyenlőségi feltétel szerepel, akkor a megoldás szintén előállítható a (41) rendszer segítségével az előbbieknek megfelelően, de ekkor a (41) azonos struktúrában értelmezett egyenleteket tartalmaz, így egyszerűbben megoldható.

A (41) rendszer csak egyenlőségi feltételeket tartalmazó esetben

$$h_r^j(\alpha)=0,\quad j=1,\dots,m,\quad r=1,\dots,\infty,$$

$$\sum_{r=1}^{\infty}\sum_{s=1}^{\infty}\left(p_{rzs}\nabla f_r(\alpha)+\sum_{j=1}^{m}p_{rzs}^j\eta_z^j\nabla h_r^j(\alpha)\right)=0, \tag{42}$$

$$\ell=1,\dots.ns,\quad i=1,\dots,\infty,\quad z=1,\dots,\infty.$$

A (42) rendszer első egyenlet-csoportját beszorozva a j és r indexeknek megfelelő $\sum_{s=1}^{\infty} p_{rzs} \neq 0$-val

$$\sum_{s=1}^{\infty} p_{rzs}^{j} h_r^j(\alpha) = 0, \quad j = 1, \ldots, m, \quad r = 1, \ldots, \infty, \tag{43}$$

$$\sum_{r=1}^{\infty}\sum_{s=1}^{\infty}\left(p_{rzs}\,\nabla f_r(\alpha) + \sum_{j=1}^{m} p_{rzs}^{j}\eta_z^j\,\nabla h_r^j(\alpha)\right) = 0, \quad \ell = 1, \ldots, ns, \quad i = 1, \ldots, \infty.$$

A (43) rendszer megoldása megegyezik a következő feladat-pár megoldásával:

$$\min \sum_{s=1}^{\infty}\sum_{r=1}^{\infty}\sum_{z=1}^{\infty} p_{rzs} f_r(\alpha) \tag{44}$$

$$\sum_{s=1}^{\infty} p_{rzs}^{j} h_r^j(\alpha) = 0, \quad j = 1, \ldots, m, \quad r = 1, \ldots, \infty, \quad i = 1, \ldots, \infty,$$

illetve:

$$\max \sum_{s=1}^{\infty}\left(\sum_{r=1}^{\infty}\sum_{z=1}^{\infty} p_{rzs} f_r(\alpha) + \sum_{j=1}^{m} p_{rzs}^{j}\eta_z^j h_r^j(\alpha)\right), \tag{45}$$

$$\sum_{r=1}^{\infty}\sum_{z=1}^{\infty}\left(p_{rzs}\,\nabla f_r(\alpha) + \sum_{j=1}^{m} p_{rzs}^{j}\,\eta_z^j\,\nabla h_r^j(\alpha)\right) = 0, \quad \ell = 1, \ldots, ns, \quad s = 1, \ldots, \infty.$$

A (37) probléma duál feladata a (38) feladat. Ha csak egyenlőségi feltételek vannak, akkor ezeknek a feladatoknak a $\boldsymbol{f}$ térben a (44), illetve (45) felel meg, amelyek végtelen dimenziós nemlineáris programozási feladatok.

Ha egyenlőtlenségi feltétel is szerepel a (37), illetve (38) feladatban, akkor a $\boldsymbol{f}$ térben a fenti szétválasztás nem lehetséges, mivel az egyenlőtlenségek transzformációja nem az izomorfia tétel alapján történik.

A gyakorlatban a végtelen dimenziós tereket végesre csonkítjuk, vagyis a $\mathcal{F}$ térben véges számú bázis függvényt tekintünk és a probléma megoldását csak ebben az altérben közelítjük. Ez a $\boldsymbol{f}$ térben is a dimenziók azonos módon történő csökkentését vonja maga után. Ilymódon egy véges dimenziós nemlineáris egyenlőtlenség rendszert, illetve matematikai programozási feladatot oldunk meg.

8. Fritz–John feltétel funkcionális deriváttal

Ugyanerre az eredményre jutunk, ha az L^2 térben a függvények funkcionális deriváltjaival írjuk fel a Fritz–John feltételt. Az állapothatározók sebessége folyamatok esetén nem nulla mindaddig, míg a folyamat egyensúlyi helyzetbe nem kerül.

Az elméleti fizikában feltétel, hogy a szerkezet egyensúlyi állapotát csak határértékben éri el, így a továbbiakban feltesszük, hogy az állapothatározók sebessége nem nulla. A (30) feltételt szorozzuk be, majd osszuk el a $\sum_{i=1}^{\infty} \alpha_{i\ell}\dot{P}_i(t) \neq 0$ értékkel, vagyis:

$$\eta(t)\left(\frac{d\left(F\sum_{i=1}^{\infty}\alpha_{i\ell}P_i(t)\right)}{d\left(\sum_{i=1}^{\infty}\alpha_{i\ell}P_i(t)\right)}\left(\sum_{i=1}^{\infty}\alpha_{i\ell}\dot{P}_i(t)\right)\right)\frac{1}{\left(\sum_{i=1}^{\infty}\alpha_{i\ell}\dot{P}_i(t)\right)}+$$

$$+\left(\sum_{j=1}^{m}\lambda_j(t)\frac{d\left(H^j\sum_{i=1}^{\infty}\alpha_{i\ell}P_i(t)\right)}{d\left(\sum_{i=1}^{\infty}\alpha_{i\ell}P_i(t)\right)}\left(\sum_{i=1}^{\infty}\alpha_{i\ell}\dot{P}_i(t)\right)\right)\frac{1}{\left(\sum_{i=1}^{\infty}\alpha_{i\ell}\dot{P}_i(t)\right)}-$$

$$-\left(\sum_{k=1}^{q}\mu_k(t)\frac{d\left(G^k\sum_{i=1}^{\infty}\alpha_{i\ell}P_i(t)\right)}{d\left(\sum_{i=1}^{\infty}\alpha_{i\ell}P_i(t)\right)}\left(\sum_{i=1}^{\infty}\alpha_{i\ell}\dot{P}_i(t)\right)\right)\frac{1}{\left(\sum_{i=1}^{\infty}\alpha_{i\ell}\dot{P}_i(t)\right)}=0,$$

$$\ell = 1,\ldots,ns$$

A nagy zárójelekben levő deriváltak rendre az F, G és H függvények paraméter szerinti deriváltjai. Így

$$(47)\quad \eta(t)\frac{\dot{F}(\boldsymbol{x}(t))}{\dot{x}(t)_\ell}+\sum_{j=1}^{m}\lambda_j(t)\frac{\dot{H}^j(\boldsymbol{x}(t))}{\dot{x}(t)_\ell}-\sum_{k=1}^{q}\mu_k(t)\frac{\dot{G}^k(\boldsymbol{x}(t))}{\dot{x}(t)_\ell}=0,\ \dot{\boldsymbol{x}}(t)\neq 0,$$

$$x(t)\in(L^2)^{sn},\ t\in[0,1],\ \ell=1,\ldots,ns.$$

A (47) egyenletekben szereplő paraméter szerinti deriváltak hányadosa az F, G illetve H függvények funkcionális deriváltjai [1].

A funkcionális derivált a Stieltjes derivált általánosítása, amelyet, mint a Stieltjes integrál inverz műveletét értelmezték az [5], [19] munkákban.

A Stieltjes derivált képzése az alábbi, feltételezve, hogy $\dot{x}(t) \neq 0$:

$$\frac{df(x(t))}{dx(t)}=\frac{\dot{f}(x(t))}{\dot{x}(t)}=\nabla f(\boldsymbol{x}(t)).$$

A funcionális derivált a függvény paraméter szerinti deriváltját viszonyítja a folyamat állapothatározó-sebességeihez.

9. Mintafeladatok

A következőkben két, egyszerű feladat megoldási módját mutatjuk be. Ezek a példák illusztratív jellegűek és az általunk ajánlott eljárás számpéldákon való követését teszik lehetővé. Mechanikai tartalmuk nincs, hiszen ebben az esetben a méretek nagyon nagyok lennének és így kézi számításra áttekinthetetlenné válnának.

a. tekintsük a következő feladatot:

$$
\begin{array}{ll}
1 & x_1(t) + 4x_2(t) - 5\sin t = 0 \\
6 & -x_1(t) \le 0 \\
7 & -x_2(t) \le 0 \\
 & \min x_1(t) - 3x_2(t)
\end{array} \tag{48}
$$

A (48) feladathoz tartozó Fritz–John optimalitási feltétel rendszer az alábbi:

$$
\begin{array}{ll}
1 & x_1(t) + 4x_2(t) = 5\sin t \\
2 & u_1(t) - u_2(t) + 1 = 0 \\
3 & 4u_1(t) - u_3(t) - 3 = 0 \\
4 & u_2(t)x_1(t) = 0 \\
5 & u_3(t)x_2(t) = 0 \\
6 & x_1(t) \ge 0 \\
7 & x_2(t) \ge 0 \\
8 & u_2(t) \ge 0 \\
9 & u_3(t) \ge 0
\end{array} \tag{49}
$$

A (2) összefüggést felhasználva a (49) Fritz–John optimalitási feltétel rendszer a következő formában írható, ha a bázist jelentő polinomrendszert a Fourier polinomoknak vesszük fel és közelítésként az első négy tagot használjuk:

$$
\begin{aligned}
1. \quad & \left(K_1 + \sum_{i=1,3} \alpha_{1i}\sin(it) + \sum_{i=2,4} \alpha_{1i}\cos(it)\right) + \\
& + 4\left(K_2 + \sum_{i=1,3} \alpha_{2i}\sin(it) + \sum_{i=2,4} \alpha_{2i}\cos(it)\right) = 0 + 5\sin t \\
2. \quad & \left(k_1 + \sum_{i=1,3} \beta_{1i}\sin(it) + \sum_{i=2,4} \beta_{1i}\cos(it)\right) -
\end{aligned} \tag{50}
$$

$$-\left(k_2+\sum_{i=1,3}\beta_{2i}\sin(it)+\sum_{i=2,4}\beta_{2i}\cos(it)\right)+1=0$$

3. $$4\left(k_1+\sum_{i=1,3}\beta_{1i}\sin(it)+\sum_{i=2,4}\beta_{1i}\cos(it)\right)-$$
$$-\left(k_3+\sum_{i=1,3}\beta_{2i}\sin(it)+\sum_{i=2,4}\beta_{3i}\cos(it)\right)-3=0$$

4. $$\left(k_2+\sum_{i=1,3}\beta_{2i}\sin(it)+\sum_{i=2,4}\beta_{2i}\cos(it)\right)$$
$$\left(K_1+\sum_{i=1,3}\alpha_{1i}\sin(it)+\sum_{i=2,4}\alpha_{1i}\cos(it)\right)=0$$

5. $$\left(k_3+\sum_{i=1,3}\beta_{3i}\sin(it)+\sum_{i=2,4}\beta_{3i}\cos(it)\right)$$
$$\left(K_2+\sum_{i=1,3}\alpha_{2i}\sin(it)+\sum_{i=2,4}\alpha_{2i}\cos(it)\right)=0$$

6. $$\left(K_1+\sum_{i=1,3}\alpha_{1i}\sin(it)+\sum_{i=2,4}\alpha_{1i}\cos(it)\right)\geq 0$$

7. $$\left(K_2+\sum_{i=1,3}\alpha_{2i}\sin(it)+\sum_{i=2,4}\alpha_{2i}\cos(it)\right)\geq 0$$

8. $$\left(k_2+\sum_{i=1,3}\beta_{2i}\sin(it)+\sum_{i=2,4}\beta_{2i}\cos(it)\right)\geq 0$$

9. $$\left(k_3+\sum_{i=1,3}\beta_{3i}\sin(it)+\sum_{i=2,4}\beta_{3i}\cos(it)\right)\geq 0$$

A Fritz–John optimalitási feltétel rendszer a $\boldsymbol{f}$ térben az alábbi, ha az egyenlőtlenségi feltételeknél polinomok függvényértékeit a $k\pi/4$, $k=0,1,\ldots,7$ pontokban tekintve és a 4., 5. komplementaritási feltételeknél a polinomszorzatok integráljait (p_{ijk}) kiszámítva

(51) 1. $K_1 + 4K_2 = 0$ 2. $k_1 - k_2 + k_3 + 1 = 0$ 3. $4k_1 - k_3 - 3 = 0$

$\alpha_{11} + 4\alpha_{21} = 5$ $\beta_{11} - \beta_{21} + \beta_{31} = 0$ $4\beta_{11} - \beta_{31} = 0$

$\alpha_{12} + 4\alpha_{22} = 0$ $\beta_{12} - \beta_{22} + \beta_{32} = 0$ $4\beta_{12} - \beta_{32} = 0$

$\alpha_{13} + 4\alpha_{23} = 0$ $\beta_{13} - \beta_{23} + \beta_{33} = 0$ $4\beta_{13} - \beta_{33} = 0$

$\alpha_{14} + 4\alpha_{24} = 0$ $\beta_{14} - \beta_{24} + \beta_{34} = 0$ $4\beta_{14} - \beta_{34} = 0$

4. $k_2K_1 + \pi(k_2\alpha_{11} + K_1\beta_{21}) + \pi/2(-\beta_{21}\alpha_{14} - \beta_{22}\alpha_{13} + \beta_{23}\alpha_{12} + \beta_{24}\alpha_{11}) = 0$

$k_2K_1 + \pi(k_2\alpha_{12} + K_1\beta_{22}) + \pi/2(-\beta_{21}\alpha_{13} + \beta_{22}\alpha_{14} - \beta_{23}\alpha_{11} + \beta_{24}\alpha_{12}) = 0$

$k_2K_1 + \pi(k_2\alpha_{13} + K_1\beta_{23}) + \pi/2(\beta_{21}\alpha_{12} - \beta_{22}\alpha_{11} + \beta_{23}\alpha_{14} - \beta_{24}\alpha_{13}) = 0$

$k_2K_1 + \pi(k_2\alpha_{14} + K_1\beta_{24}) + \pi/2(\beta_{21}\alpha_{11} + \beta_{22}\alpha_{12} - \beta_{23}\alpha_{13} - \beta_{24}\alpha_{14}) = 0$

5. $k_3K_2 + \pi(k_3\alpha_{21} + K_2\beta_{31}) + \pi/2(-\beta_{31}\alpha_{24} - \beta_{32}\alpha_{23} + \beta_{33}\alpha_{22} + \beta_{34}\alpha_{21}) = 0$

$k_3K_2 + \pi(k_3\alpha_{22} + K_2\beta_{32}) + \pi/2(-\beta_{31}\alpha_{23} + \beta_{32}\alpha_{24} - \beta_{33}\alpha_{21} + \beta_{34}\alpha_{22}) = 0$

$k_3K_2 + \pi(k_3\alpha_{23} + K_2\beta_{33}) + \pi/2(\beta_{31}\alpha_{22} - \beta_{32}\alpha_{21} + \beta_{33}\alpha_{24} - \beta_{24}\alpha_{23}) = 0$

$k_3K_2 + \pi(k_3\alpha_{24} + K_2\beta_{34}) + \pi/2(\beta_{31}\alpha_{21} + \beta_{32}\alpha_{22} - \beta_{33}\alpha_{23} - \beta_{34}\alpha_{24}) = 0$

6. $K_1 + \alpha_{12} + \alpha_{14} \geq 0$

$K_1 + \sqrt{2}/2\alpha_{11} + \sqrt{2}/2\alpha_{12} + \alpha_{13} \geq 0$

$K_1 + \alpha_{11} - \alpha_{14} \geq 0$

$K_1 + \sqrt{2}/2\alpha_{11} - \sqrt{2}/2\alpha_{12} - \alpha_{13} \geq 0$

$K_1 - \alpha_{12} + \alpha_{14} \geq 0$

$K_1 - \sqrt{2}/2\alpha_{11} - \sqrt{2}/2\alpha_{12} + \alpha_{13} \geq 0$

$K_1 - \alpha_{11} - \alpha_{14} \geq 0$

$K_1 - \sqrt{2}/2\alpha_{11} + \sqrt{2}/2\alpha_{12} - \alpha_{13} \geq 0$

7. $K_2 + \alpha_{22} + \alpha_{24} \geq 0$

$K_2 + \sqrt{2}/2\alpha_{21} + \sqrt{2}/2\alpha_{22} + \alpha_{23} \geq 0$

$K_2 + \alpha_{21} - \alpha_{24} \geq 0$

$K_2 + \sqrt{2}/2\alpha_{21} - \sqrt{2}/2\alpha_{22} - \alpha_{23} \geq 0$

$K_2 + \alpha_{22} + \alpha_{24} \geq 0$

$K_2 - \sqrt{2}/2\alpha_{21} - \sqrt{2}/2\alpha_{22} + \alpha_{23} \geq 0$

$K_2 - \alpha_{21} + \alpha_{24} \geq 0$

$K_2 - \sqrt{2}/2\alpha_{21} + \sqrt{2}/2\alpha_{22} - \alpha_{23} \geq 0$

$$
\begin{array}{lrrrl}
8. & K_2 & & +\beta_{22} & +\beta_{24} \geq 0 \\
& k_2 & +\sqrt{2}/2\beta_{21} & +\sqrt{2}/2\beta_{22} & +\beta_{23} \geq 0 \\
& k_2 & +\beta_{21} & & -\beta_{24} \geq 0 \\
& k_2 & +\sqrt{2}/2\beta_{21} & -\sqrt{2}/2\beta_{22} & -\beta_{23} \geq 0 \\
& k_2 & & -\beta_{22} & +\beta_{24} \geq 0 \\
& k_2 & -\sqrt{2}/2\beta_{21} & -\sqrt{2}/2\beta_{22} & +\beta_{23} \geq 0 \\
& k_2 & -\beta_{21} & & -\beta_{24} \geq 0 \\
& k_2 & -\sqrt{2}/2\beta_{21} & +\sqrt{2}/2\beta_{22} & -\beta_{23} \geq 0
\end{array}
$$

$$
\begin{array}{lrrrl}
9. & k_3 & & +\beta_{32} & +\beta_{34} \geq 0 \\
& k_3 & +\sqrt{2}/2\beta_{31} & +\sqrt{2}/2\beta_{32} & +\beta_{33} \geq 0 \\
& k_3 & +\beta_{31} & & -\beta_{34} \geq 0 \\
& k_3 & +\sqrt{2}/2\beta_{31} & -\sqrt{2}/2\beta_{32} & -\beta_{33} \geq 0 \\
& k_3 & & -\beta_{32} & +\beta_{34} \geq 0 \\
& k_3 & -\sqrt{2}/2\beta_{31} & -\sqrt{2}/2\beta_{32} & -\beta_{33} \geq 0 \\
& k_3 & -\beta_{31} & & +\beta_{34} \geq 0 \\
& k_3 & -\sqrt{2}/2\beta_{31} & +\sqrt{2}/2\beta_{32} & -\beta_{33} \geq 0
\end{array}
$$

Az $(F1)$ feladat megoldásának közelítését az (F4) rendszerből meghatározott α együtthatókkal és K konstansokkal adjuk meg

$$
x_1 = \left(K_1 + \sum_{i=1,3} \alpha_{1i} \sin(it) + \sum_{i=2,4} \alpha_{1i} \cos(it) \right),
$$

$$
x_2 = \left(K_2 + \sum_{i=1,3} \alpha_{2i} \sin(it) + \sum_{i=2,4} \alpha_{2i} \cos(it) \right) \quad \text{formában.}
$$

b. tekintsük a következő feladatot:

$$
\begin{array}{ll}
(52) & 1 \quad x_1(t) + 4x_2(t) - 5\sin t = 0 \\
& \min x_1(t)x_2(t)
\end{array}
$$

Az (52) feladathoz tartozó Fritz–John optimalitási feltételek

$$
\begin{array}{ll}
(53) & 1 \quad x_1(t) + 4x_2(t) = 5\sin t, \\
& 2 \quad u_1(t) + x_2(t) = 0, \\
& 3 \quad 4u_1(t) + x_1(t) = 0.
\end{array}
$$

A (2) összefüggést felhasználva a (53) Fritz–John optimalitási feltételek a bázist jelentő polinomrendszert a Fourier polinomoknak véve és közelítésként az első négy tagot használva:

$$
\begin{aligned}
(54)\quad 1 \quad & \left(K_1 + \sum_{i=1,3} \alpha_{1i} \sin(it) + \sum_{1=2,4} \alpha_{1i} \cos(it)\right) + \\
& + 4\left(K_2 + \sum_{i=1,3} \alpha_{2i} \sin(it) + \sum_{i=2,4} \alpha_{2i} \cos(it)\right) = 0 + 5 \sin t, \\
2 \quad & \left(k_1 + \sum_{i=1,3} \beta_{1i} \sin(it) + \sum_{i=2,4} \beta_{1i} \cos(it)\right) + \\
& + \left(K_2 + \sum_{i=1,3} \alpha_{2i} \sin(it) + \sum_{i=2,4} \alpha_{2i} \cos(it)\right) = 0, \\
3 \quad & 4\left(k_1 + \sum_{i=1,3} \beta_{1i} \sin(it) + \sum_{i=2,4} \beta_{1i} \cos(it)\right) + \\
& + \left(K_1 + \sum_{i=1,3} \alpha_{1i} \sin(it) + \sum_{i=1,4} \alpha_{1i} \cos(it)\right) = 0.
\end{aligned}
$$

Az (54) optimalitási rendszer a $\boldsymbol{f}$ térben

$$
\begin{array}{lllll}
(55) & 1. & K_1 + 4K_2 = 0 & 2. \quad k_1 + K_2 = 0 & 3. \quad 4k_1 + K_1 = 0 \\
& & \alpha_{11} + 4\alpha_{21} = 5 & \beta_{11} + \alpha_{21} = 0 & 4\beta_{11} + \alpha_{11} = 0 \\
& & \alpha_{12} + 4\alpha_{22} = 0 & \beta_{12} + \alpha_{22} = 0 & 4\beta_{12} + \alpha_{12} = 0 \\
& & \alpha_{13} + 4\alpha_{23} = 0 & \beta_{13} + \alpha_{23} = 0 & 4\beta_{13} + \alpha_{13} = 0 \\
& & \alpha_{14} + 4\alpha_{24} = 0 & \beta_{14} + \alpha_{24} = 0 & 4\beta_{14} + \alpha_{14} = 0
\end{array}
$$

Az (55) rendszer a következő matematikai programozási feladat optimalitási feltételeivel egyezik meg:

$$
\begin{aligned}
(56)\quad 1. \quad & K_1 + 4K_2 = 0 \\
& \alpha_{11} + 4\alpha_{21} = 5 \\
& \alpha_{12} + 4\alpha_{22} = 0 \\
& \alpha_{13} + 4\alpha_{23} = 0 \\
& \alpha_{14} + 4\alpha_{24} = 0 \\
& \min\left(K_1 K_2 + \alpha_{11}\alpha_{21} + \alpha_{12}\alpha_{22} + \alpha_{13}\alpha_{23} + \alpha_{14}\alpha_{24}\right)
\end{aligned}
$$

illetve az (56) feladat duáljának optimalitási feltétele szintén az (55) rendszer:

$$(57) \qquad \begin{array}{lll} 2. & k_1 + K_2 = 0 & 3. \quad 4k_1 + K_1 = 0 \\ & \beta_{11} + \alpha_{21} = 0 & \quad 4\beta_{11} + \alpha_{11} = 0 \\ & \beta_{12} + \alpha_{22} = 0 & \quad 4\beta_{12} + \alpha_{12} = 0 \\ & \beta_{13} + \alpha_{23} = 0 & \quad 4\beta_{13} + \alpha_{13} = 0 \\ & \beta_{14} + \alpha_{24} = 0 & \quad 4\beta_{14} + \alpha_{14} = 0 \end{array}$$

$$\max\left(-5\beta_{11} - K_1K_2 - \alpha_{11}\alpha_{21} - \alpha_{12}\alpha_{22} - \alpha_{13}\alpha_{23} - \alpha_{14}\alpha_{24}\right)$$

Ez a munka részben az OTKA-5313 és OTKA-2568 számú szerződések támogatásával készült.

Összefoglalás

Lokális egyensúlyi állapotban az állapotváltozás vizsgálatára adtunk egy módszert, amely akkor is használható, ha az állapothatározók sebességei függnek az időtől.

Az „útkövető" eljárások helyett a problémát parametrikus optimalizálási feladatként fogalmazzuk meg. Beláttuk, hogy a L^2 térben felírt feladat esetében igaz a Fritz–John tétel. A feladatot transzformáltuk a ℓ^2 térbe az izomorfia tétel alapján. Az egyenlőtlenségi feltételek átvitelére csak becslést tudunk adni, mivel az egyenlőtlenség nem leképezés tartó. Végül az egyenlőtlenségeket is tartalmazó L^2 térbeli feladatot a ℓ^2 térben egy semi-infinit egyenlőtlenség rendszer megoldására vezettük vissza, míg a csak egyenlőségi feltételeket tartalmazó feladatokhoz egy semi-infinit matematikai programozási feladatpár rendelhető a ℓ^2 térben.

IRODALOM

[1] ABRAHAM, R., MARSDEN, J. E., RATIU T., „Manifolds", *Tensor Analysis and Applications* (Springer Verlag, New York, Berlin, 1988).

[2] BAZARAA, M. S., SHETTY C. M., *Nonlinear Programming Theory and Algorithms* (John Wiley & Sons, New York, 1979).

[3] BHAKTA, P. C., ROYCHANDHURI S., „Optimization in Banach Spaces", *Jour of Math. Analysis and Applications* **134** (1988), 460–470.

[4] COHN, M. Z., MAIER G., *Engineering Plasticity by Mathematical Programming* (Pergamon Press Inc. Waterloo, 1979).

[5] DANIELL P. J., „Differentiation with Respect to a Function of Limited Variation", *Trans. Amer. Math. Soc.* **19** (1918), 353–362.

[6] DANCS I., *Az L^2 és ℓ^2 terek közötti izomorfia nem rendezéstartó*, Privát közlemény (1992).

[7] FIACCO, A. V., MCCORMICK, G. P., „Nonlinear Programming", *Sequential Unconstrained Minimization Techniques* (Wiley and Sons, New York, 1968).

[8] HIRSCH, M. W., *Differencial Topology* (Springer-Verlag, 1976).

[9] JONGEN, H. TH., JONKER, P., TWILT, F., „Critical Sets in Parametric Optimization", *Mathematical Programming* **34** (1986), 333–353.

[10] KALISZKY S., VÁSÁRHELYI A., LÓGÓ J., „The Time History Analysis of Viscoelastic Structures by Mathematical Programming", *Advances in Continuum Mechanics* (O. Bruller, V. Mannl, J. Najar, eds.) (Springer Verlag, 1991), 488–499.
[11] KALISZKY S., *Képlékenységtan* (Akadémiai Kiadó, Budapest, 1975).
[12] KESTIN, J., *A Course in Thermodynamics*, vol. 1 sect. 8.4.9 (Hemisphere Publ. Coop., Washington, 1979).
[13] KOJIMA, M., HIRABAYASHI, R., „Continuous Deformation of Nonlinear Programs", *Mathematical Programming Study* **21** (1984), 150–198.
[14] MILNOR, J., *Morse Thery* (Pronceton University Press, Princeton, 1963).
[15] MIKOLÁS M., *Valós függvénytan és ortogonális sorok* (Akadémiai Kiadó, Budapest, 1970).
[16] MUSHIK, W., *Aspects of Nonequilibrium Thermodynamics*, sets 1.1.1 (World Scientific, Singapore, 1990).
[17] ROOS, E. C., VIAL PH. J. , (eds), „Interior Point Methods for Linear Programming", Theory and Practice, *Mathematical Programming, Series B.* **52** (1991).
[18] SZÖKEFALVI-NAGY B., *Valós függvények és függvénysorok* (Tankönyvkiadó, Budapest, 1961).
[19] YOUNG, W. H., „On Integrals and Derivates with Respect to a Function", *Proc. London Math. Soc.* (1914–15), 35–63.

(Beérkezett: 1992. május 21.)

(Átdolgozva beérkezett: 1994. május 18.)

VÁSÁRHELYINÉ SZABÓ ANNA
BME ÉPÍTŐIPARI SZÁMÍTÓKÖZPONT
1111 BUDAPEST, MŰEGYETEM RKP. 3.

MATHEMATICAL MODEL OF ANALYSIS OF STRUCTURES IN THE CASE OF LOCAL EQUILIBRIUM PROCESSES

VÁSÁRHELYINÉ, A. SZABÓ

Generally, path-following algorithms are used for historical analysis of structures. Now, a new approach is presented for solving the problem by parametric optimization.

The optimization problem is solved in a direct product of function spaces. The necessary condition of the stationarity of a curve are examined. A method is presented for determining a piece of a continuous component of the stationarity curve depending on one parameter which transforms the problem into the space ℓ^2.

Alkalmazott Matematikai Lapok **17** (1993) 321–328

VÉGES HALMAZON ÉRTELMEZETT FÜGGVÉNYEK *PR*-MAXIMÁLIS ÉS *PR*-TELJES KLÓNJAI*

BAGYINSZKI JÁNOS

Gödöllő-Budapest

A k-értékű logika függvényei halmazának hatványhalmazán egy — pr-lezárásnak nevezett — operációt értelmezünk. A k-értékű logika függvényeinek egy halmazát pr-zártnak, vagy pr-klónnak nevezzük, ha az egy olyan klón, amely zárt a primitív rekurzióra. (Klón a vetítőfüggvényeket tartalmazó olyan függvényhalmaz, amely zárt az összetett függvények képzésére.) Megmutatjuk, hogy az igazság-függvények ($k = 2$) esetén pontosan két pr-klón van. Az általános esetben a legfontosabb eredményünk a — Rosenberg-féle teljességi tételnek megfelelő — pr-teljességi alaptétel kimondása és bizonyítása.

1. Bevezetés

Az 1970-es és 80-as években a klónokkal kapcsolatos vizsgálatok iránt a számítástudomány és az algebra területén is megnőtt az érdeklődés. Bár az igazságfüggvények zárt osztályainak teljes tartalmazás-struktúráját Post 1941-ben leírta, $k > 2$ esetén a legfontosabb eredmények sorát Rosenberg teljességi tételének megjelenése (1965) indította el. (Itt k az alaphalmaz elemszáma.) Hogy a problémák lényegesen nehezebbek $k > 2$ esetén, az abból is sejthető, hogy $k = 2$ esetén a klónok halmaza megszámlálhatóan végtelen számosságú, $k > 2$ esetén ez a számosság nem megszámlálható. A klón-lezárásnál erősebb pr-lezárást abból a célból vezettük be [1,2], hogy a klón-háló bizonyos részhálójának vizsgálatával teljesebb képet kapjunk a klón-hálóról is. Másrészről a pr-lezárás bevezetésével a számítástudományban természetes rekurzióval való függvényképzési módot emeltünk át a k-értékű logikába, szorosabb kapcsolatot teremtve ezzel a két terület között. A klónokra fölvethető legtöbb kérdés megfogalmazható pr-klónokra (és r-klónokra) is. Következő dolgozatokban szeretnénk foglalkozni a klón-háló számosságával, ill. a pr-klónok kvázi-primál jellemzésével.

A 2. részben a definíciók és jelölések megadása után utalunk az általánosítás lehetőségére (rekurzív lezárás), ill. a primitív rekurziónak egy másik lehetséges megfogalmazására (ciklikus rekurzió) véges alaphalmaz esetén.

A 3. részben megadjuk az igazság-függvények pr-klónjainak hálóját, amely meglepően egyszerű: kételemű lánc. Megmutatjuk, hogy mindkét osztály egy-egy elemmel generálható, sőt, minden igazság-függvény generálja a két osztály egyikét.

A 4. részben előkészítjük a fő eredmény bizonyítását, amelyet az 5. rész tartalmaz.

*A dolgozat az OTKA T4295/92 sz. téma keretében készült.

Az 5. részben megadunk egyváltozós *pr*-Sheffer függvényeket, valamint a Słupecki-tétel egy *pr*-megfelelőjét. Megmutatjuk továbbá azt a (nem meglepő) tényt, hogy az O_k függvényosztályban minden függvény primitív rekurzív. Fő eredményként meghatározzuk a *pr*-maximális osztályokat és megadjuk a *pr*-teljesség alaptételét.

Megjegyezzük, hogy a [2] dolgozatban a 3.8 tétel hibás és ezért részben hibás a 3.9 és 3.10 tétel is. A helyes eredményeket a jelen dolgozat 1. tétele tartalmazza.

Végezetül köszönetet szeretnék mondani Csákány Bélának és Czédli Gábornak értékes észrevételeikért.

2. Definíciók és jelölések

Legyenek k, m, n, i, ℓ nem-negatív egész számok, $k > 1$ rögzített, $m, n, i > 0$, $\ell \geq 0$. Jelölje $O_K^{(n)}$ a véges vagy végtelen K alaphalmazon értelmezett n-változós függvények halmazát: $O_K^{(n)} := \{f : K^n \to K\}$, és legyen $O_K := \bigcup_{n>0} O_K^{(n)}$. Tetszőleges $O \subseteq O_K$ részhalmazra legyen $O^{(n)} := O \cap O_K^{(n)}$ és jelölje $f_{|M}$ az f függvény leszűkítését az $M(\subset K)$ halmazra: $f_{|M} : M^n \to K$, azaz $f_{|M}(\tilde{x}) = f(\tilde{x})$, ha $\tilde{x} := (x_1, \ldots, x_n) \in M^n$. Néhányszor alkalmazzuk az $(\tilde{x}, x_{n+1}) := (x_1, \ldots x_n, x_{n+1})$ jelölést, $K = \{0, 1, \ldots, k-1\}$ esetén $\oplus$ a mod k összeadást $K = \{0, 1, \ldots\}$ esetén az összeadást jelenti. A továbbiakban *függvényen*, ill. O halmazon az O_K egy elemét, ill. egy részhalmazát értjük. A következőkben néhány speciális függvénynek külön nevet és jelet adunk. Az $s \in O_K^{(1)}$ *ciklikus permutáció-függvényt*, a c_ℓ^n $(\ell \in K)$ *n-változós állandófüggvényeket* és az e_i^n *i-edik vetítőfüggvényt* a következőképp definiáljuk: minden $x \in K$, $\tilde{x} \in K^n$ esetén legyen $s(x) = x \oplus 1$, $c_\ell^n(\tilde{x}) = \ell$ és $e_i^n(\tilde{x}) = x_i$. A felső indexeket $n = 1$ esetén elhagyjuk: $e_1 := e_1^1$, $c_\ell := c_\ell^1$. Jelölje rendre az állandó-függvények és a vetítő-függvények halmazát C és E: $C^{(n)} := \{c_\ell^n \mid \ell \in K\}$, $E^{(n)} := \{e_i^n \mid 1 \leq i \leq n\}$. Legyen továbbá minden $x, y \in K$, $\tilde{x} \in K^n$ esetén

$$\max(x,y) = \begin{cases} x, & \text{ha } x \geq y, \\ y, & \text{ha } x < y; \end{cases} \qquad \min(x,y) = \begin{cases} y, & \text{ha } x \geq y, \\ x, & \text{ha } x < y; \end{cases}$$

$$\text{és} \quad w'(\tilde{x}, y) = \begin{cases} x_{\ell+1}, & \text{ha } y = \ell \in K, \\ y, & \text{ha } y = \ell \notin K. \end{cases}$$

Azt mondjuk, hogy az f függvény *megőrzi a* (nem üres) $M(\subset K)$ *halmazt*, ha $f_{|M}$ függvényértékei M-beliek: minden $\tilde{x} \in M^n$ esetén $f_{|M}(\tilde{x}) \in M$. Jelölje $(O \mid M)$ az M *halmazt megőrző* O-beli függvények halmazát:

$$(O \mid M)^{(n)} := \{f \in O^{(n)} \mid f_{|M}(\tilde{x}) \in M, \text{ ha } \tilde{x} \in M^n\}.$$

Jelöljön $[\![\ ]\!]$ az O_K hatványhalmazán értelmezett lezárási operációt és O egy $[\![\ ]\!]$-re nézve *zárt osztályt*: $[\![O]\!] = O \subseteq O_K$. Az O osztály egy O' részhalmaza $[\![\ ]\!]$-*teljes* az

O halmazban, ha $[\![O']\!] = O$. Ha O'' és O' $[\![\ ]\!]$-zárt részhalmazai az O halmaznak és $O'' \subset O' \subseteq O$ esetén $O' = O$, akkor az O'' osztályt O *halmazban* $[\![\ ]\!]$-*maximális* osztálynak nevezzük.

Ha B $[\![\ ]\!]$-teljes az O osztályban és minden $B' \subset B$ valódi részhalmazra $[\![B']\!] \neq O$, akkor B halmazt az O *bázisának* nevezzük.

A vetítő-függvények E halmazát tartalmazó $[\![\ ]\!]$-zárt osztályokat $[\![\ ]\!]$-*klónoknak* nevezzük.

A klónok, a véges algebrák és a k-értékű logikák elméletében alkalmazott lezárási operációk nyerhetők az $\mathfrak{N} := \{\zeta, \tau, \Delta, \nabla, *, \$, \varepsilon\}$ művelethalmaz bizonyos részhalmazai által meghatározott lezárásokként. A $\mathfrak{N}$ művelethalmaz elemeit a következő egyenlőségek definiálják. Legyen minden $f \in O_K^{(n)}$ és $g \in O_K^{(m)}$ függvényre és $\tilde{x} \in K^n$ n-esre igaz, hogy:

(1)(*a*) $n = 1$ eset: $(\zeta f) = (\tau f) = (\Delta f) = f$,

$n > 1$ eset:

(*b*) $(\zeta f)(\tilde{x}) = f(x_2, x_3, \ldots, x_n, x_1)$,

(*c*) $(\tau f)(\tilde{x}) = f(x_2, x_1, x_3, \ldots, x_n)$,

(*d*) $(\Delta f)(\tilde{x}) = f(x_2, x_2, x_3, \ldots, x_n)$,

(2) $(\nabla f)(\tilde{x}, x_{n+1}) = f(\tilde{x})$,

(3) $(g * f)(\tilde{x}, x_{n+1}, \ldots, x_{n+m-1}) = g(f(\tilde{x}), \ldots, x_{n+m-1})$,

(4) $(g\$f)(\tilde{x}, 0) = f(\tilde{x})$,

$(g\$f)(\tilde{x}, s(i)) = g(\tilde{x}, (g\$f)(\tilde{x}, i), i)$, $s(i) \in K$, $m = n + 2$,

(5) $(\varepsilon f) = e_1^2$.

A (3), (4) és (5) alatt megadott operációkat rendre az O_K halmazon ható *kompozíciónak, primitív rekurziónak, konstans-operációnak* nevezzük.

A következő táblázatban néhány lezárást ($\mathfrak{M} \subseteq \mathfrak{N}$) és a megfelelő $[\ ]_{\mathfrak{M}}$-zárt osztályt írjuk le.

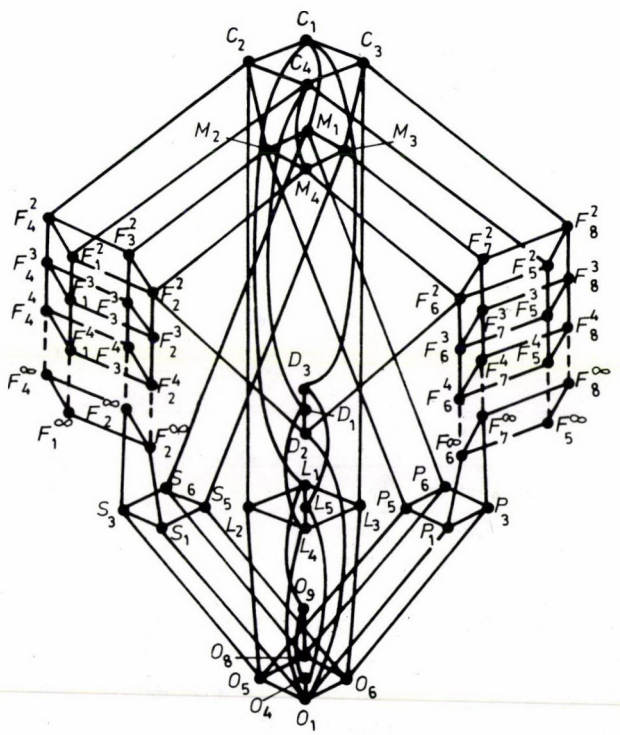
$\mathfrak{M}$	$[\ ]_{\mathfrak{M}}$-zárt osztály	$[\ ]_{\mathfrak{M}}$ jelölése
$\{\zeta, \tau, \Delta, *\}$	Post-zárt osztály	$[\ ]_{\mathrm{p}}$
$\{\zeta, \tau, \Delta, \nabla, *\}$	zárt osztály vagy: iteratív osztály	$[\ ]_{\mathrm{i}}$
$\{\zeta, \tau, \Delta, \nabla, \varepsilon\}$	klón	$[\ ]_{\mathrm{cl}}$
$\mathfrak{N}$	*pr*-zárt osztály vagy: *pr*-klón	$[\ ]_{pr}$

A definíciókból következik, hogy bármely nem-üres $O \subset O_K$ halmazra érvényesek az $[O]_{\mathrm{p}} \subseteq [O]_{\mathrm{i}} \subseteq [O]_{\mathrm{cl}} \subseteq [O]_{pr}$ tartalmazások, mert például $(\Delta f)(\tilde{x}) = (f * e_2^2)(\tilde{x})$. Az ε operációt a klón definíciójának egyszerűsítése céljából vezettük be. Bár ezek a definíciók véges és megszámlálhatóan végtelen halmazra egyaránt

érvényesek, az eredményeket csak a $K := \{0, 1, \ldots, k-1\}$ rögzített véges halmazra fogalmazzuk meg, annak ellenére, hogy néhány eredmény általánosabban is igaz. A szóbanforgó O_K helyett O_k-t írunk.

3. Az igazság-függvények *pr*-klónjainak hálója ($k = 2$)

A *pr*-lezárás hatékonyságának érzékeltetése céljából összehasonlításként megadjuk az igazság-függvények (más szavakkal: logikai függvények) klónjainak Post-féle hálóját (1. ábra) [3] és a *pr*-klónjainak hálóját (2. ábra). A Post-hálón megtartottuk Post eredeti jelöléseit, így C_1 a logikai függvények halmazát, C_3 pedig a $\{0\}$-őrző logikai függvények halmazát jelöli. Valójában Postnál nem klónok, hanem Post-zárt osztályok szerepelnek, s azok diagramja nem háló. Azonban a mi szempontunkból az eltérés nem lényeges.



1. ábra

2. ábra

A *pr*-zárt klónok (két-elemű) háló-diagramját a következő tétel alapján rajzoltuk fel.

1. TÉTEL. (1) *Az* $(O \mid \{0\})$ *halmaz pr-maximális az* O_2 *osztályban.*
(2) *Minden egyes* $\{0\}$*-őrző igazság-függvény pr-teljes az* $(O_2 \mid \{0\})$ *halmazban.*
(3) *Ha egy igazság-függvény nem* $\{0\}$*-őrző, akkor pr-teljes (az* O_2 *osztályban).*

Bizonyítás. (1) Ismeretes [4], hogy az $(O_2 \mid \{0\})$ osztály (klón-zárt, és) klón maximális (az O_2 osztályban). Továbbá, ez az osztály zárt a primitív rekurzióra is, mert $(g\$f)(\tilde{0}, 0) = f(\tilde{0}) = 0$, ha $f \in (O_2 \mid \{0\})$. Ezért $(O_2 \mid \{0\})$ *pr*-maximális klón (O_2-ben).
(2) Legyen $g_2, g_3 \in O_2$ a következő módon primitív rekurzióval definiálva (az E halmazon!): $g_2(x,0) = e_1(x)$, $g_2(x,1) = e_3^3(x, e_1(x), 0)$ és $g_3(x,y,0) = e_1^2(x,y)$, $g_3(x,y,1) = e_2^4(x,y,x,0)$. Könnyen látható, hogy $g_2(x,y) = x \oplus xy$, $g_3(x,y,z) = x \oplus xz \oplus yz$, ezért $g_2(x,x) = c_0(x)$, $g_3(x,y,x) = xy$ és $g_3(x, g_2(y,x), y) = x \oplus y$, így fennállnak a következő tartalmazások:

$$(O_2 \mid \{0\}) = [\{x \oplus y, xy\}]_{\mathrm{cl}} \subseteq [\{x \oplus y, xy\}]_{pr} \subseteq [E]_{pr} \subseteq [(O_2 \mid \{0\})]_{pr} = (O_2 \mid \{0\}).$$

Ez azt jelenti, hogy mindenhol egyenlőség van, vagyis $[E]_{pr} = (O_2 \mid \{0\})$.
(3) Definíció szereint a $g \in O_2 \setminus (O_2 \mid \{0\})$ logikai függvényekre igaz, hogy $g(\tilde{0}) = 1$. Ezért $c_1 = g(c_0, \ldots, c_0) \in [\{g\}]_{pr}$, így $g_3(e_1, c_1) = e_1 \oplus c_1 \in [\{g\}]_{pr}$. Azonban jól ismert, hogy $\{xy, x \oplus 1\}$ klón-bázis O_2-ben, így $\{g\}$ *pr*-teljes az O_2 osztályban. □

4. A *pr*-lezárás néhány tulajdonsága

Bevezetünk néhány függvényt és megmutatjuk, hogy — egyikük kivételével — minden *pr*-zárt osztály tartalmazza ezeket a függvényeket. Legyenek r_1, r_2, r_3, r, $w, j_\ell \in O_k$ azok a függvények, amelyeket a minden $x, y, z, x_1, \ldots, x_n \in K$ esetén fennálló alábbi egyenlőségek definiálnak:

$$r_1(x) = \begin{cases} 0, & \text{ha } x = 0, \\ x - 1, & \text{ha } x \neq 0; \end{cases} \qquad r_2(x,y) = \begin{cases} x, & \text{ha } y = 0, \\ y - 1, & \text{ha } y \neq 0; \end{cases}$$

$$r_3(x,y,z) = \begin{cases} x, & \text{ha } z = 0, \\ y, & \text{ha } z \neq 0; \end{cases} \qquad r(x,y) = \begin{cases} x, & \text{ha } y = 0, \\ y, & \text{ha } y \neq 0 \text{ és } x = 0, \\ x - 1, & \text{ha } y \neq 0 \neq x; \end{cases}$$

$$j_\ell(x) = \begin{cases} k - 1, & \text{ha } x = \ell \\ 0, & \text{ha } x \neq 0 \end{cases} \qquad (\ell \in K).$$

Világos, hogy az $n = k$, $w(\tilde{x}, z) = w'_{|K}(\tilde{x}, z)$ egyenlőségek által definiált $w \in O_k^{(k+1)}$ függvényre a $w(\tilde{x}, \ell) = x_{\ell+1}$, $0 \leq \ell < k$ egyenlőségek fennálnak.

Megjegyzés. Szokásos (pl. a változók száma szerinti teljes indukciós bizonyításban) egy $f \in O_k^{(n+1)}$ függvényt $f(\tilde{x}, z) = \max(\min(j_\ell(z), f(\tilde{x}, \ell)) \mid \ell \in K)$ kifejezéssel reprezentálni. Azonban ez a reprezentáció a w függvény alkalmazásával a következő kifejezéssel is nyerhető: $f(\tilde{x}, z) = w(f(\tilde{x}, 0), \ldots, f(\tilde{x}, k-1), z)$. Ez utóbbi reprezentációt ebben a dolgozatban *sztenderd reprezentációként* fogjuk idézni.

1. SEGÉDTÉTEL. (α) *Minden pr-klón tartalmazza a* c_0, r_1, r_2, r_3, r *és* w *függvényeket.* (β) $c_0 \in [r_1]_{\mathrm{cl}}$, $r_1 \in [r_2]_{\mathrm{cl}}$, $r, w \in [r_2, r_3]_{\mathrm{cl}}$.

Bizonyítás. Definiáljuk a vetítő-függvények halmaza felett primitív rekurzióval az f_1 és f_2 függvényt a következő módon: $f_j(x, y, 0) = e_1^2(x, y)$, $f_j(x, y, s(i)) = e_{2j}^4(x, y, f(x, y, i), i)$, $i = 0, 1, \ldots, k-2$ (vagy a kompaktabb, operációs jelöléssel: $f_j := e_{2j}^4 \$ e_1^2$). Kapjuk: $r_2(x, y) = f_2(x, y, y)$, $r_3 = f_1$, $r_1(x) = r_2(x, x)$, $r(x, y) = r_3(x, r_2, (y, x), y)$. Az $r_1^{*1} := r_1$, $r_1^{*m+1} = r_1 * r_1^{*m}$ $(m \geq 1)$ kompozíció-hatványozás (iterálás) a c_0 függvényt eredményezi: $c_0 = r_1^{*k-1}$. A $w_2 := r_3$, $w_{i+1}(x_1, \ldots, x_i, y, z) := r_3(x_1, w_i(x_2, \ldots, x_i, y, r_1(z)), z)$, $i = 2, 3, \ldots$ rekurzív kompozíció (nem primitív rekurzió!) viszont a w függvényt eredményezi, mert könnyen láthatóan $w(x_1, \ldots, x_k, z) = w_{k+1}(x_1, \ldots, x_k, z, z)$. A konstrukciókból leolvasható a (β) állítás is. □

5. A *pr*-maximális halmazok és a *pr*-teljességi tétel

Az 1. tétel alapján érzékeltettük, hogy a *pr*-lezárás lényegesen erősebb, mint a klón-lezárás: amíg az igazság-függvények klón-hálója megszámlálhatóan végtelen számosságú (lásd [3]), addig a *pr*-klónok száma az igazságfüggvények esetében (azaz, ha $k = 2$) mindössze 2 (lásd 1. és 2. ábra). A következő tétel tetszőleges véges $k \geq 2$ esetén világít rá a *pr*-lezárás erejére (emlékeztetünk rá, hogy klón-lezárás esetén a Sheffer-függvények legalább két-változósak!).

2. TÉTEL. *A* c_{k-1} *és az* s *függvények mindegyike pr-Sheffer függvény (azaz, mindegyike önmagában pr-teljes).*

Bizonyítás. Mindkét függvényből előállíthatjuk az állandó függvények mindegyikét a következő egyszerű konstrukciókkal: $c_i = r_1^{*k-i-1}(c_{k-1})$ és $c_i = s^{*i}(c_0)$, $i = 0, 1, \ldots, k-1$, $(r_1^{*0} := s^{*0} := e_1)$. (Az 1. segédtételt alkalmaztuk.) A változók n számára vonatkozó indukcióval következik az állítás, mert bármely $f \in O_k$ függvény sztenderd reprezentációs alakban felírható: $f(\tilde{x}, z) = w(f(\tilde{x}, 0), \ldots, f(\tilde{x}, k-1), z)$. □

KÖVETKEZMÉNY. *Az állandó függvények* C *osztálya pr-teljes.*

Megjegyzések. A 2. tétel bizonyítása és a következmény szerint $g \in O^{(n)}$ esetén $g(r_1^{*\alpha_1}, \ldots, r_n^{*\alpha_n}) \in O^{(n)}$, ezért $\alpha_i = \ell - a_i \geq 0$ választással adódik:

1. A Słupecki-tétel *pr*-megfelelőjének tekinthető a következő *állítás*: A c_ℓ állandófüggvényt tartalmazó $O \subseteq O_k$ függvényhalmaz *pr*-teljes, ha létezik $g \in O^{(n)}$ és $\tilde{a} \in K^n$, amelyekre $g(\tilde{a}) = k-1$ és minden $1 \leq i \leq n$ esetén $a_i \leq \ell$.

2. Ha a $C \cup \{s\}$ és a $C \cup \{s\}]_{pr}$ halmazokat rendre az O_k-beli *elemi függvények*, ill. a *primitív rekurzív függvények* halmazának nevezzük, akkor az O_k függvényosztályban minden függvény primitív rekurzív.

A \$ (primitív rekurzió) definíciójából következik, hogy $(g\$f)(\tilde{0}, 0) = f(\tilde{0})$, ezért az $(O_k \mid \{0\})$ 0-megőrző osztály $k \geq 2$ esetén is *pr*-klón. Minthogy ez az osztáy klónmaximális [4], a következő megállapítás adódik:

ÁLLÍTÁS. *A 0-megőrző függvények osztálya pr-maximális az O_k osztályban.*

Ismeretes, hogy az $(O_k \mid M)$ osztály minden $M \subseteq K$ részhalmaz esetén klón [4]. Azonban, ezek általában nem *pr*-zárt osztályok, mert pl. az 1. segédtétel szerint $r_1 \in [(O_k \mid \{1\})]_{pr}$, de $r_1(1) = 0$, így $r_1 \notin (O_k \mid \{1\})$. Legyen $K_i := \{0, 1, \ldots, i\}$, ha $0 \leq i \leq k-1 (K_{k-1} = K)$.

A következő tételben a részhalmaz-megőrző klónok vannak jellemezve *pr*-teljesség és *pr*-maximalitás szempontjából.

3. TÉTEL. *A K alaphalmaz egy (nem-üres) M valódi részhalmazára az $(O_k \mid M)$ klón:*

(i) *pr-maximális klón (O_k-ban), ha $M = K_i$, $0 \leq i \leq k-2$,*

(ii) *pr-teljes osztály (O_k-ban), minden más esetben.*

Bizonyítás. Legyen M egy valódi részhalmaza K-nak és legyen $b \in K \setminus M$. Létezik egy $g \in (O_k^{(1)} \mid M)$ függvény, amelyre $g(b) = k-1$ teljesül.

(i) Az $(O_k \mid K_i)$, $i \neq k-1$ osztályok *pr*-klónok, amint az a \$ művelet definíciójából látható. Ezek az osztályok azonban nem tartalmazzák a c_{k-1} függvényt, így nem *pr*-teljesek (nem azonosak O_k-val). Bármely $f \in O_k^{(n)} \setminus (O_k \mid M)$ függvényhez létezik $\tilde{a} = (a_1, \ldots, a_n) \in M^n$, amelyekre $f(\tilde{a}) = b (\in K \setminus M)$. Ezért $c_{k-1} = g(f(c_{a_1}, \ldots, c_{a_n})) \in [(O_k \mid M) \cup f]_{pr}$, minthogy $c_{a_1}, \ldots, c_{a_n}, g \in (O_k \mid M)$. Így a 2. tétel alapján az $(O_k \mid M)$ osztály egy *pr*-maximális klón.

(ii) Tegyük fel, hogy M nem azonos a $K_0, K_1, \ldots, K_{k-1}$ halmazok egyikével sem. Ekkor létezik $b+1 \in M$, amelyre $b \in K \setminus M$. Minthogy $c_{b+1} \in (O_k \mid M)$, így $r_1(c_{b+1}) = c_b \in [(O_k \mid M)]_{pr}$, ezért $g(c_b) = c_{k-1} \in [(O_k \mid M)]_{pr}$. Tehát a 2. tétel alapján az $(O_k \mid M)$ osztály *pr*-teljes. □

A következő tétel a Rosenberg-féle teljességi tétel [4] *pr*-megfelelője, ezért alaptételnek nevezzük. Bizonyítása azonban — a felhasznált segédtételekkel együtt is sokkal egyszerűbb, annak köszönhetően, hogy a *pr*-lezárás „erősebb" a klónlezárásnál.

4. TÉTEL (a *pr*-teljesség alaptétele). *Minden $k \geq 2$ egész számra az $O \subseteq O_k$ függvényhalmaz pontosan akkor pr-teljes, ha az $O \setminus (O_k \mid K_i)$, $0 \leq i < k-1$ halmazok egyike sem üres halmaz.*

Bizonyítás. Világos, hogy a feltételek *szükségesek*, minthogy az $(O_k \mid K_i)$ halmazok *pr*-maximális klónok a 3. tétel szerint.

Elégségesség. Tegyük fel, hogy $O \setminus (O_k \mid K_i)$ nem-üres halmazok, $0 \leq i < k-1$. Ekkor létezik egy $g \in O$ függvény, amelyre $g(c_0, \ldots, c_0) = c_j \neq c_0$, és így $c_j \in [O]_{pr}$ az 1. segédtétel szerint. Ha $j < k-1$, akkor az előbbihez hasonló módon adódik a $\{g(c_{i_1}, \ldots, c_{i_n}) \mid g \in O, \{i_1, \ldots, i_n\} \subseteq K_j\}$ halmaz, amelynek létezik egy c_ℓ eleme, ahol $\ell > j$. Legfeljebb $k-1$ lépésben így generáljuk a c_{k-1} függvényt. Ezért az 1. segédtétel és a 2. tétel alapján az O osztály *pr*-teljes. □

IRODALOM

[1] BAGYINSZKI J., „A prím-értékű logikák zárt lineáris függvényosztályainak diagramja és a *pr*-lezárás tulajdonságai", *Kandidátusi értekezés* (1991), 1–88.

[2] BAGYINSZKI J., „Logikai függvények *pr*-zárt osztályainak hálója", *Alkalmazott Matematikai Lapok* **15** (1990–91), 289–302.

[3] PÖSCHEL, R., KALUZSNYIN, L. A., *Funktionen- und Relationenalgebren* (Deutcher Verlag der Wissenschaften, 1979).

[4] ROSENBERG, I. G., „Completeness properties of multiple-valued logic algebras", *Computer Science and Multiple-valued Logic.* (D. C. Rine, ed.) (North-Holland Publ. Co., 1977), 144–186.

(Beérkezett: 1994. május 11.)

BAGYINSZKI JÁNOS
KANDÓ KÁLMÁM MŰSZAKI FŐISKOLA
1034 BUDAPEST, NAGYSZOMBAT U. 19.

PR-MAXIMAL AND *PR*-COMPLETE CLONES OF FUNCTIONS DEFINED ON A FINITE SET

J. BAGYINSZKI

On the power-set of the set of k-valued logical functions there is defined a closure operation, called *pr*-closure. A set of k-valued logical functions is called *pr*-closed (or *pr*-clone), if it was a clone, closed under primitive recursion. (Clone is a set of functions containing the projections, which is closed under composition of functions.) We prove, that in the case of truth-functions ($k = 2$) there are exactly two *pr*-clone. In the general case, our main result is to present and prove the basic theorem of *pr*-completeness (similar to the theorem of Rosenberg).

Alkalmazott Matematikai Lapok 17 (1993) 329–345

MEGJEGYZÉSEK LÁMER G.: A SZÜKSÉGES ÉS ELÉGSÉGES ÖSSZEFÉRHETŐSÉGI PEREMFELTÉTELEK MEGHATÁROZÁSA CÍMŰ CIKKÉHEZ

KOZÁK IMRE

Miskolc

A tanulmány összefoglaló jelleggel ismerteti a szilárd kontinuumok mechanikája összeférhetőségi feladatának megoldását mind térfogati, mind felületi tartományon, az alakváltozás nemlineáris és lineáris elmélete esetében is. Áttekinti a lineáris rugalmasságtan duál egyenlet-rendszerének és az összeférhetőségi feltételeknek a kapcsolatát, különös tekintettel a peremfeltételekre. A tanulmány ezek után LÁMER, G. cikkének egyes megállapításaira reagál, korántsem kitérve a cikk egészére.

1. Bevezetés

1.1. LÁMER [1] elemzés tárgyává teszi a WASHIZU [2] és KOZÁK [3] tanulmányokat, amelyek mindegyikének a kontinuummechanika összeférhetőségi (kompatibilitási) feltételei a tárgya, és kritikai észrevételeket fűz KOZÁK [3]-hoz.

Az összeférhetőségi feltételek a szilárd kontinuumok mechanikájának sajátos problémája. Beszélhetünk összeférhetőségi feltételekről térfogati vagy felületi tartományon, az alakváltozás nemlineáris vagy lineáris elmélete alapján, továbbá szerepükről a rugalamasságtan duál egyenletrendszerében. A problémakör egy része (összeférhetőségi feltételek térfogati tartományon) monográfiák standard szakaszait képezi, más része azonban (pl. összeférhetőségi feltételek felületi tartományon) a közelmúltban is érdeklődés tárgya volt, sőt — amint azt LÁMER [1] mutatja — még napjainkban is az.

Jelen tanulmány szerzőjét akkor kezdték el érdekelni az összeférhetőségi feltételek részletei, amikor olyan általános, háromdimnziós lineáris héjelmélet felépítésén dolgozott, amelynek a feszültség koordináták az alapváltozói (nem pedig az elmozdulás koordináták, amint az a héjelméletekben addig kizárólagos volt). Szerző ezután a témakörrel [3]-on kívül több más munkájában is foglalkozott, vagy legalábbis érintette azt, KOZÁK [4]–[10] és BÉDA–KOZÁK–VERHÁS [11].

A következő, 2. szakasz az összeférhetőségi feltételek feladatot ismerteti. A 3. szakasz összefoglalja a feladatra vonatkozó megoldásokat, először térfogati majd felületi tartományon. Tárgyalja az összeférhetőségi feltételeket az alakváltozás nemlineáris és lineáris elmélete szerint is. A megkülönböztetés azért indokolt, mert az alakváltozás nemlineáris elméletéből következnek ugyan a lineáris elmélet eredményei, mégis a lineáris elmélet általában ettől független más eljárásokat használ.

Felmerül a kérdés, miért kell az összeférhetőségi feltételeket felületi tartományon is vizsgálni. Nem csak azért, mert azok szorosan kapcsolódnak a rugalmasságtan duál rendszerében a mezőegyenletek és peremfeltételek számához, hanem azért is, mert önálló szerepük lehet a középfelület vonatkozásában a kétdimenziós héjelméletek felépítésében.

A 4. szakasz az összeférhetőségi feltételeknek a rugalmasságtan duál egyenletrendszerében betöltött szerepét részletezi a mezőegyenletek és peremfeltételek tekintetében, ez esetben csak az alakváltozás lineáris elmélete alapján.

Az 5. szakaszban a szerző LÁMER [1] említett kritikai észrevételeire tesz megjegyzéseket, mintegy válaszul azokra.

A 2.–4. szakaszokat a szerző azért tartja szükségesnek, hogy az 5. szakasz megjegyzései jól áttekinthetőek és egyértelműek legyenek.

Feltételezzük, hogy a vizsgált kontinuum V térfogati tartománya egyszeresen összefüggő és S pereme egyetlen zárt, sima felület. Alkalmazzuk az indexes jelölést.

Legyenek az *euklideszi térben* értelmezett, tetszőleges görbevonalú koordinátarendszer koordinátái x^k, bázisvektorai g_k, metrikus tenzora $g_{k\ell}$, első- és másodfajú *Christoffel-szimbólumai* $\Gamma_{k\ell,m}$ és $\Gamma^m_{k\ell}$, permutációs tenzora $\varepsilon_{k\ell m}$, *Riemann-Christoffel görbületi tenzora* $R_{k\ell pq}$. Jelölje a tenzor parciális deriváltját vessző utáni index az alsó indexsorban (kivétel a szabály alól a $\Gamma_{k\ell,m}$ szimbólum), kovariáns deriváltját pedig ugyanott a pontosvessző utáni index.

Amikor kontinuum alakváltozásának elméletéről van szó, a *teljes Lagrange-féle leírási módot* alkalmazzuk és az x^k koordináták a *Lagrange-féle (más néven materiális, vagy együttmozgó) koordinátákat* jelentik. A *teljes Lagrange-féle leírási módban* a mennyiségeket a kontinuum kezdeti (alakváltozás előtti) konfigurációjának pontjaihoz kötjük. A koordináta-rendszernek a kontinuum alakváltozása következtében megváltozott jellemzőit felülvonással különböztetjük meg. Pl. $\bar{g}_{k\ell}$ és $g_{k\ell}$ az alakváltozás utáni és az alakváltozás előtti metrikus tenzor.

Legyen $e_{k\ell}$, $e_{\ell k} = e_{k\ell}$ a szimmetrikus *Green–Cauchy alakváltozási tenzor*, $a_{k\ell}$, $a_{\ell k} = a_{k\ell}$ a szintén szimmetrikus *Green–Lagrange alakváltozási tenzor* és u_k az elmozdulásvektor.

A továbbiakban szövegközi környezetben az alábbi rövidítésekkel élünk:
ANLE $\equiv$ az alakváltozás nemlineáris elmélete, ALE $\equiv$ az alakváltozás lineáris elmélete.

A hivatkozás nélkül közölt összefüggések monográfiákban megtalálhatók, pl. ERINGEN [12]-ben vagy BÉDA–KOZÁK–VERHÁS [11]-ben.

2. Szilárd kontinuumok összeférhetőségi feladata

Az u_k az elmozdulás vektormezőből (továbbiakban: elmozdulásmező) az alábbiak szerint számítható az $a_{k\ell}$ alakváltozási tenzormező (továbbiakban: alakváltozásmező):

(2.1) ANLE: $$a_{k\ell} = \tfrac{1}{2}\left(u_{k;\ell} + u_{\ell;k} + g^{mn}u_{m;k}u_{n;\ell}\right),$$

(2.2) ALE: $$a_{k\ell} = \tfrac{1}{2}\left(u_{k;\ell} + u_{\ell;k}\right) .$$

A fordított esetben, amikor az $a_{k\ell}$ alakváltozásmező koordinátái adottak és az u_k elmozdulásmező a keresett, a 3 elmozdulás koordináta meghatározása a (2.1), vagy a (2.2) alatti 6 parciális differenciálegyenlet megoldását igényli. A feladat túlhatározott, és ahhoz, hogy — a kontinuum adottnak feltételezett merevtestszerű mozgása estén — folytonos és egyértékű elmozdulásmező létezzen, az $a_{k\ell}$ tenzormező bizonyos korlátozásoknak (feltételeknek) kell eleget tegyen. Ezek a feltételek az összeférhetőségi (kompatibilitási) feltételek.

3. Szilárd kontinuumok összeférhetőségi feltételei

3.1. Térfogati tartományon

3.1.1. Az összeférhetőségi feltételekhez a (2.1–2) egyenletekből ki kell küszöbölni az u_k elmozdulás koordinátákat. ALE-ben ez relatíve egyszerű. Ekkor (2.2) alapján

(3.1) ALE: $$\eta^{ab} = \varepsilon^{ak\ell}\varepsilon^{bpq}a_{kp;\ell q} = 0$$

adódik, mint összeférhetőségi feltétel, ahol η^{ab}; $\eta^{ba} = \eta^{ab}$ a szimmetrikus inkompatibilitási (összeférhetőségi) tenzor ALE-ban. (3.1) a *Saint-Venant összeférhetőségi egyenletek (feltételek).*

ALE-ben az összeférhetőségi feltétel meghatározására egy közvetlen út is létezik. Az $a_{k\ell}$ alakváltozásmező és az egész kontinuum merevtestszerű mozgásának ismeretében ugyanis a *Casaro-formulával* görbe menti integrálással bármely pontban meghatározható a forgásvektor és az elmozdulásvektor. Ahhoz, hogy az eredmény az integrálási görbétől független legyen a (3.1) feltételnek kell teljesülnie.

A fentiek miatt is, szokás az összeférhetőségi feltételeket integrálhatósági feltételeknek nevezni.

3.1.2. ANLE-ben az u_k elmozdulás koordináták kiküszöbölése (2.1)-ből rendkívül fáradtságos és szinte reménytelenül komplikált és a *Casaro-formula* nem működik. Helyettük egy alternatív módszer, a *Riemann-féle elmélet* alkalmazása terjedt el. Az alábbiak először — a kontinuum alakváltozásától függetlenül — ismertetik a *Riemann-féle elmélet* vonatkozó részeit.

A Riemann-féle elmélet szerint
ahhoz, hogy egy szimmetrikus $g_{k\ell}$ tenzor euklideszi tér metrikus tenzora legyen, szükséges és elégséges, hogy $g_{k\ell}$ szimmetrikus pozitív definit tenzor legyen, és identikusan kielégítse az $R_{k\ell pq}$ Riemann-Christoffel görbületi tenzort:

(3.2) $$R_{k\ell pq} = 0 .$$

Értelmezés szerint

$$(3.3)\quad R_{k\ell pq}=\tfrac{1}{2}\left(g_{kq,\ell p}+g_{\ell p,kq}-g_{kp,\ell q}-g_{\ell q,kp}\right)+g^{mn}\left(\Gamma_{kq,m}\Gamma_{\ell p,n}-\Gamma_{kp,m}\Gamma_{\ell q,n}\right),$$

$$(3.4)\qquad R_{k\ell pq}=-R_{\ell kpq}\qquad R_{k\ell pq}=-R_{k\ell qp}\qquad R_{k\ell pq}=R_{pqk\ell},$$

$$(3.5)\qquad \Gamma_{k\ell,m}=\frac{1}{2}\left(g_{km,\ell}+g_{\ell m,k}-g_{k\ell,m}\right),$$

és *euklideszi térben* fennáll a

$$(3.6)\qquad P^{ab}=\varepsilon^{ak\ell}\varepsilon^{bpq}R_{k\ell pq}=0\qquad P^{ab}=P^{ba}$$

egyenlet is. Legutóbbi egyenlet (3.3–4) alapján az alábbi alakra hozható:

$$(3.7)\qquad P^{ab}=-2\varepsilon^{ak\ell}\varepsilon^{bpq}\left(g_{kp,\ell q}+\Gamma_{kp,m}\Gamma_{\ell q,n}g^{mn}\right)=0.$$

A (3.2) és (3.6) egyenletek (3.4) miatt csak 6 egymástól különböző skaláris egyenletet jelentenek:

$$(3.8)\qquad R_{1212}=R_{2323}=R_{3131}=R_{1223}=R_{2331}=R_{3112}=0,$$

$$(3.9)\qquad P^{11}=P^{22}=P^{33}=P^{12}=P^{23}=P^{31}=0.$$

Az $R_{k\ell pq}$ *Riemann-Christoffel görbületi tenzorra* fennáll a *Bianchi-azonosság*:

$$(3.10)\qquad R_{k\ell pq;m}+R_{k\ell qm;p}+R_{k\ell mp;q}=0,$$

illetve más alakban:

$$(3.11)\qquad P^{ab}{}_{;b}=\varepsilon^{ak\ell}\varepsilon^{bpq}R_{k\ell pq;b}=0.$$

3.1.3 Abból a követelményből, hogy a kontinuum metrikája az alakváltozás után is *Euklideszi teret* írjon le, (3.2) és (3.6–7) analógiája alapján a

$$(3.12)\ \text{ANLE:}\qquad \bar{R}_{k\ell pq}=0,$$

vagy a

$$(3.13)\ \text{ANLE:}\quad \bar{P}^{ab}=\bar{\varepsilon}^{ak\ell}\bar{\varepsilon}^{bpq}\bar{R}_{k\ell pq}=-2\bar{\varepsilon}^{ak\ell}\bar{\varepsilon}^{bpq}\left(\bar{g}_{kp,\ell q}+\bar{\Gamma}_{kp,m}\bar{\Gamma}_{\ell q,n}\bar{g}^{mn}\right)=0$$

egyenletnek kell teljesülnie (jelölésbeli megállapodásunk szerint $\bar{R}_{k\ell pq}$ és $\bar{P}^{ab}$ az alakváltozás utáni állapotra vonatkozik), ahol $\bar{g}_{kp}$ a megváltozott metrikus tenzor, és azonos a nemszinguláris, pozitív definit e_{kp} *Green-Cauchy alakváltozási tenzorral.*

Fennállnak az alábbiak :

$$\bar{g}_{kp} = e_{kp} = g_{kp} + 2a_{kp} \,. \tag{3.14}$$

ahol a_{kp} a *Green-Lagrange alakváltozási tenzor*, továbbá (3.5)-öt is figyelembe véve

$$\bar{\Gamma}_{kp,m} = \Gamma_{kp,m} + (a_{km,p} + a_{pm,k} - a_{kp,m}) \,. \tag{3.15}$$

Hozzávéve még (3.14–15)-höz a $\bar{g}^{mn} = \left(e^{-1}\right)^{mn}$ megváltozott felsőindexes metrikus tenzort, megállapíthatjuk, hogy (3.12), vagy (3.13) tulajdonképpen az e_{kp}, illetve a_{kp} tenzor koordinátáira jelent differenciálegyenlet-rendszert.

(3.12) vagy (3.13) ANLE-ben a *teljes Lagrange-féle leírási mód* szerinti összeférhetőségi (kompatibilitási) feltétel.

Az összeférhetőségi feltétel a

$$\varrho^{ab} = -\tfrac{1}{4}\left(\bar{P}^{ab} - P^{ab}\right) = 0 \tag{3.16}$$

alakban is megfogalmazható, ahol ϱ^{ab}, $\varrho^{ba} = \varrho^{ab}$ a szimmetrikus inkompatibilitási (összeférhetetlenségi) tenzor ANLE-ben.

3.1.4. Meg kell jegyeznünk, hogy sem a (3.1) sem a (3.12–13) összférhetőségi feltételek nem függetlenek egymástól. (3.1) alapján azonnal belátható a

$$\text{(3.17)}\quad \text{ALE:} \qquad \eta^{ab}{}_{;b} = \varepsilon^{ak\ell}\varepsilon^{bpq}a_{kp;\ell qb} = 0$$

egyenlet teljesülése, míg a megváltozott $\bar{R}_{k\ell pq}$ *Riemann-Christoffel görbületi tenzorra* és $\bar{P}^{ab}$ tenzorra fennáll a *Bianchi-azonosság:*

$$\text{(3.18)}\quad \text{ANLE:} \qquad \bar{P}^{ab}{}_{;b} = \varepsilon^{ak\ell}\varepsilon^{bpq}\bar{R}_{k\ell pq;b} = 0 \,.$$

Szokás (3.17)-et is *Bianchi-azonosságnak* nevezni.

A (3.18) *Bianchi-azonosság* — tekintettel (3.16)-ra — a

$$\text{(3.19)}\quad \text{ANLE:} \qquad \varrho^{ab}{}_{;b} = 0$$

alakban is felírható.

3.1.4. ALE-ről akkor beszélünk, ha az elmozdulásmező $u_{k;\ell}$ gradiensének koordinátái (globális merevtestszerű mozgást nem számítva), következőleg az $a_{k\ell}$ alakváltozási tenzor koordinátái is, abszolút értékeiket tekintve nagyságrendekkel (pl. 10^{-3} szorosan) kisebbek, mint a $g_{k\ell}$ metrikus tenzor koordinátái. Ilyenkor (lásd pl. ERINGEN [12], kissé más formalizmussal) lineáris közelítéssel:

$$\text{(3.20)}\quad \text{ANLE} \implies \text{ALE:} \qquad \varrho^{ab} = \eta^{ab} = \varepsilon^{ak\ell}\varepsilon^{bpq}a_{kp;\ell q} = 0 \,,$$

és a *Bianchi-azonosságra* nézve is fennáll:

$$\text{(3.21)}\quad \text{ANLE} \implies \text{ALE:} \qquad \varrho^{ab}{}_{;b} = \eta^{ab}{}_{;b} = 0 \,.$$

3.2. Felületi tartományon

3.2.1. A zárt g görbére kifeszített sima S_g felületen is használható a *Cesaro-formula* ALE-ben a forgásvektor és az elmozdulásvektor meghatározására. Ekkor az eredménynek az integrálási görbétől való függetlensége az alábbi feltételhez vezet:

(3.22) ALE: $\qquad \eta^{k\ell} n_\ell = 0 \quad$ az S_g-n.

(n_ℓ a felület normálisa). Szerző szerint (3.22) az összeférhetőségi peremfeltétel ALE-ban. Igazolásához a *Stokes-tételt* is fel kell használni (lásd pl. BÉDA–KOZÁK–VERHÁS [11]).

A V tartomány S peremén az

(3.23) ALE: $\qquad \eta^{k\ell} n_\ell = 0 \quad$ az S-en.

feltételen túl még annak is teljesülnie kell, hogy a perem valamely zárt g görbéjén

(3.24) az alakváltozásmező és a merevtestszerű forgásmező görbementi deriváltja a g-n a görbével két részre vágott perem egyik és másik része felöl nézve is azonos legyen.

Szükséges hangsúlyozni, hogy a (3.23) és (3.24) feltételekhez nem elegendő S-en csupán az alakváltozásmező ismerete, mivel a szereplő formulákban az alakváltozási koordináták felületre merőleges deriváltjai is fellépnek.

3.2.2. ANLE-ben más utat kell járnunk. Ehhez — a kontinuum alakváltozásától függetlenül — a felületelmélet ide vonatkozó főbb eredményeit kell előbb összefoglalnunk.

Legyen most a koordináta-rendszer olyan, hogy a zárt g görbére kifeszített sima S_g felületen $x^3 = 0$, és x^3-at, mint előjeles távolságot a felület normálisa mentén mérjük (felületre épített koordináta-rendszer). Vezessük be, hogy a görög betűs indexek csak az 1,2 értékeket veszik fel, és jelöljük „$\circ$"-rel a felületen kívül is értelmezett mennyiség felületen felvett értékét, pl. $g_{\alpha\beta} = \overset{\circ}{g}_{\alpha\beta}\left(x^1, x^2, x^3 = 0\right)$. A felület $\overset{\circ}{n}_k$ normálisának koordinátái: $\overset{\circ}{n}_\alpha = 0$, $\overset{\circ}{n}_3 = \overset{\circ}{n}{}^3 = 1$.

A felületen, pl. BÉDA–KOZÁK [13] szerint:

(3.25) $$\overset{\circ}{\Gamma}{}^m_{33} = 0 \qquad \overset{\circ}{\Gamma}{}^3_{k3} = 0 \qquad \overset{\circ}{\Gamma}{}^3_{33} = 0$$

és

(3.26) $$\overset{\circ}{\Gamma}{}^3_{\kappa\lambda} = b_{\kappa\lambda} \qquad \overset{\circ}{\Gamma}{}^\mu_{3\lambda} = -b^\mu_\lambda\,,$$

ahol $b_{\kappa\lambda}$, $b^{\mu\kappa} = \overset{\circ}{g}{}^{\mu\kappa} b_{\kappa\lambda}$ a felület görbületi tenzora.

Szokás $\overset{\circ}{g}_{\alpha\beta}$ koordinátáit első alapmennyiségeknek, $b_{\kappa\lambda}$ koordinátáit pedig második alapmennyiségeknek nevezni.

Fennállnak az alábbi összefüggések:

$$g_\kappa = \underset{\circ}{g}_\kappa - x^3 b_\kappa^\mu \underset{\circ}{g}_\mu \tag{3.27}$$

$$g_{\kappa\lambda} = \underset{\circ}{g}_{\kappa\lambda} - 2x^3 b_{\kappa\lambda} + (x^3)^2 b_\kappa^\mu b_{\mu\lambda} \quad g_{\kappa 3} = \underset{\circ}{g}_{\kappa 3} = 0 \quad g_{33} = \underset{\circ}{g}_{33} = 1\,, \tag{3.28}$$

(3.28) szerint

$$(g_{\kappa\lambda,3})_0 = -2b_{\kappa\lambda}\,. \tag{3.29}$$

A *Riemann-Chrisltffel* görbületi tenzor koordinátáinak felületen felvett értékei a (3.3), (3.5) és a (3.25–29) képletekből adódnak:

$$\underset{\circ}{R}_{1212} = \frac{1}{2}\left(2\underset{\circ}{g}_{12,12} - \underset{\circ}{g}_{11,22} - \underset{\circ}{g}_{22,11}\right) + \tag{3.30}$$

$$+ \underset{\circ}{g}^{\mu\nu}\left(\underset{\circ}{\Gamma}_{12,\mu}\underset{\circ}{\Gamma}_{12,\nu} - \underset{\circ}{\Gamma}_{11,\mu}\underset{\circ}{\Gamma}_{22,\nu}\right) + (b_{12})^2 - b_{11}b_{22} = 0$$

$$\underset{\circ}{R}_{12\beta 3} = b_{1\beta,2} - b_{2\beta,1} + \underset{\circ}{\Gamma}^\mu_{1\beta} b_{\mu 2} - \underset{\circ}{\Gamma}^\mu_{2\beta} b_{\mu 1} = 0\,, \tag{3.31}$$

$$\underset{\circ}{R}_{3\alpha 3\beta} = -b^\mu_\alpha b_{\mu\beta} + b^\mu_\beta b_{\mu\alpha} \equiv 0\,. \tag{3.32}$$

Összegezve megállapíthatjuk, hogy felületen a *Riemann-Christoffel görbületi tenzor* egymástól különböző 6 koordinátája közül

- (3.32) szerint $\underset{\circ}{R}_{2323}, \underset{\circ}{R}_{3131}$ és $\underset{\circ}{R}_{2331}$ identikusan zérus,
- (3.30) szerint $\underset{\circ}{R}_{1212} = 0$ a *Gauss-féle Theorema egregium* állítását jelenti és
- (3.31) szerinti $\underset{\circ}{R}_{1223} = 0$ és $\underset{\circ}{R}_{3112} = 0$ egyenletek a *Mainardi-Codazzi-féle formulát* adják meg.

$\underset{\circ}{R}_{k\ell pq}$ nem identikusan zérus koordinátái (3.6) és az előzőek szerint összefoghatók a

$$\underset{\circ}{P}^{3b} = \underset{\circ}{n}_a \underset{\circ}{P}^{ab} = \underset{\circ}{\varepsilon}^{3\kappa\lambda} \underset{\circ}{\varepsilon}^{bpq} \underset{\circ}{R}_{\kappa\lambda pq} = 0 \tag{3.33}$$

egyenletbe, amelyben valóban csak az $\underset{\circ}{R}_{1212}$ és $\underset{\circ}{R}_{12\beta 3}$ koordináták szerepelnek.

Meg kell jegyeznünk, hogy a

$$\underset{\circ}{P}^{\alpha\beta} = \underset{\circ}{\varepsilon}^{\alpha k\ell} \underset{\circ}{\varepsilon}^{\beta pq} \underset{\circ}{R}_{k\ell pq} \tag{3.34}$$

kifejezés csak az identikusan zérus $\underset{\circ}{R}_{3\alpha 3\beta}$ egymástól különböző koordinátákat tartalmazza, vagyis maga is identikusan zérus.

Bonnet-tétele, vagy más néven a *felületelmélel főtélele* (SZŐKEFALVI-NAGY, GEHÉR és NAGY [14]) szerint (a jelen tanulmány jelöléseivel):

ha a $g_{\kappa\lambda}$ *és* $b_{\kappa\lambda}$ *függvények eleget tesznek a Theorema egregium állitásának és a Mainardi-Codazzi egyenleteknek, továbbá* $\det \overset{\circ}{g} > 0$, *akkor térbeli mozgásoktól és tükrözésektől eltekintve egy és csak egy olyan felület létezik, amelynek ezek az első és második alapmennyiségei.*

A (3.33)-ban összefoglalt eredmények tehát, vagyis a *Riemann-Christoffel görbületi tenzor* meghatározott 3 koordinátájának zérus volta a felületen, figyelembe véve azt is, hogy $\det \overset{\circ}{g}_{\kappa\lambda} > 0$ mindig fennáll, biztosítják a Bonnet-tétel leljesülését.

Ugyancsak [14] szerint: a *Bonnet-tétel* bizonyítása arra épül, hogy a differenciálegyenletként felfogott *Gauss- és Weingarten-egyenletek* integrálhatók a tételbeli feltételek mellett. E feltételeket ezért integrálhatósági feltételeknek is nevezik. (A *Gauss-egyenletek* a felület helyvektorának második, a *Weingarten-egyenletek* a felület normálisának első deriváltjai az x^α koordináták, mint a felület paraméterei szerint.)

Tekintettel arra, hogy (3.33)-ban csupa tenzormennyiség szerepel, a képlet nem csak felületre épített koordináta-rendszerben, hanem bármely koordináta-rendszerben felírható, fennáll az

$$n_a P^{ab} = n_a \varepsilon^{ak\ell} \varepsilon^{bpq} R_{k\ell pq} = 0 \quad \text{az } S_g\text{-n} \tag{3.35}$$

egyenlet (n_a az S_g felület normálisa).

A (3.35) alatti 3 skaláris egyenlet tetszőleges koordináta-rendszerben is biztosítja, hogy a *Bonnet-tétel* az S_g felületen teljesüljön (a $\det g_{k\ell} > 0$ mindig fennáll).

A (3.33) és (3.35) egyenletek akkor is fennállnak, ha azokat egyszeresen összefüggő, sima, zárt S felületre alkalmazzuk.

3.2.3. Térjünk vissza a kontinuum alakváltozásához. A 3.1.3. pontban elmondottak szerint (3.35) a felület megváltozott geometriájára is igaz:

$$\text{ANLE:} \qquad \bar{n}_a \bar{P}^{ab} = \bar{n}_a \bar{\varepsilon}^{ak\ell} \bar{\varepsilon}^{bpq} \bar{R}_{k\ell pq} = 0 \quad \text{az } S_g\text{-n}, \tag{3.36}$$

ahol $\bar{n}_a$ a megváltozott felület normálisa (a $\det \bar{g}_{k\ell} > 0$ mindig fennáll).

Megkapjuk ezáltal a *teljes Lagrange-féle leírási mód* szerinti összeférhetőségi peremfeltételt ANLE-ban:

$$\text{ANLE:} \qquad \bar{n}_a \varrho^{ab} = \bar{n}_a \left(\bar{P}^{ab} - P^{ab} \right) = 0 \quad \text{az } S_g\text{-n}. \tag{3.37}$$

A (3.36), vagy (3.37) alatti 3 skaláris egyenlet biztosítja, hogy a *Bonnet-tétel* az alakváltozás utáni felületen is teljesüljön, más szóval, hogy — a felület egészének merevtestszerű mozgásától eltekintve — a felületen egyértékű elmozdulásmező legyen előállítható.

(3.36) és (3.37) is fennáll, ha azokat egyszeresen összefüggő, sima, zárt S felületre alkalmazzuk.

A *Bonnet-tétel* akkor is teljesül, ha (3.35)-ben P^{ab}, vagy (3.37)-ben ϱ^{ab} egymástól különböző 6 koordinátája közül előírjuk 1-, 2-, ..., 6-nak a zérus voltát. Az ilyen esetekben a képletekből további következtetések is levonhatók, mindezek azonban nem érintik magát a tételt. Pl., ha előírjuk, hogy a felületen $\varrho^{12} = \varrho^{23} = \varrho^{31} = 0$ legyen, és emellett $\bar{n}_1 \neq \bar{n}_2 \neq \bar{n}_3 \neq \bar{n}_1$, akkor (3.37)-ből ugyanott $\varrho^{11} = \varrho^{22} = \varrho^{33} = 0$ következik, ezzel szemben, ha $\varrho^{12} = \varrho^{23} = \varrho^{31} = 0$ mellett $\bar{n}_1 = \bar{n}_2 = 0$ és $\bar{n}_3 = 1$, akkor (3.37)-ből a felületen csak a $\varrho^{33} = 0$ egyenlet adódik.

3.2.4. Az ANLE szerinti (3.37) összeférhetőségi peremfeltételből a 3.1.4. pont szerinti lineáris közelítéssel ALE-ban is megkapjuk (3.20) alapján az összeférhetőségi peremfeltételt. Ekkor ugyanis $\bar{n}_a \Rightarrow n_a$, $\varrho^{ab} \Rightarrow \eta^{ab}$, tehát fennáll:

$$(3.38)\quad \text{ANLE} \Rightarrow \text{ALE:}\quad \bar{n}_a \varrho^{ab} = n_a \varrho^{ab} = n_a \eta^{ab} = n_a \varepsilon^{akl} \varepsilon^{bpq} a_{kp;lq} = 0 \quad \text{az } S_g\text{-n.}$$

Az egymástól különböző módokon kapott (3.22) és (3.38) összeférhetőségi peremfeltételek — amint azt (3.38) is mutatja — megegyeznek.

(3.38) is fennáll egyszeresen összefüggő, zárt, sima S felületen.

4. Összeférhetőségi feltételek a lineáris rugalmasságtan duál rendszerében

4.1. Előzetesen szólni kell a lineáris rugalmasságtan primál és duál egyenletrendszeréről, amelyek — mint egymástól független rendszerek — elvileg egyaránt alkalmasak a lineáris rugalmasságtan peremérték feladatainak a megoldására.

A dolgok természetéből következően a lineáris rugalmasságtanban csak ALE szerepelhet.

Primál rendszerben az elmozdulásmező 3 koordinátája az alapváltozó, amelyekre nézve a mérlegegyenletek (egyensúlyi egyenletek) 3 másodrendű differenciálegyenletet szolgáltatnak. A peremfeltételek száma (elmozdulási, vagy feszültségi): 3.

Duál rendszerben 3 feszültségfüggvény, mint a szimmetrikus feszültségfüggvény tenzor 3 alkalmasan kiválasztott koordinátája az alapváltozó, amelyekre nézve a mérlegegyenletek (alkalmasan kiválasztott 3 összeférhetőségi mezőegyenlet) 3 negyedrendű differenciálegyenletet adnak. A peremfeltételek száma 6, amelyekből 3 lehet feszültségi peremfeltétel.

A duál rendszerben nem szerepel az elmozdulásmező, szükséges tehát, hogy a duál egyenletrendszer biztosítsa az alakváltozásmező összeférhetőségét (integrálhatóságát). Ehhez, mint mérlegegyenlet, a (3.1) alatti 6 *Saint-Venant összeférhetőségi egyenletből* (mint összeférhetőségi mezőegyenletből), csak 3 jöhet szóba (3-nál több mérlegegyenlet nem lehetséges), a többi összeférhetőségi feltétel tehát a peremen keresendő (úgyis hiányzik a szükséges 6 peremfeltételből legalább 3).

A duál rendszer vázolt problémája úgy is felmerült, hogy SOUTHWELL [15] feszültségfüggvények alkalmazásával csak 3 összeférhetőségi mezőegyenletet kapott a rugalmasságtan *Castigliano-féle (a teljes kiegészítő energiára vonatkozó) variációs elvéből.*

Szükséges hangsúlyozni, hogy 3 összeférhetőségi mezőegyenlet nem biztosítja az alakváltozásmező integrálhatóságának (összeférhetőségének) (3.1) alatti feltételét.

A továbbiakban (3.20) és (3.38) alapján — és LÁMER [1] jelölésrendszeréhez igazodva — az inkompatibilitási tenzor jelölésére akkor is a ϱ^{ab} jelölést használjuk, ha egyébként, mint a jelen szakaszban is, ALE-ről van szó.

4.2. A *Southwell paradoxonnak* is elnevezett problémára WASHIZU [2] adott megoldást, amennyiben — *Descartes-féle koordináta-rendszert* feltételezve és kihasználva a (3.21) *Bianchi azosságot* — kimutatta, hogy valóban csak 3 összeférhetőségi mezőegyenletre van szükség, de ezen túl elő kell írni összeférhetőségi feltételeket a peremen. Konkrétan, WASHIZU kimutatta, hogy ha

$$\varrho^{11} = \varrho^{22} = \varrho^{33} = 0 \quad \text{a } V\text{-n} \tag{4.1}$$

és

$$\varrho^{12} = \varrho^{23} = \varrho^{31} = 0 \quad \text{az } S\text{-en}, \tag{4.2}$$

vagy

$$\varrho^{12} = \varrho^{23} = \varrho^{31} = 0 \quad \text{a } V\text{-n} \tag{4.3}$$

és

$$\varrho^{11} = \varrho^{22} = \varrho^{33} = 0 \quad \text{az } S\text{-en}, \tag{4.4}$$

akkor a (4.1)-ből, illetve (4.3)-ból hiányzó másik 3, vagyis a (3.1) alatti összes összeférhetőségi mezőegyenlet teljesül.

Arra is rámutatott WASHIZU, hogy a

$$\varrho^{k\ell} = 0 \quad \varrho^{12} = \varrho^{22} = \varrho^{33} = \varrho^{12} = \varrho^{23} = \varrho^{31} = 0 \quad \text{az } S\text{-en}. \tag{4.5}$$

feltételek esetén akár (4.1), akár (4.3) szolgálhat összeférhetőségi mezőegyenletként. Egyúttal felhívta a figyelmet, hogy a (4.5) feltételek egy része feleslegesnek látszik. A probléma megoldására WASHIZU a variációs eljárással nyerhető peremfeltételek gondos elemzését javasolta.

Összegezve: WASHIZU feloldotta a *Southwell-paradoxont*, mivel kimutatta, hogy valóban csak 3 összeférhetőségi mezőegyenletre van szükség, az általános — az összeférhetőségi mezőegyenletektől független — összeférhetőségi peremfeltételek kérdését azonban nyitva hagyta.

4.3. A duál rendszer összeférhetőségi egyenleteinek a problémájára szerző háromféle módon is kereste a megoldást, mindhárom esetben tetszőleges görbevonalú koordináta-rendszert feltételezve:

- a *Bianchi-azonosság* alapján KOZÁK [3], [5], [7] és BÉDA–KOZÁK–VERHÁS [11],

- a virtuális munka elv segítségével KOZÁK [7] és [8]
- és a *Castigliano-féle variációs elv* alkalmazásával KOZÁK [4], [6], [7] és [9], a két utóbbi esetben a peremfeltételek WASHIZU által javasolt gondos elemzésével.

Szerző mindhárom módon azonos eredményre jutott. Nevezetesen, ha

$$\varrho^{XY} = 0 \quad \text{a } V\text{-n} \tag{4.6}$$

és

$$\varrho^{k\ell} n_\ell = 0 \quad \text{az } S\text{-en} \tag{4.7}$$

teljesül, akkor

$$\varrho^{AB} = 0 \quad \text{a } V\text{-n} \tag{4.8}$$

is teljesül. Az XY és AB három–három indexpár kiadja a teljes $k\ell$ $(k, \ell = 1, 2, 3)$ kombinációt, vagyis (4.6) és (4.8) együtt (3.1) összes egyenletét jelenti. Az XY indexpárok meghatározott szabály szerint választhatók, illetve választandók ki, lásd pl. KOZÁK [3] és BÉDA-KOZÁK-VERHÁS [11].

A (4.6–7) összeférhetőségi feltételeket már GRYCZ [16] is megadta.

A *Bianchi-azonosság* és a virtuális munka elv független a kontinuum anyagegyenletétől, (4.6) és (4.7) tehát nem csak a lineáris rugalmasságtanban érvényes.

A korábbi 3.1. és 3.2. szakaszokban és a jelen 4.3. szakaszban elmondottak között az alábbiak teremtenek szoros kapcsolatot. A 4.3. szakasz (4.6–7) összeférhetőségi feltételei, amelyeket szerző a felsorolt, egymástól különböző három módon nyert, együtt a (4.8) képletekkel, azonosak a 3.1. és 3.2. szakaszok más gondolatmenet alapján kapott (3.1) és (3.20) összeférhetőségi mezőegyenleteivel, illetve (3.22) és (3.38) összeférhetőségi peremfeltételeivel.

4.4. Legyenek olyanok a peremfeltételek, hogy a kontinuum peremének egyszeresen összefüggő S_t, részén feszültségi, szintén egyszeresen összefüggő S_u részén pedig elmozdulási peremfeltételt írunk elő és $S_t \cup S_u = S$. Ez esetben a virtuális munka elvből és a *Castigliano-féle variációs elvből* (4.7)-en túl további összeférhetőségi feltételek is következnek a peremen.

Nevezetesen

– az S_u peremrészen összesen 6, ún. alakváltozási peremfeltétel adódik, amelyek egyrészt biztosítják ugyanott a (4.7) összeférhetőségi peremfeltétel teljesülését, másrészt azt, hogy az alakváltozásmezőből és az alakváltozási koordináták felületre merőleges deriváltjaiból integrálással valóban az előírt elmozdulásmezőt kapjuk meg (merevtestszerű mozgást leszámítva), lásd KOZÁK [7],

– megkapjuk továbbá az S_u és S_t peremrész közös g görbéjén az (1.5) összeférhetőségi feltételt.

Ilyen módon a lineáris rugalmasságtan duál rendszerében mind az S_t mind az S_u peremrészen 6 skaláris peremfeltétel áll rendelkezésünkre. S_t-n a 3 feszültségi és

a (4.7) szerinti 3 összeférhetőségi peremfeltétel, míg S_u-n az említett 6 alakváltozási peremfeltétel.

ABOVSKIJ–ANDREEV–DERUGA [17] is megad az S_u peremrészre alakváltozási peremfeltételeket, ezek azonban nem egyeznek meg a KOZÁK [7]-beliekkel.

4.5. Az összeférhetőségi feltételeknek — a fentiekben vázolt eléggé körültekintő — vizsgálata után szerzőnek sikerült felépítenie egy általános, háromdimenziós, lineáris héjelméletet feszültségi koordinátákkal és egyúttal aszimptotikus módszert megadnia a kapott egyenletrendszer integrálására KOZÁK [7], [10].

4.6. LÁMER [1] nem tér ki sem a 4.4. szakaszban elmondottakra sem a 4.3. szakasznak a virtuális munka elvet és a *Castigliano-féle variációs elvet* érintő részeire.

4.7. Megemlítjük még, hogy LÁMER [1] gondolatmenete több helyütt támaszkodik az alábbi lemmára (idézet a 108. oldalról):

„LEMMA. Legyen az f folytonos függvény a V (egyszeresen összefüggő) tartományon azonosan zérus értékű. Ekkor az f függvény a V tartomány S peremén is zérus értéket vesz fel."

5. Megjegyzések

Szerző az alábbiakban nem tér ki a LÁMER [1] tanulmány egészére, mindössze annak a saját korábbi munkáit érintő és ebből a szempontból általa jelentősebbeknek ítélt részeihez fűz megjegyzéseket.

5.1. LÁMER [1] fixa ideája, hogy a kontinuum peremén fenn kell álljon a $\varrho^{k\ell} = 0$ 6 skaláris feltétel, mind ALE, mind ANLE esetén. Eme állítását több helyen is rögzíti, pl. a 99. oldalon („Az inkompatibilitási tenzor komponenseinek eltűnését az S peremen összeférhetőségi peremfeltételnek nevezzük."), a 103. oldalon („A következő paragrafusban bebizonyítjuk, hogy a szükséges és elégséges összeférhetőségi feltétel egy felületen az inkompatibilitási tenzor eltűnése a felületen."), a 105. oldalon („ ... egy reguláris felületen megadott alakváltozási mezőből egyértékű eltolódásmező előállításának szükséges és elégséges feltétele az inkompatibilitási tenzor komponenseinek eltűnése a felületen."), vagy a 109. oldalon („2°. teljesül mind a hat összeférhetőségi peremfeltétel").

A 103. oldalról vett idézetben beígért és a felületelmélet egyes eredményeinek felhasználásával próbálkozó bizonyítás nem állja meg a helyét. Elegendő ehhez ANLE esetén a *Bonnet-tételre* (3.2.1. pont) és a *teljes Lagrange-féle leírási mód* ehhez csatlakozó (3.36) összeférhetőségi peremfeltételére (integrálhatósági feltételére) hivatkozni, míg ALE esetén a 3.2.1. és 3.2.4. pontokra. Mindezekből egyértelműen látszik, hogy felületen az $n_a \varrho^{ab} = 0$ típusú, 3 skaláris összeférhetőségi peremfeltétel a szükséges és elégséges feltétel az elmozdulásmező egyértelmű meghatározásához (merevtestszerű mozgást nem számítva).

Szerző azért iktatta be a 3.1. és 3.2. szakaszoknak ANLE-re vonatkozó részeit, hogy LÁMER [1] fixa ideája állításának helytelenségét a saját korábbi munkáitól füg-

getlenül is bemutassa, de azért is, hogy szemléltesse, ANLE-ből mind térfogati, mind felületi tartományon kiadódnak lineáris közelítéssel ALE összeférhetőségi feltételei. Meg kell ehhez még jegyezni, hogy szerző összes hivatkozott munkája csak ALE-val foglalkozott és ALE-ban az összeférhetőségi feltételek képzéséhez az ANLE-beli módszerektől eltérő eljárások a szokásosak.

5.2. A 105. oldalról vett fenti idézethez meg kell még jegyeznünk, hogy felületen megadott alakváltozásmezőből semmi módon nem lehet elmozdulásmezőt előállítani, mivel ehhez — amint a 3.2.1. pont is felhívja rá a figyelmet — az alakváltozási koordináták felületre merőleges deriváltjai is kellenek.

5.3. LÁMER [1] a 106. oldalon megállapítja (nem teljes egészében szószerint idézve): a *Kozák féle peremfeltélel* nem tartalmazza az alábbi két feltételt:

„1°. a feladatnak a megadott alakváltozási tenzormezőből számított első alapformája pozitív definit, valamint

2°. adott a felületnek a második alapformája."

A megállapítás, mivel szerző korábbi munkássága kizárólag ALE-re vonatkozott, értelmetlen. Egyrészt $\det g_{\alpha\beta} = \det g_{ab} > 0$ mindig fennáll (g_{ab} metrikus tenzor), másrészt (3.14)-ből is beláthatóan ALE esetén $\det \bar{g}_{k\ell} > 0$ is. Ugyanakkor, amint (3.1)-ből és a vele azonos (3.38)-ból is látszik az ún. *Kozák-féle összeférhetőségi peremfeltételhez* (ALE-ről van szó) nem kell a felület második alapformája.

LÁMER fenti megállapítása *teljes Lagrange féle leirási módban* ANLE esetén sem állja meg a helyét. Mivel $\bar{g}_{k\ell} = e_{k\ell}$ és $e_{k\ell}$ nem más, mint a nem szinguláris, pozitív definit *Green-Cauchy alakváltozási tenzor* (lásd pl. ERINGEN [12]), $\det \bar{g}_{k\ell} > 0$ mindig teljesül. Az is rögzíthető a (3.13–15) és a (3.36) képletek alapján, hogy a (3.37) összeférhetőségi peremfeltételhez nem szükséges a megváltozott felület második alapformája (kontinuummechanikában az eredeti felületet, így annak második alapformáját is adottnak vesszük).

5.4. A (3.22) integrálhatósági (összeférhetőségi) feltétel képzéséhez, amint ezt a 3.2.1. pontban említettük, a *Stokes-tételt* másodrendű tenzorra kell alkalmazni. Ehhez LÁMER [1], hogy fixa ideáján sérelem ne essék, két észrevételt is tesz:

1. A 107. oldalon: „De csak formailag kaptuk meg KOZÁK eredményét: ... a *Stokes-tétel* csak vektormezőre értelmezett és nem tenzormezőre (lásd pl. KORN–KORN [2] 164. old.)."

2. A 108. oldalon: „Amennyiben a (klasszikus) *Stokes-tétel* általánosítható tetszőleges tenzorra, úgy annak bizonyítását célszerű lett volna mellékelni."

A KORN–KORN [13] könyv LÁMER [1] által idézett 164. oldalán a *Stokes-tétel* valóban vektormezőre van értelmezve, hiszen az 5. Vektoranalízis fejezet 5.6.2. A Stokes-tétel és kapcsolódó tételek szakaszában található. Tovább lapozva azonban KORN–KORN a 16. Matematikai modellek reprezentációja fejezet a 16.10.11. A másodrendű tenzorok differenciál invariánsai; integráltételek szakaszban, az 505. oldalon értelmezi a *Stokes-tételt* másodrendű tenzorra is.

Hasonlóan jár el LAGALLY [14], aki a Kapitel 3. Theorie der Felder 88. Stokesscher Integralsatz szakaszában vektormezőre, 89. Allgemeiner Stokesscher Integral-

satz szakaszában, a (36") képlet utolsó sorában bármely Feldgrösse (tenzormező) esetére értelmezi a *Stokes-tételt.*

5.5. LÁMER [1] 101. oldala szerint: „KOZÁK [3] ... az összeférhetőségi peremfeltételeket WASHIZUétól eltérő alakban veszi fel."

Szerző nem felveszi az összeférhetőségi peremfeltételeket, hanem azokat többoldalúan is igazolja.

5.6. LÁMER [1] a 102. oldalon megállapítja: „ ... KOZÁK igazolta WASHIZU bizonyítás nélkül közölt állítását ...". Itt WASHIZU (4.5) alatti felvetéséről (nem állításáról) van szó.

Szerző munkáiban éppenhogy nem igazolta a (4.5) alatti felvetést, hanem ezzel ellenkezően azt igazolta, hogy az összeférhetőségi peremfeltételek száma nem 6, hanem 3, és azok (4.7) vagy a vele azonos (3.1) alakban adhatók meg.

5.7. LÁMER [1] a 100. oldalon a következőket mondja: „3°. A primál rendszerben is fennálló pontonkénti három peremfeltételhez milyen további három peremfeltételt kell csatolni ahhoz, hogy a duál feladat ... peremérték-feladatában a szükséges hat peremfeltétel adott legyen."

A kérdés ilyetén felvetése csak akkor indokolt, ha a teljes S peremen feszültségi peremfeltételt írunk elő. Az általánosabb esetben, amikor is a teljes perem egy részén feszültségi, más részén elmozdulási peremfeltételt írunk elő, a 4.4. szakasz szerint egy sajátos probléma jelentkezik. E probléma elkerülte LÁMER [1] figyelmét.

5.8. Végezetül még egy probléma. A LÁMER [1] tanulmány a 109. oldalon — a *Bianchi-azonosság* felhasználásával — az összeférhetőségi feltételek alábbi változatát adja meg a lineáris rugalmasságtan duál rendszeréhez. Ha

$$\varrho^{XY} = 0 \quad \text{a } V\text{-n} \tag{5.1}$$

és

$$\varrho^{AB} = 0 \quad \text{az } S\text{-en} \tag{5.2}$$

teljesül, akkor

$$\varrho^{AB} = 0 \quad \text{a } V\text{-n} \tag{5.3}$$

is teljesül (az indexkiválasztás a 4.3. szakaszban említettek szerint történhet). Nevezzük a fentieket LÁMER-féle megoldásnak és a (4.6–7) alattiakat KOZÁK-féle megoldásnak.

A LÁMER-féle megoldás helyességéhez nem férhet kétség, a kérdés mindössze az (5.2) peremfeltételek szükséges volta.

A duál rendszer összeférhetőségi feltételeinek a *Bianchi-azonosságot* felhasználó előállításában döntő szerepe van az S peremen felírt

$$\int_S \nu_k \varrho^{k\ell} n_\ell dS \tag{5.4}$$

integrál zérussá tételének, ha ν_k tetszőleges vektormező.

A KOZÁK-féle és a LÁMER-féle megoldás is azt állítja, hogy (4.6) és a vele azonos (5.1) fennállása esetén, ha az (5.4) integrál zérussá tehető, meghatározott kiválasztási szabály teljesülése esetén (4.8) és a vele azonos (5.3) is fennáll. Az eltérés a két megoldásban ott van, ahogy az (5.4) integrál zérus voltát biztosítják.

A KOZÁK-féle megoldás szerint az (5.4) integrál zérus voltához szükséges és elégséges, ha (4.7), vagyis a $\varrho^{k\ell}n_\ell = 0$ az S-en feltétel teljesül. Ezzel szemben a LÁMER-féle megoldás az integrál eltűnéséhez a (4.5), vagyis a $\varrho^{k\ell} = 0$, nem szükségszerű feltételt szabja (ismét a fixa idea).

A lemmát is figyelembe véve (4.7)-ből a

$$\varrho^{AB}n_B = 0 \quad \text{az } S\text{-en}, \tag{5.5}$$

míg (4.5)-ből az (5.2) alatti $\varrho^{AB} = 0$ az S-en összeférhetőségi peremfeltétel következik.

LÁMER [1] az (5.5) képlethez is fűz kritikai észrevételt (lásd a jelen szakasz utolsó bekezdését), ezért szerző részletesebben is foglalkozik (5.5)-tel.

Ha (5.2) teljesül, (5.5) is teljesül. Ezen felül (5.5) olyan megoldásokat is tartalmaz, amelyek (5.2)-ből nem következnek.

Pl., ha $(X,Y) = (1,2),(2,3),(3,1)$, vagyis $(A,B) = (1,1),(2,2),(3,3)$ és valamely peremrészen $n_1 = n_2 = 0$, (5.5)-ből az illető peremrészre két identikusan zérus és a $\varrho^{33}n_3 = 0$ skaláris egyenlet adódik. A peremrészen tehát $\varrho^{11}, \varrho^{22}$ tetszőleges volta mellett az egyetlen $\varrho^{33} = 0$ egyenlet az előírandó összeférhetőségi peremfeltétel. Lásd még a 3.2.3. pont utolsó bekezdését.

Egy másik esetben, ha $(X,Y) = (1,1),(2,2),(2,3)$, $(A,B) = (3,3),(3,1),(1,2)$ és valamely peremrészen $n_1 = 0$, $n_2 \neq 0$, $n_3 \neq 0$, az illető peremrészre (5.5)-ből a $\varrho^{33} = 0$ és a $n_2\varrho^{12} + n_3\varrho^{31} = 0$ egyenletek adódnak, egy egyenlet pedig identikusan zérus. Ugyanezen indexkiválasztásnál, de az $n_1 \neq 0$, $n_2 = 0$, $n_3 \neq 0$ normálisú peremrészen (5.5) megoldása: $\varrho^{33} = \varrho^{31} = \varrho^{12} = 0$.

Legyen egy további példában a gömbalakú testnél $x^1 = \varphi$, $x^2 = \vartheta$, $x^3 = r$ a három koordináta. A gömb teljes S felületén $n_\alpha = 0$ és $n_3 = 1$ a normális, és így $\varrho^{k3} = 0$ a teljes S-en az összeférhetőségi peremfeltétel. Az $(X,Y) = (1,1),(2,2),(1,3)$, $(A,B) = (1,2),(2,3),(3,3)$ indexpárokkal (5.5)-ből most a teljes peremre a $\varrho^{23} = 0$, $\varrho^{33} = 0$ az S-en, tehát két előírandó összeférhetőségi peremfeltétel adódik, míg a harmadik identikusan zérus. Természetesen teljesülnek a példában a (4.8) összeférhetőségi mezőegyenletek.

Érdemes a példák kapcsán külön hangsúlyozni, hogy 1 illetve 2 előírandó összeférhetőségi peremfeltétel mellett is teljesülhetnek, illetve teljesülnek a (4.6), (4.8) összeférhetőségi mezőegyenletek

Az előzőek szerint a KOZÁK-féle megoldás tartalmazza a LÁMER-féle megoldást, a fordított eset azonban nem áll fenn, következőleg a LÁMER-féle megoldás (5.2) összeférhetőségi peremfeltétele nem szükséges.

A felvetett példákhoz kapcsoljunk egy más jellegű példát is, és nézzük meg a

$$t^{k\ell}n_\ell = p^k \quad \text{az } S\text{-en}$$

feszültségi peremfeltételt ($t^{k\ell} = t^{\ell k}$ a feszültségi tenzor, p^{ℓ} a felületi terhelés), éspedig sík alakváltozási feladatnál (ahol $t^{11} \neq 0, t^{22} \neq 0, t^{33} \neq 0, t^{12} \neq 0, t^{13} \equiv 0$, $t^{23} \equiv 0$) az $n_1 = n_2 = 0$, $n_3 = 1$ normálisú S_{33} jelű peremrészen. Ekkor az identikusan zérus

$$t^{1\ell} n_\ell \equiv 0 \quad t^{2\ell} n_\ell \equiv 0 \qquad \text{az } S_{33}\text{-on}$$

és az érdemi

$$t^{3\ell} n_\ell = t^{33} = p^3 \qquad \text{az } S_{33}\text{-on}$$

feszültségi peremfeltételeket kapjuk.

A példák kapcsán szerző azt szeretné hangsúlyozni, ha a skaláris peremfeltételek egy része egyes peremrészeken (vagy akár a teljes peremen) identikusan zérus, ez nem jelenti azt, hogy ott nem áll a rendelkezésünkre elegendő számú peremfeltétel.

Térjünk vissza (5.5)-höz, amely a ϱ^{AB} koordinátákra nézve homogén, lineáris háromismeretlenes egyenletrendszer. (5.5) Δ determinánsa általában zérustól különböző. Ez esetekben (5.5) megoldása $\varrho^{AB} = 0$, vagyis azonos (5.2)-vel. Speciális esetekben $\Delta = 0$, amikor is az előírható összeférhetőségi peremfeltételek száma az illető peremrészen 2-re vagy 1-re csökken, míg a maradó 2 vagy 1 peremfeltétel identikusan teljesül.

Nem osztja tehát szerző LÁMER [1] állítását (113. oldal 3. bekezdés), amely szerint ha (5.5) determinánsa zérus, akkor (5.5) nem szolgáltat elegendő összeférhetőségi peremfeltételt, és így azt nem lehet használni. Mindig teljesül ugyanis (5.5)-tel együtt a teljes S peremen a (4.7) összeférhetőségi peremfeltétel, és így a (4.6) egyenletekkel együtt a (4.8) egyenlet is.

IRODALOM

[1] LÁMER, G., „A szükséges és elégséges összeférhetőségi peremfeltételek meghatározása", *Alkalmazott Matematikai Lapok* **16** (1992), 99–113.

[2] WASHIZU, K., „A note on the conditions of compatibility", *J. of Math. and Physics* **36 (4)** (1958), 306–312.

[3] KOZÁK, I., „A lineáris elasztostatika feszültségekkel felírt mezőegyenleteiről és peremfeltételeiről", *Műszaki Tudomány* **57** (1979), 423–446.

[4] KOZÁK, I., „Észrevételek és kiegészítések a lineáris elasztostatika feszültségfüggvényekkel felírt variációs elveihez", *Műszaki Tudomány* **57** (1979), 361–379.

[5] KOZÁK, I., „Notes on the field equations with stresses and on the boundary conditions in the linearized theory in elastostatics,", *Acta Technica Sci. Hung.* **90 (3-4)** (1980), 221–245.

[6] KOZÁK, I., „Determination of compatibility boundary conditions in linear elastostatics with the aid of the principle of minimum comlementary energy", *Publ. Techn. Univ. Heavy Industry Ser. D. Natural Sciences* **34** (1980), 83–98.

[7] KOZÁK, I., *Vékony héjak feszültségmezővel felépített elmélete*, Akadémiai doktori értekezés (Miskolc, 1980), 246.

[8] KOZÁK, I., „Principle of virtual work in terms of the stress functions", *Publ. Techn. Univ. Heavy Industry Ser. D. Natural Sciences* **34** (1981), 147–163.

[9] KOZÁK, I., „Remarks and contributions to the variational principles of the linearized theory of elasticity in terms of the stress functions", *Acta Technica Sci. Hung.* **92 (1-2)** (1981), 45–65.

[10] KOZÁK, I., „Construction of an approximate linear shell theory by asymptotic integration of the equations of elasticity in term of stresses", *Advances in Mechanics* **8 (1/2)** (1983), 91–110.

[11] BÉDA, GY., KOZÁK, I. és VERHÁS, J., *Kontinuummechanika* (Műszaki Könyvkiadó, Budapest, 1986), 246; angol nyelven: *Continuum Mechanics* (Akadémiai Kiadó, Budapest, 1995), 314.
[12] ERINGEN, A. C., *Nonlinear theory of continuous media* (McGraw-Hill Book Company, Inc., New York, San Francisco, Toronto, London, 1962), 477.
[13] BÉDA, GY. és KOZÁK, I., *Rugalmas testek mechanikája* (Műszaki Könyvkiadó, Budapest, 1987), 264.
[14] SZŐKEFALVI-NAGY, GY., GEHÉR, L. és NAGY, P., *Differenciálgeometria* (Műszaki Könyvkiadó, Budapest, 1979), 256.
[15] SOUTHWELL, R. V., „Castigliano's principle of minimum strain energy and the conditions of compatibility of strains", *S. Timoshenko 60th Anniversary volume* (The McMillan Co., 1983), 211–217.
[16] GRYCZ, J., „On the compatibility conditions in the classical theory of elasticity", *Archiwum Mechaniki Stosowanej* **6** (**19**) (1967), 883–891.
[17] ABOVSKIJ, N. P., ANDREEV, N. P. és DERUGA, A. P., *Variacionnüe principi teorii uprugosti i teorii obolocek* (Izd. Nauka, Moskva, 1978), 287.
[18] KORN, G. A. és KORN, T. M., *Matematikai kézikönyv műszakiaknak* (Műszaki Könyvkiadó, Budapest, 1975), 996.
[19] LAGALLY, M., *Vorlesungen über Vektorrechnung 5. Auflage* (Akademische Verlagsgesellschaft, Leipzig, 1956), 462.

(Beérkezett: 1995. március 1.)

KOZÁK IMRE
3525 MISKOLC, DÓZSA GY. U. 14.

REMARKS ON THE PAPER "DETERMINATION OF THE NECESSARY AND SUFFICIENT COMPATIBILITY CONDITIONS ON THE BOUNDARY" WRITTEN BY G. LÁMER

I. KOZÁK

The paper shortly presents the solution to the compabitibility problem of the mechanics of solid bodies both for a volume region and for a surface including the linear and the non-linear cases. Relations between the dual equation system of linear theory of elasticity and the compatibility conditions have also been investigated with a special regard to the boundary conditions. Then the paper reacts to some conclusions of the paper Lámer, but not to the whole of Lámer's paper.

Alkalmazott Matematikai Lapok **17** (1993) 347–371

KLASSZIKUS ÉS SÚLYOZOTT TUDÁSBÁZISOK TRANSZFORMÁCIÓI*

BENCZÚR ANDRÁS, B. NOVÁK ÁGNES, RÉVÉSZ Z. PÉTER

Ebben a cikkben a már ismert nulladrendű kijavító és felfrissítő operátorok ismertetése mellett részletesen elemezzük a kijavítással kapcsolatos problémákat. Ezen problémák elkerülésére bővítjük az eredeti axiómarendszert. A nulladrendben alkalmazható modell-illesztő operátor értelmezése után a kijavítás, modell-illesztés egy-egy lehetséges elsőrendű alkalmazására adunk példákat. Végül bevezetjük a súlyozott tudásbázis fogalmát, és az említett transzformációk közül a kijavítást és a modell-illesztést értelmezzük súlyozott tudásbázisokra is.

1. Bevezetés

Az utóbbi években számos cikk foglalkozott tudásbázisok módosításának problémájával. A tudásbázisban a „világot" leíró ismeretek rendszerint valamely nulladvagy elsőrendű logikai nyelv formuláival, ill. mondataival adottak. E formulák/mondatok összessége alkotja a tudásbázist. Így a tudásbázis egyetlen formulának tekinthető, amely az eredetileg szereplő formulák konjunkciója. Ennek megfelelően az egyszerűbb tárgyalás kedvéért a továbbiakban *klasszikus tudásbázis*, röviden *tudásbázis* alatt egy adott logikai nyelven leírt formulát értünk.

A feladat a következő: A φ formulával leírt tudásbázisba kell beépíteni egy (ugyanazon nyelven leírt) új ismeretet, melyet a μ formula (tudásbázis) reprezentál. A φ tudásbázis μ szerinti módosítása azonban sokféleképpen lehetséges.

ALCHOURRÓN, GÄRDENFORS és MAKINSON 1985-ben írott cikkükben [AGM85] elsőként fogalmaztak meg egy olyan axiómarendszert (a továbbiakban AGM axiómák), amelyet célszerű a módosító operátoroknak kielégíteni. Eszerint az axiómarendszer szerint pl. ha φ és μ konzisztensek, akkor a φ tudásbázis μ szerinti módosítása a $\varphi \wedge \mu$ formulával ekvivalens. Ha azonban φ és μ inkonzisztens, akkor olyan tudásbázist kell keresni, amely valamilyen szempontból hasonlít az eredetihez, ill. a megadott hasonlóság szempontjából legközelebb áll hozzá, és amelyben a μ formula igaz. Ez az elvárás lényegében az eredeti tudásbázis minimális változtatásának igényét fejezi ki. Bizonyos esetekben azonban az AGM axiómákat kielégítő operátorral kapott eredmény nem felel meg az intuitív elvárásoknak. Ezért KELLER és WINSLETT a gyakorlati alkalmazások szempontjából legfontosabbnak tartott két operátor közti különbséget fogalmazta meg informálisan [KW85]-ben. Ennek alapján KATSUNO és MENDELZON [KM91a] nulladrendű esetben formalizálták az operátorok e két családját. E cikkben részletesen foglalkoznak a korábbi cikkekben már megadott

*A dolgozat az OTKA T2149 szerződés támogatásával készült.

konkrét operátorok (pl. [D88], [S88], [We86], [Wi88]) ismertetésével, és a megadott axiómákhoz való viszonyukkal is.

Az egyik típusra vonatkozó axiómarendszer az AGM posztulátumok nulladrendű átírásával keletkezett. Ezeket az axiómákat kielégítő operátorokat *kijavító* (*revision*) operátoroknak nevezték el. A tudásbázis kijavításakor az az alapfeltevés, hogy a tudásbázis által leírt világ változatlan. Erről az állandó világról érkezik új információ, amelynek figyelembevétele akár gyökeresen is megváltoztathatja az eredeti tudásbázist.

A másik típusra megadott axiómarendszert kielégítő operátorok a tudásbázist *felfrissítő* (*update*) operátorok. Ezek állnak legközelebb a szokásos adatbázis alkalmazásokhoz, ui. a tudásbázis felfrissítésekor az a feltevés, hogy a tudásbázis által leírt világ változik, és ezt a változást kell a tudásbázisba beépíteni. Ilyen típusú változás pl.: bizonyos cikk ára 8 %-kal nőtt.

Mind a kijavítás, mind a felfrissítés esetében a tudásbázisok szerepe nem szimmetrikus. Ha φ az eredeti tudásbázis és μ az azt módosító, új ismeretet leíró formula, μ szerepe elsődleges abban az értelemben, hogy a módosítás után kapott tudásbázisnak μ logikai következménye, összhangban azzal az intuitív elvárással, hogy az új ismeretet feltétel nélkül elfogadjuk.

[R93]-ban nulladrendű tudásbázisokra új axiómarendszer bevezetésére került sor, amely súlyozott tudásbázisok esetében is értelmezhető. Ezen axiómarendszert kielégítő operátorok a *modell-illesztő* operátorok. Itt is igaz az új tudás elsőrendűsége a fentebb említett értelemben. A modell-illesztés egy alkalmazása a szimmetrikus modell-illesztés. E transzformációnál a tudásbázisok szerepe szimmetrikus.

[R93]-ban a fent említett operátorok alkalmazására a következő példa szerepel:

Egy bizonyítási eljárás során a bíróság a tanúk vallomásai alapján változtatja a bűntényre vonatkozó ismereteit.

Kijavításkor feltehető, hogy az új ismeret fontosabb és megbízhatóbb mint az adott pillanatban rendelkezésre álló tudásbázis. Pl. ha a tanúk sorrendje a kevésbé megbízhatótól a megbízható felé halad. A távoli rokon szerint a vádlott szociális ivó, míg a közeli hozzátartozó szerint alkoholista.

Felfrissítéskor az adat újabb, mint a pillanatnyilag rendelkezésre álló tudás. Pl. ha a tanúkat mondandójuk tartalmára vonatkozó időrendi sorrendben vonultatják fel. Az első tanú szerint a vádlott januárban fegyvert vásárolt, a második szerint pedig februárban eladta.

Szimmetrikus modell-illesztéskor az eredeti és az újonnan tudomásunkra hozott ismeret egyenrangúan kezelendő. Minden tanúvallomás egyformán fontos. Ez az eset pl. ha egy utcai baleset szemtanúi közül 9 azt állítja, hogy sárga volt a lámpa amikor az autó áthajtott a kereszteződésen, 5 szerint pedig már piros.

A továbbiakban a 2. fejezetben az alapfogalmak tisztázására kerül sor. A 3. fejezetben a [KM91]-ben bevezetett kijavító operátorok axiómarendszerét, és az azzal kapcsolatos problémákat elemezzük. A kijavító operátorok axiómarendszeréhez egy új axiómát csatolva a problémák egy része kiküszöbölhető. A 4. fejezet az [R93]-ban értelmezett modell-illesztés egy módosított változatáról szól. A felfrissítő

operátorok családját ismertetjük [KM91] alapján az 5. fejezetben. A 6. fejezetben az egyes operátorok elsőrendű alkalmazásának lehetőségeit vizsgáljuk. A 6.1 fejezet a [GMR92]-ben megadott elsőrendű felfrissítő operátorról szól. A 6.2 fejezetben egy új, konkrét elsőrendű kijavító operátort adunk meg. A 7.1 fejezetben [R93b]-ben bevezetett súlyozott tudásbázis fogalmát úgy módosítjuk, hogy az negáció kifejezésére is alkalmas legyen. A súlyozott tudásbázisok transzformációi közül azok kijavítását és modell-illesztését értelmezzük a 7.2, 7.3 fejezetekben. Végül a 8. fejezetben az ismertetett operátorokkal kapcsolatban felmerülő, még megoldatlan problémákat soroljuk fel.

2. Alapfogalmak

2.1 Szintaxis és szemantika

Legyen L_0 nulladrendű nyelv. Az L_0-beli ítéletváltozók halmaza T.

Jólformált formulák (a továbbiakban *formulák*) a szokásos módon a $\neg$, $\wedge$, $\vee$ szimbólumokkal képezhetők. Az $\alpha \rightarrow \beta$ formula a $\neg\alpha \vee \beta$ formula rövidítése, az $\alpha \leftrightarrow \beta$ pedig az $(\alpha \rightarrow \beta) \wedge (\beta \rightarrow \alpha)$ formuláé, ahol α és β formulák. A φ formulában előforduló ítéletváltozók halmaza $\mathrm{var}(\varphi)$.

A φ nulladrendű formula *interpretációjának* nevezzük a $\mathrm{var}(\varphi)$-n értelmezett, {igaz, hamis} halmazba képező függvényt. A φ formula interpretációit egyértelműen megadhatjuk azon ítéletváltozók halmazával, amelyeken az interpretáció az igaz értéket veszi fel. Ezért a továbbiakban a T részhalmazait interpretációnak nevezzük. Az összes interpretáció halmaza $\Im$.

A φ formula *igaz* egy adott interpretációban, ha a logikai összekötők szokásos értelmezésével kiértékelve igaz eredményt kapunk. Egyébként a formula *hamis*. Valamely formula *kielégíthetetlen*, ha minden interpretációban hamis. A formula *kielégíthető*, ha van olyan interpretáció, amelyben igaz. Azokat az interpretációkat, amelyekben a formula igaz, a *formula modelljeinek* nevezzük.

A φ formula modelljeinek halmazát $\mathrm{Mod}(\varphi)$ jelöli. Amennyiben A ítéletváltozó, minden A-t tartalmazó interpretáció modellje A-nak. Összetett formula modelljei a következőképpen származtathatók:

$$\mathrm{Mod}(\neg\varphi) = \Im \setminus \mathrm{Mod}(\varphi)$$
$$\mathrm{Mod}(\varphi \vee \mu) = \mathrm{Mod}(\varphi) \vee \mathrm{Mod}(\mu)$$
$$\mathrm{Mod}(\varphi \wedge \mu) = \mathrm{Mod}(\varphi) \wedge \mathrm{Mod}(\mu)$$

Ha $I_1, I_2, \ldots, I_k$ interpretációk, $\mathrm{form}(I_1, I_2, \ldots, I_k)$ jelöli azt a formulát és egyben azt a tudásbázist, amelynek pontosan $I_1, I_2, \ldots, I_k$ a modelljei. A TB rövidítés a lehetséges tudásbázisok halmazát jelöli.

Azt mondjuk, hogy φ *implikálja* μ-t, ha $\mathrm{Mod}(\varphi) \subseteq \mathrm{Mod}(\mu)$.

2.2 Interpretációk előrendezése

A minimális változtatás igényének megfelelően az eredeti tudásbázishoz bizonyos értelemben legközelebb álló modellek keresése a cél. A legközelebbi modellek megkereséséhez az interpretációk közötti előrendezés szükséges.

A $\leq$ reláció előrendezés az $\Im$-n, ha reflexív, tranzitív, és a következő tulajdonsággal rendelkezik: $I, J \in \Im$-re $I < J$ akkor és csak akkor, ha $I \leq J$ és $J \not\leq I$. Az előrendezés totális, ha minden $I, J \in \Im$-re vagy $I \leq J$ vagy $J \leq I$, egyébként az előrendezés parciális. Azt mondjuk, hogy $I = J$, ha $I \leq J$ és $J \leq I$ egyidejűleg teljesül.

Az I interpretáció a $\leq$ előrendezés szerint minimális a $H \subseteq \Im$ halmazon, ha $I \in H$, és bármely $J \in H$ esetén $I \leq J$. A H halmazon a $\leq$ előrendezés szerint minimális interpretációk halmazát $\mathrm{Min}\{H, \leq\}$ jelöli.

A későbbiekben látni fogjuk, hogy valamely operátor osztályba sorolását megkönnyítik az alábbi definiált függvények. A definíciókban az $\Im$-n értelmezett előrendezések halmazát ER-rel, a totális előrendezések halmazát TER-rel jelöljük. Szokásos módon D_f és R_f az f függvény értelmezési tartományát illetve értékkészletét jelenti.

2.2.1 Definíció. Az f függvényt *globálisan megbízható* függvénynek nevezzük, ha teljesülnek az alábbiak:

1. $D_f = TB$
2. $R_f \subseteq TER$
3. Ha $I \in M(\varphi)$ és $J \notin \mathrm{Mod}(\varphi)$, akkor $I <_\varphi J$ minden $\varphi \in D_f$ esetén
4. Ha $I, J \in \mathrm{Mod}(\varphi)$ akkor $I =_\varphi J$, minden $\varphi \in D_f$ esetén
5. Ha $\varphi_1 \longleftrightarrow \varphi_2$ akkor $f(\varphi_1) = f(\varphi_2)$.

Az f függvény lokálisan megbízható, ha az alábbi tulajdonságokkal rendelkezik:

1. $D_f = \Im$.
2. $R_f \subseteq ER$
3. Ha $I \not\equiv J$ akkor $I <_I J$ minden $I, J \in D_f$ esetén
4. Ha $I \equiv J$ akkor $I =_I J$ minden $I, J \in D_f$ esetén

($\mathrm{A} \equiv$ jel az azonosságot jelöli.)

Az eddigi definíciók a tudásbázisok, illetve az egyes interpretációkhoz rendelt előrendezések tulajdonságait rögzítették. Az alábbi definíció az egyes előrendezések egymáshoz való kapcsolatára vonatkozik.

2.2.2 Definíció. A $f : TB \to TER$ függvény *lojális*, ha teljesül, hogy

1.) Minden $\varphi_1, \varphi_2 \in D_g$-re, ha $g(\varphi_1) =\leq_{\varphi_1}, g(\varphi_2) =\leq_{\varphi_2}$ és $I \leq_{\varphi_1} J$, $I \leq_{\varphi_2} J$ akkor $I \leq_{\varphi_1 \vee \varphi_2} J$, ahol $g(\varphi_1 \vee \varphi_2) =\leq_{\varphi_1 \vee \varphi_2}$
2.) Ha $\varphi_1 \longleftrightarrow \varphi_2$ akkor $f(\varphi_1) = f(\varphi_2)$

3. Tudásbázis kijavítása

Jelölje φ az eredeti tudásbázist, μ pedig a φ által leírt világról új ismeretet adó formulát.

Jelölje $^\circ$ a kijavító operátort. A φ tudásbázis μ szerinti kijavításakor eredményül kapott új tudásbázis $\varphi^\circ\mu$.

KATSUNO és MENDELZON [KM91a]-ban az [AGM] posztulátumokat kielégítő nulladrendű tudásbázisokra alkalmazható $^\circ : TB \times TB \to TB$, ún. *kijavító* operátorokra a következő axiómarendszert fogalmazták meg:

[K1] $\varphi^\circ\mu$ implikálja μ-t

[K2] Ha $\varphi \wedge \mu$ kielégíthető, akkor $\varphi^\circ\mu \longleftrightarrow \varphi \wedge \mu$

[K3] Ha μ kielégíthető, akkor $\varphi^\circ\mu$ is kielégíthető.

[K4] Ha $\varphi_1 \longleftrightarrow \varphi_2$ és $\mu_1 \longleftrightarrow \mu_2$, akkor $(\varphi_1{}^\circ\mu_1) \longleftrightarrow (\varphi_2{}^\circ\mu_2)$

[K5] $(\varphi^\circ\mu) \wedge \nu$ implikálja $\varphi^\circ(\mu \wedge \nu)$-t

[K6] Ha $(\varphi^\circ\mu) \wedge \nu$ kielégíthető, akkor $\varphi^\circ(\mu \wedge \nu)$ implikálja $(\varphi^\circ\mu) \wedge \nu$-t

A [K1] axióma szerint az új ismeret visszakapható a kijavított tudásbázisból. Ez az axióma fejezi ki az új μ tudásnak az eredeti φ tudásbázishoz képest feltételezett „igazabb" voltát (feltétel nélküli elfogadását). Amennyiben φ és μ konzisztensek, akkor az eredmény az eredeti és az új tudásbázis közös modelljeire húzódva pontosabbá válik. A [K3] axióma a kijavítás igen lényeges tulajdonságát rögzíti, éspedig ha (az eredeti tudásbázis konzisztens voltától függetlenül) az új tudást hordozó μ formula kielégíthető, az eredményül kapott tudásbázis is az lesz. Ez tehát azt is biztosítja, hogy a konzistens tudásbázissal való kijavításkor nem keletkezhet inkonzisztencia. Ez is indokolja az operátor elnevezését. A [K4] axióma Dalal szintaxis-függetlenségi alapelvét fejezi ki, amely szerint a kijavítás eredményéül adódó tudásbázis jelentése független kell hogy legyen éppúgy az eredeti tudásbázis, mint maga a kijavítás szintaxisától. Bármely, tudásbázist módosító operátorról legyen szó, természetes az a törekvés, hogy az eredeti tudásbázisból a lehető legtöbb információ megmaradjon, másképpen, hogy az eredeti tudásbázis csak minimálisan változzon. A [K5]-[K6] axiómák azt fejezik ki, hogy φ és μ inkonzisztenciája esetén az eredeti tudásbázishoz legközelebbi modellek lesznek az új tudásbázis modelljei.

KATSUNO és MENDELZON [KM91b]-ben bizonyította az alábbi tételt:

3.1 TÉTEL. *A $^\circ : TB \times TB \to TB$ tudásbázist módosító operátor akkor és csak akkor elégíti ki a* [K1]-[K6] *axiómákat, ha van olyan globálisan megbízható f függvény, amelyre*

$$\mathrm{Mod}(\varphi^\circ\mu) = \mathrm{Min}\{\mathrm{Mod}(\mu), f(\varphi)\}. \qquad \square$$

Tehát a φ tudásbázis μ szerinti kijavításakor kapott új tudásbázis modelljei mindazok a μ modellek, amelyek a φ-hez rendelt $f(\varphi) = \leq_\varphi$ előrendezés szerint

minimálisak. Ebben az értelemben a kijavítás valóban a φ tudásbázis minimális megváltoztatásával, a φ-hez legközelebb álló modelleket választja ki. Mivel e közelség fogalma az előrendezésen múlik, a jól használható operátor megadásának feltétele a „jó " előrendezés megadása. Az alábbiakban egy-egy példát adunk a „jó ", ill. „rossz" előrendezésekre.

3.2 Példa. DALAL [D88a], [D88b] cikkeiben megadott operátor kielégíti a [K1]-[K6] axiómákat. Az előrendezés megadásához Dalal az interpretációk különbözőségét azok szimmetrikus differenciájának számosságával jellemezte:

$$\text{kül}(I, J) := |I \oplus J|,$$

ahol $\oplus$ jelöli a szimmetrikus differenciát: $I \oplus J = (I \setminus J) \cup (J \setminus I)$. Valamely interpretáció távolságát a φ tudásbázistól táv(φ, I) jelöli:

$$\text{táv}(\varphi, I) := \operatorname*{Mod}_{J \in \text{Mod}(\varphi)} \text{kül}(J, I).$$

Az előrendezés a fentiek alapján a következőképpen adható meg:

Az f_D függvény minden φ tudásbázishoz rendelje azt a $\leq_\varphi$ előrendezést, amely szerint $I \leq_\varphi J$ akkor és csak akkor, ha táv$(\varphi, I) \leq_\varphi$ táv(φ, J). Nyilvánvaló, hogy f_D globálisan megbízható függvény, ezért a

$$\text{Mod}(\varphi \circ \mu) = \text{Min}\{\text{Mod}(\mu), f_D(\varphi)\}$$

összefüggéssel megadott $\circ$ operátor valóban kijavító operátor. □

Ebben a példában látott globálisan megbízható függvény nemcsak a formális, hanem az intuitív elvárásoknak is megfelel. A 3.1 tétel segítségével azonban könnyen megadhatunk olyan globálisan megbízható függvényt, amely bizonyos tudásbázisok esetében éppen a Dalal-féle f_D függvény szerinti φ-től maximális távolságra lévő modelleket adja meg. A következő példa ezt illusztrálja.

3.3 Példa. Az f^* függvény legyen egyenlő az f_D függvénnyel azokon a φ tudásbázisokon, amelyekre $|\text{Mod}(\varphi)|$ páratlan. Ha pedig $|\text{Mod}(\varphi)|$ páros, akkor a következő, $\leq^*_\varphi$ előrendezést rendelje a φ tudásbázishoz:

$$I \leq^*_\varphi J \begin{cases} \text{ha táv}(\varphi, I) \geq \text{táv}(\varphi, J) \text{ és } I, J \notin \text{Mod}(\varphi) \\ \text{vagy} \\ I \in \text{Mod}(\varphi) \text{ és } J \notin \text{Mod}(\varphi) \end{cases}$$

$$I =^*_\varphi J \text{ ha } I, J \in \text{Mod}(\varphi).$$

f^* globálisan megbízható, és ha $|\text{Mod}(\varphi)|$ páros, akkor a kijavítás eredményeképpen kapott modellek éppen $\leq_\varphi$ szerint a φ-től legtávolabbi interpretációk, amennyiben nincsenek közös modellek. □

A kijavításra vonatkozó axiómarendszer szerinti függvényosztály tehát túlságosan tág, hiszen az előző példából látható, hogy csaknem tetszőleges előrendezéssel kielégíthető a globális megbízhatóság kritériuma, jóformán csak arra kell ügyelni, hogy ekvivalens tudásbázisokhoz ugyanazt az előrendezést adja a függvény. Ezen előnytelen tulajdonságot további axiómák bevezetésével lehet javítani.

Például ha a 3.1 tételben az f_D függvény segítségével adjuk meg a kijavító operátort, akkor az kielégíti a következő tulajdonságot is:

[K7] $(\varphi_1 \circ \mu) \wedge (\varphi_2 \circ \mu)$ implikálja $(\varphi_1 \vee \varphi_2) \circ \mu$-t.

Mivel az f_D függvény a gyakorlatban jól alkalmazható, célszerűnek látszik ezt a speciális tulajdonságot megkövetelni, és 7. axiómaként bevezetni. Az új, [K7] axióma független az eredeti [K1]-[K6] axiómáktól, hiszen pl. az f^* választással olyan függvényt adunk meg, amelyik a globális megbízhatósága miatt kielégíti a [K1]-[K6] axiómákat, de a [K7] axiómát nem. Ennek belátása a következő tétel felhasználásával igen egyszerű, ezért a tétel bizonyítása után térünk rá.

3.4 TÉTEL. *A* $\bullet : TB \times TB \to TB$ *tudásbázist módosító operátor akkor és csak akkor elégíti ki a* [K1]-[K7] *axiómákat, ha van olyan globálisan megbízható és lojális* f *függvény, amelyre*

$$\mathrm{Mod}(\varphi \bullet \mu) = \mathrm{Min}\{\mathrm{Mod}(\mu), f(\varphi)\}.$$

Bizonyítás. Az alábbi bizonyításokban a [KM91b]-ben található 3.1 tétel bizonyításában szereplő ötleteket, ill. [R93]-ban szereplő bizonyításokat használtuk fel.

I. Tegyük fel, hogy létezik egy operátor, amelyik kielégíti a [K1]-[K7] axiómákat. Az f függvény minden φ tudásbázishoz rendelje hozzá azt a $\leq_\varphi$ előrendezést, amelyre $I \leq_\varphi J$ akkor és csak akkor, ha $I \in \mathrm{Mod}(\varphi \bullet \mathrm{form}(I, J))$. Bizonyítandó, hogy

1.) f globálisan megbízható függvény
2.) $\mathrm{Mod}(\varphi \bullet \mu) = \mathrm{Min}\{\mathrm{Mod}(\mu), \leq_\varphi\}$.
3.) f lojális

1.) a.) $R_f \subseteq \mathrm{TER}$ ([R93] alapján)

A fent definiált $\leq_\varphi$ reláció **totális**: Ugyanis ha μ kielégíthető, akkor a [K1] és [K3] axiómák alapján $\mathrm{Mod}(\varphi \bullet \mathrm{form}(I, J))$ nem üres részhalmaza $\{I, J\}$-nek. Ezért minden I, J interpretáció összehasonlítható.

$\leq_\varphi$ **reflexív**: Az előzőhöz hasonlóan a [K1] és [K3] axiómák alapján $\mathrm{Mod}(\varphi \bullet \mathrm{form}(I))$ nem üres részhalmaza $\{I\}$-nek.

$\leq_\varphi$ **tranzitív**: Tegyük fel, hogy φ kielégíthető, és a $\leq_\varphi$ reláció nem tranzitív, vagyis van olyan I, J, K interpretáció, amelyre $I \leq_\varphi J$, $J \leq_\varphi K$ és $I \not\leq_\varphi K$. A $\leq_\varphi$ reláció definíciója miatt $I \notin \mathrm{Mod}(\varphi \bullet \mathrm{form}(I, K))$, ezért a [K5] axióma miatt $I \notin \mathrm{Mod}(\varphi \bullet \mathrm{form}(I, J, K)) \wedge \{I, K\}$, tehát $I \notin \mathrm{Mod}(\varphi \bullet \mathrm{form}(I, J, K))$.

Ha $J \in \mathrm{Mod}(\varphi \bullet \mathrm{form}(I,J,K))$, akkor $(\varphi \bullet \mathrm{form}(I,J,K)) \wedge \mathrm{form}(I,J)$ kielégíthető. Mivel $I \notin \mathrm{Mod}(\varphi \bullet \mathrm{form}(I,J,K))$, ezért $I \notin \mathrm{Mod}(\varphi \bullet \mathrm{form}(I,J,K)) \wedge \{I,J\}$. A [K6] axióma miatt így $I \notin \mathrm{Mod}(\varphi \bullet \mathrm{form}(I,J))$. Ez pedig ellentmond az $I \leq_\varphi J$ feltevésnek.

Ha $J \notin \mathrm{Mod}((\varphi \bullet \mathrm{form}(I,J,K))$, akkor a [K1] valamint a [K3] axiómákból $K = \mathrm{Mod}((\varphi \bullet \mathrm{form}(I,J,K))$ következik. Így a $(\varphi \bullet \mathrm{form}(I,J,K) \wedge \mathrm{form}(J,K))$ kielégíthető. [K6]-ot felhasználva, mivel $J \notin \mathrm{Mod}(\varphi \bullet \mathrm{form}(I,J,K)) \wedge \{J,K\}$, ezért $J \notin \mathrm{Mod}((\varphi \bullet \mathrm{form}(J,K))$, ami ellentmond a $J \leq_\varphi K$ feltevésnek.

1.) b) $I, J \in \mathrm{Mod}(\varphi)$, akkor $I =_\varphi J$:
Mivel mindkét interpretáció modellje φ-nek, ezért a [K2] axióma miatt $\mathrm{Mod}(\varphi \bullet \mathrm{form}(I,J)) = \{I,J\}$. Tehát $I, J \in \mathrm{Mod}(\varphi \bullet \mathrm{form}(I,J))$, ami a $\leq_\varphi$ definíciója alapján azt jelenti, hogy $I \leq_\varphi J$ és $J \leq_\varphi I$.

1.) c.) $I \in \mathrm{Mod}(\varphi)$, és $J \notin \mathrm{Mod}(\varphi)$, akkor $I <_\varphi J$:
A [K2] axióma miatt $\mathrm{Mod}(\varphi \bullet \mathrm{form}(I,J)) = \{I\}$, így $I \leq_\varphi J$ teljesül. Mivel $J \notin \mathrm{Mod}(\varphi \bullet \mathrm{form}(I,J))$, így $J \not\leq_\varphi I$, tehát valóban $I <_\varphi J$ fennáll.

1.) d.) Ha $\varphi_1 \longleftrightarrow \varphi_2$, akkor $f(\varphi_1) = f(\varphi_2)$:
Nyilvánvalóan teljesül a [K4] axióma miatt.

2.) Amennyiben μ kielégíthetetlen, a [K1] axióma felhasználásával $\mathrm{Mod}(\varphi \bullet \mu) = \emptyset = \mathrm{Min}\{\mathrm{Mod}(\mu), \leq_\varphi\}$, tehát igaz az állítás. A továbbiakban feltesszük, hogy a μ kielégíthető. Az egyenlőséget a kétirányú tartalmazás bizonyításával látjuk be.

2.) a.) $\mathrm{Mod}(\varphi \bullet \mu) \subseteq \mathrm{Min}\{\mathrm{Mod}(\mu), \leq_\varphi\}$:
Tegyük fel az állítással ellentétben, hogy $I \in \mathrm{Mod}(\varphi \bullet \mu)$, és $I \notin \mathrm{Min}\{\mathrm{Mod}(\mu), \leq_\varphi\}$. Akkor létezik legalább egy olyan $J \in \mathrm{Mod}(\mu)$ interpretáció, amelyre $J < I$ teljesül. A $\leq_\varphi$ előrendezés definíciója szerint ekkor $\mathrm{Mod}(\varphi \bullet \mathrm{form}(I,J)) = \{J\}$. Mivel I és J mindegyike modellje μ-nek, $\mu \wedge \mathrm{form}(I,J) = \mathrm{form}(I,J)$. A [K5] axióma szerint $\mathrm{Mod}(\varphi \bullet \mu) \wedge \{I,J\} \subseteq \mathrm{Mod}(\varphi \bullet (\mu \wedge \mathrm{form}(I,J))) = \mathrm{Mod}(\varphi \bullet \mathrm{form}(I,J) = \{J\}$. Vagyis, $I \notin \mathrm{Mod}(\varphi \bullet \mu)$, ami ellentmondás.

2.) b.) $\mathrm{Min}\{\mathrm{Mod}(\mu), \leq_\varphi\} \subseteq \mathrm{Mod}(\varphi \bullet \mu)$:
Tegyük fel az állítással ellentétben, hogy $I \in \mathrm{Min}\{\mathrm{Mod}(\mu), \leq_\varphi\}$, és $I \notin \mathrm{Mod}(\varphi \bullet \mu)$. A [K3] axiómából következik, hogy van $(\varphi \bullet \mu)$-nek modellje, legyen ez J. [K1] miatt J a μ-nek is modellje. Így $\mu \wedge \mathrm{form}(I,J) = \mathrm{form}(I,J)$. A [K5] és [K6] axiómákat felhasználva $\mathrm{Mod}((\varphi \bullet \mu) \wedge \mathrm{form}(I,J)) \subseteq \mathrm{Mod}(\varphi \bullet \mu) \wedge \{I,J\} = \mathrm{Mod}(\varphi \bullet \mathrm{form}(I,J))$. A [K1] és [K3] axiómák szerint $\mathrm{Mod}(\varphi \bullet \mathrm{form}(I,J))$ nemüres részhalmaza $\{I,J\}$-nek. Mivel $I \notin \mathrm{Mod}(\varphi \bullet \mu)$, ezért $I \notin \mathrm{Mod}(\varphi \bullet \mathrm{form}(I,J))$, tehát $I \not\leq_\varphi J$. Ez a kiindulási feltétellel ellentétben éppen azt jelenti, hogy I nem lehet $\mathrm{Mod}(\mu)$-nek a $\leq_\varphi$ előrendezés szerinti minimális eleme.

3.) Mivel f globálisan megbízható függvény, ezért csak a lojalitás 1.) tulajdonságát kell bizonyítani. Tegyük fel, hogy $I \leq_{\varphi_1} J$ és $I \leq_{\varphi_2} J$.

Akkor $I \in \mathrm{Mod}(\varphi_1 \bullet \mathrm{form}(I,J))$, és $I \in \mathrm{Mod}(\varphi_2 \bullet \mathrm{form}(I,J))$. A [K7] axióma szerint $I \in \mathrm{Mod}((\varphi_1 \vee \varphi_2) \bullet \mathrm{form}(I,J))$, vagyis $I \leq_{\varphi_1 \vee \varphi_2} J$, ami éppen azt jelenti, hogy az f függvény lojális.

II. Tegyük fel most, hogy van olyan f lojális és globálisan megbízható függvény, amely minden φ tudásbázishoz a $\leq_\varphi$ előrendezést rendeli. Ekkor bizonyítandó, hogy a

$$\mathrm{Mod}(\varphi \bullet \mu) = \mathrm{Min}\{\mathrm{Mod}(\mu), \leq_\varphi\}$$

összefüggéssel definiált $\bullet$ operátor kielégíti a [K1]-[K7] axiómákat.

A [K1] axióma teljesülése a $\bullet$ operátor definíciójából közvetlenül látható.

[K2]-t a kétirányú tartalmazás bizonyításával látjuk be. Feltesszük, hogy $\varphi \wedge \mu$ kielégíthető.

Először azt bizonyítjuk, hogy

$$\mathrm{Mod}(\varphi \wedge \mu) \subseteq \mathrm{Min}\{\mathrm{Mod}(\mu), \leq_\varphi\}.$$

Legyen $I \in \mathrm{Mod}(\varphi \wedge \mu)$. Mivel f globálisan megbízható függvény, a 3. tulajdonságból következik, hogy ha $I \in \mathrm{Mod}(\varphi)$, akkor $I <_\varphi J$ teljesül minden olyan J interpretációra, amelyre $J \notin \mathrm{Mod}(\varphi)$. Másrészt $I \in \mathrm{Mod}(\mu)$ is igaz, ezért ha $I \in \mathrm{Mod}(\varphi \wedge \mu)$, akkor $I \in \mathrm{Min}\{\mathrm{Mod}(\mu(, \leq_\varphi\}$.

Most azt látjuk be, hogy $\mathrm{Min}\{\mathrm{Mod}(\mu), \leq_\varphi\} \subseteq \mathrm{Mod}(\varphi \wedge \mu)$. Tegyük fel, hogy van olyan I interpretáció, amelyre $I \in \mathrm{Min}\{\mathrm{Mod}(\mu), \leq_\varphi\}$ fennáll, de $I \notin \mathrm{Mod}(\varphi \wedge \mu)$. Mivel $\varphi \wedge \mu$ kielégíthető, van modellje, legyen ez J. A globális megbízhatóság miatt $J <_\varphi I$, hiszen $J \in \mathrm{Mod}(\varphi)$ és $I \notin \mathrm{Mod}(\varphi)$. Ezért I nem lehet minimális eleme a $\mathrm{Mod}(\mu)$ halmaznak, ami ellentmondás.

A [K3] axióma teljesülése a definíció alapján nyilvánvaló.

A [K4] axióma a globális megbízhatóság 5. (ill. a lojalitás 2.) tulajdonságának egyenes következménye.

A [K5] és [K6] axiómák belátásához tegyük fel, hogy $(\varphi \bullet \mu) \wedge \nu$ kielégíthető, hiszen ellenkező esetben a [K5] axióma nyilvánvalóan teljesül.

Bizonyítandó, hogy $\mathrm{Mod}((\varphi^\circ \mu) \wedge \nu) \subseteq \mathrm{Mod}(\varphi^\circ(\mu \wedge \nu))$, vagyis, hogy $\mathrm{Min}\{\mathrm{Mod}(\mu), \leq_\varphi\} \cap \mathrm{Mod}(\nu) \subseteq \mathrm{Min}\{\mathrm{Mod}(\mu \wedge \nu), \leq_\varphi\}$. Tegyük fel, hogy $I \in \mathrm{Min}\{\mathrm{Mod}(\mu), \leq_\varphi\} \cap \mathrm{Mod}(\nu)$. Ekkor $I \in \mathrm{Min}\{\mathrm{Mod}(\mu \wedge \nu), \leq_\varphi\}$ is teljesül, hiszen ha nem így lenne, létezne olyan $J \in \mathrm{Mod}(\mu \wedge \nu)$, amelyre $J <_\varphi I$. Mivel egyúttal $J \in \mathrm{Mod}(\mu)$, ezért I nem lehetne minimális $\mathrm{Mod}(\mu)$-ben.

Most a másik irányú tartalmazást bizonyítjuk:
$\mathrm{Mod}(\varphi^\circ(\mu \wedge \nu)) \subseteq \mathrm{Mod}((\varphi^\circ \mu) \wedge \nu)$, ami azt jelenti, hogy
$\mathrm{Min}\{\mathrm{Mod}(\mu \wedge \nu), \leq_\varphi\} \subseteq \mathrm{Min}\{\mathrm{Mod}(\mu), \leq_\varphi\} \cap \mathrm{Mod}(\nu)$.
Tegyük fel, hogy $I \in \mathrm{Min}\{\mathrm{Mod}(\mu \wedge \nu), \leq_\varphi\}$, és $I \notin \mathrm{Min}\{\mathrm{Mod}(\mu), \leq_\varphi\} \cap \mathrm{Mod}(\nu)$. Mivel $I \in \mathrm{Mod}(\nu)$, ezért $I \notin \mathrm{Min}\{\mathrm{Mod}(\mu), \leq_\varphi\}$. A $(\varphi^\circ \mu) \wedge \nu$ tudásbázis kielégíthetősége miatt van olyan J interpretáció, amelyre $J \in \mathrm{Min}\{\mathrm{Mod}(\mu), \leq_\varphi\} \cap \mathrm{Mod}(\nu)$. Így $J \in \mathrm{Mod}(\mu \wedge \nu)$, és így $I \in \mathrm{Min}\{\mathrm{Mod}(\mu \wedge \nu), \leq_\varphi\}$ miatt $I \leq_\varphi J$. Ez ellentmond a $I \notin \mathrm{Min}\{\mathrm{Mod}(\mu), \leq_\varphi\}$ állításnak.

A [K7] axióma pedig a lojalitás alapján a következőképpen látható be: Ha $I \in \mathrm{Min}\{\mathrm{Mod}(\mu), \leq_{\varphi_1}\}$, és $I \in \mathrm{Min}\{\mathrm{Mod}(\mu), \leq_{\varphi_2}\}$, akkor $I \leq_{\varphi_1} J$ és $I \leq_{\varphi_2} J$ minden más $J \in \mathrm{Mod}(\mu)$ interpretációra. Ekkor a lojalitás miatt $I \leq_{\varphi_1 \vee \varphi_2} J$ fennáll, tehát $I \in \mathrm{Min}\{\mathrm{Mod}(\mu), \leq_{\varphi_1 \vee \varphi_2}\}$, így a [K7] axióma teljesül. □

Most rátérünk annak bizonyítására, hogy a [K7] axióma független a [K1]-[K6] axiómáktól.

A 3.3 példához hasonló, a jogos elvárásokat nem kielégítő megoldások egy részét tehát a [K7] axiómával ki lehet küszöbölni. A 3.3 példabeli f^* függvény esetében legyen φ is és μ is olyan tudásbázis, amelyeknek modellhalmaza páratlan elemszámú, és tegyük fel továbbá, hogy nincsen közös modelljük. Legyen $I \leq^*_\varphi J$ és $I \leq^*_\mu J$, ahol $\leq^*_\varphi$ és $\leq^*_\mu$ az f^* függvény által a φ és a μ tudásbázisokhoz rendelt előrendezések. f^* definíciója szerint akkor $I \leq_\varphi J$ és $I \leq_\mu J$ is teljesül, ahol $\leq_\varphi$ és $\leq_\mu$ az f_D függvény megfelelő értékei φ-n és μ-n. Mivel f_D lojális, ezért $I \leq_{\varphi \vee \mu} J$. De $|\mathrm{Mod}(\varphi \vee \mu)|$ nyilván páros, így $J \leq^*_{\varphi \vee \mu} I$. Ezért f^* nem lehet lojális. Ezzel beláttuk, hogy a [K7] axióma független a [K1]-[K6] axiómáktól.

4. Tudásbázisok modell-illesztése

[R93] alapján egy új operátort adunk meg. A kijavító operátorra vonatkozó axiómák szigorúak abban az értelemben, hogy a $\varphi \wedge \mu$ kielégíthetősége esetén csakis ez lehet a transzformáció eredménye. Többek között ezen követelményt elhagyva és más tulajdonságokat megkövetelve, a gyakorlatban szintén jól alkalmazható transzformációhoz jutunk.

Az alábbi, [M1]-[M7] axiómák megfogalmazásával az a cél, hogy lehetőség legyen adott φ, μ tudásbázisokhoz olyan $\varphi \nabla \mu$ operátor megadására, amelynek eredménye legjobban illeszkedik az eredeti φ tudásbázishoz úgy, hogy a μ tudásbázist implikálja. Ez utóbbi indokolja az [M1] axiómát, amely az eddig ismertetett operátornál is szerepel. Az [M2] axióma új: azt fejezi ki, hogyha a φ tudásbázis kielégíthetetlen, akkor nem lehet hozzá illeszteni semmilyen más modellt. Az [M3] axióma a konzisztencia megmaradását biztosítja. A többi axióma a kijavításnál ismertetett [K3]-[K7] axiómáknak felelnek meg.

A $\nabla : TB \times TB \to TB$ operátort *modell-illesztő* operátornak nevezzük, ha kielégíti az alábbi [M1]-[M7] axiómákat:

[M1] $\varphi \nabla \mu$ implikálja μ-t

[M2] Ha φ kielégíthetetlen, akkor $\varphi \nabla \mu$ is az

[M3] Ha φ és μ mindegyike kielégíthető, akkor $\varphi \nabla \mu$ is az

[M4] Ha $\varphi_1 \longleftrightarrow \varphi_2$ és $\mu_1 \longleftrightarrow \mu_2$, akkor $\varphi_1 \nabla \mu_1 \longleftrightarrow \varphi_2 \nabla \mu_2$

[M5] $(\varphi \nabla \mu) \wedge \nu$ implikálja $\varphi \nabla (\mu \wedge \nu)$-t

[M6] Ha $(\varphi \nabla \mu) \wedge \nu$ kielégíthető, akkor $\varphi \nabla (\mu \wedge \nu)$ implikálja $(\varphi \nabla \mu) \wedge \nu$-t

[M7] $(\varphi_1 \nabla \mu) \wedge (\varphi_2 \nabla \mu)$ implikálja $(\varphi_1 \vee \varphi_2) \nabla \mu$-t

Az előző transzformációnál ismertetett 3.4 tételhez hasonlóan e transzformációra is érvényes a következő tétel:

4.1 TÉTEL. *A $\nabla : TB \times TB \to TB$ operátor akkor és csak akkor elégíti ki az* [M1]-[M7] *axiómákat, ha van olyan lojális függvény, amely a φ tudásbázishoz a $\leq_\varphi$ előrendezést rendeli, és amelyre*

$$\mathrm{Mod}(\varphi \nabla \mu) := \mathrm{Min}\{\mathrm{Mod}(\mu), \leq_\varphi\} \text{ teljesül.}$$

Bizonyítás. I. Tegyük fel, hogy a ∇ operátor kielégíti az [M1]-[M7] axiómákat. Az f függvény minden φ tudásbázishoz rendelje hozzá azt a $\leq_\varphi$ előrendezést, amelyre $I \leq_\varphi J$ akkor és csak akkor, ha $I \in \mathrm{Mod}(\varphi \nabla \mathrm{form}(I, J))$. Bizonyítandó, hogy

1.) f lojális

2.) $\mathrm{Mod}(\varphi \nabla \mu) = \mathrm{Min}\{\mathrm{Mod}(\mu), \leq_\varphi\}$.

Ezen tulajdonságok bizonyítása megegyezik a 3.4 tétel bizonyításának megfelelő I.1.a.), I.1.d), 2., 3. pontjával.

II. Tegyük fel most, hogy van olyan f lojális függvény, amely minden φ tudásbázishoz a $\leq_\varphi$ előrendezést rendeli. Ekkor bizonyítandó, hogy a $\nabla : TB \times TB \to TB$,

$$\mathrm{Mod}(\varphi \nabla \mu) = \mathrm{Min}\{\mathrm{Mod}(\mu), \leq_\varphi\}$$

összefüggéssel definiált ∇ operátor kielégíti az [M1]-[M7] axiómákat.

Az [M1] axióma teljesülése a ∇ operátor definíciójából közvetlenül látható.

Az [M2] axióma teljesül, hiszen ha φ kielégíthetetlen, akkor a minimális modellek halmaza üres.

Az [M3] axióma abból következik, hogy ha φ és μ kielégíthető, akkor mindig van olyan modellje μ-nek, amely minimális.

Az [M4] axióma a lojalitás 2. tulajdonságának közvetlen következménye.

Az [M5], [M6], [M7] axiómák teljesülésének bizonyítása megegyezik a 3.4 tétel II. részében a [K5], [K6], [K7] axiómák teljesülésének bizonyításával. □

4.2 Példa.

a.) Az f_D függvény szerint képezett operátor modell-illesztő operátor, hiszen f_D lojális.

b.) Az f_D függvény [R93]-beli alább ismertetett módosítása lojális, ezért az eszerint képezett operátor modell-illesztő operátor:

$$\mathrm{kül}(I, J) := |I \oplus J|$$

$$\mathrm{táv}^*(\varphi, I) := \mathrm{Max}\left\{ \underset{J \in \mathrm{Mod}(\varphi)}{\mathrm{kül}}(J, I) \right\}$$

Pl. tegyük fel, hogy egy csoportot tanító tanár vagy csak Datalogot (D), vagy Datalogot (D) és SQL-t (S) akar tanítani. A hallgatók viszont (egyenlő számban) vagy csak SQL-t, vagy csak Datalogot, vagy SQL-t, Datalogot és Query-by-Example-t (Q) akarnak tanulni.
A tanár álláspontja a $\mu = (\neg S \wedge D \wedge \neg Q) \vee (S \wedge D \wedge \neg Q)$ formulával, a hallgatók igénye pedig a $\varphi = (S \wedge \neg D \wedge \neg Q) \vee (\neg S \wedge D \wedge \neg Q) \vee (S \wedge D \wedge Q)$ formulával írható le. A megfelelő modellek:

$$\mathrm{Mod}(\mu) = \{\{D\}, \{S, D\}\},$$
$$\mathrm{Mod}(\varphi) = \{\{S\}, \{D\}, \{S, D, Q\}\}.$$

Az egyes távolságok kiszámításával az eredmény:

$$\mathrm{Mod}(\varphi \nabla \mu) = \mathrm{Min}\{\mathrm{Mod}(\mu), \leq_\varphi\} = \{\{S, D\}\},$$

ami azt jelenti, hogy a tanár akkor dönt helyesen, ha mind SQL-t, mind Datalogot tanít. □

5. Tudásbázis felfrissítése

A 3. fejezetben ismertetett kijavításkor amennyiben az eredeti φ tudásbázis és az új μ tudásbázis inkonzisztens, úgy a φ modelljei közül az inkonzisztenciát okozó modellek az eredményben nem szerepelnek, semmilyen módon nem is lehet rájuk következtetni. Többek között ez a tulajdonság az, ami miatt a tudásbázis felfrissítésének igénye jelentkezett. Frissítéskor az a törekvés, hogy a lehetséges világokat (vagyis az eredeti tudásbázis modelljeit) egymától függetlenül vizsgálva eljuthassuk az eredményig, és így minden modellből tartalmazzon valamilyen információt a felfrissített tudásbázis. Ezért a φ tudásbázis módosításakor annak külön-külön minden egyes modelljéhez a legközelebb álló μ-modelleket keresünk. A felfrissítés így a pontonként legközelebbi μ-modelleket adja eredményül, míg a kijavítás a tudásbázis egészéhez viszonyított legközelebbi modellt választja ki.

Jelentse φ az eredeti tudásbázist, μ pedig az új ismeretet leíró tudásbázist. A $\diamond : TB \times TB \to TB$ *felfrissítő* operátorra vonatkozó axiómák [KM91a] cikke nyomán a következők:

[F1] $\varphi \diamond \mu$ implikálja μ-t

[F2] Ha φ implikálja μ-t, akkor $\varphi \diamond \mu$ ekvivalens φ-vel

[F3] Ha φ és μ mindegyike kielégíthető, akkor $\varphi \diamond \mu$ is az

[F4] Ha $\varphi_1 \longleftrightarrow \varphi_2$ és $\mu_1 \longleftrightarrow \mu_2$, akkor $\varphi_1 \diamond \mu_1 \longleftrightarrow \varphi_2 \diamond \mu_2$

[F5] $(\varphi \diamond \mu) \wedge \nu$ implikálja $\varphi \diamond (\mu \wedge \nu)$-t

[F6] $\varphi \diamond \mu_1$ implikálja μ_2-t, és $\varphi \diamond \mu_2$ implikálja μ_1-et, akkor $\varphi \diamond \mu_1 \longleftrightarrow \varphi \diamond \mu_2$

[F7] Ha $|\mathrm{Mod}(\varphi)| = 1$, akkor $(\varphi \diamond \mu_1) \wedge (\varphi \diamond \mu_2)$ implikálja $\varphi \diamond (\mu_1 \vee \mu_2)$-t

[F8] $(\varphi_1 \vee \varphi_2) \diamond \mu$ ekvivalens $(\varphi_1 \diamond \mu) \vee (\varphi_2 \diamond \mu)$

Az [F1], [F4], [F5] axiómák rendre a [K1], [K4], [K5] axiómák megfelelői. Az [F2] axióma azonban lényegesen különbözik [K2]-től. [F2] következményeképpen, ha az eredeti tudásbázis inkonzisztens, akkor a felfrissített tudás is az. Tehát a tudásbázisba valamilyen módon bekerült ellentmondás felfrissítéssel nem hozható helyre. Ez összhangban van azzal az intuitív eljárással, hogy a felfrissítés a (φ által tükrözött) világ változásait viszi be a tudásbázisba. Ha azonban nincsen olyan lehetséges világ, amely megfelelne a φ tudásbázisban rögzítetteknek, akkor nyilván nincs mód arra, hogy a valós világ μ-vel (helyesen) megfogalmazott változásait tükrözze a felfrissített tudásbázis. A kijavítás esetében ez a probléma nem áll fenn, hiszen a μ formula mintegy felülírja a φ tudásbázist. Az [F3] axióma a konzisztencia megmaradását biztosítja. A [K6] axióma helyett három új axióma szerepel. [F6] szerint ha a μ_1 formula szerinti felfrissítés maga után vonja a μ_2 szerintit, és viszont, akkor a μ_1 és a μ_2 szerinti felfrissítések ekvivalensek. Az [F7] axióma csak szingletonokra alkalmazható. Eszerint, a résztudások alapján felfrissítéssel kapott közös világok a résztudásokkal vagylagosan felfrissített a tudásbázisban is megmaradnak. Végül az [F8] axióma biztosítja az egyes modellek független felfrissítését.

[KM91a]-ban a szerzők többek között a következő tételt bizonyították:

5.1 TÉTEL. *A* $\diamond : TB \times TB \to TB$ *operátor akkor és csak akkor elégíti ki az* [F1]-[F8] *axiómákat, ha létezik olyan* f *lokálisan megbízható függvény, amelyre*

$$\mathrm{Mod}(\varphi \diamond \mu) = \bigcup_{I \in \mathrm{Mod}(\varphi)} \mathrm{Min}\{\mathrm{Mod}(\mu), f(I)\}. \qquad \square$$

5.2 Példa. M. WINSLETT [Wi88]-ban elsőrendű nyelven megadott tudásbázisok módosítására adott meg egy operátort, amelyet nulladrendre átírva kielégíti az [F1]-[F8] axiómákat. Két interpretáció különbözőségét jelölje $\mathrm{kül}(I, J) := I \oplus J$. Az f függvény minden egyes I interpretációhoz a következő $f(I) := \leq_I$ előrendezést rendelje hozzá: $J_1 \leq_I J_2$ akkor és csak akkor, ha $\mathrm{kül}(I, J_1) \subseteq \mathrm{kül}(I, J)_2$). Mivel az f függvény nyilván lokálisan megbízható minden $I \in \mathfrak{I}$-n ezért a

$$\mathrm{Mod}(\varphi \diamond \mu) = \bigcup_{I \in \mathrm{Mod}(\varphi)} \mathrm{Min}\{\mathrm{Mod}(\mu), f(I)\}$$

összefüggéssel megadott $\diamond$ operátor valóban felfrissítő operátor. $\square$

Ha nem áll rendelkezésre egyértelmű utalás arra vonatkozóan, hogy a régi tudásbázis helyesen írja-e le a világot, és az új ismeret a világ változását hordozza-e, vagy pedig az eredeti tudásbázis helytelenségére derül-e fény az új információból, akkor nehéz eldönteni, hogy pl. a kijavító és felfrissítő operátorok közül melyiket célszerű alkalmazni. Ezt illusztrálja az alábbi példa, GINSBERG és SMITH [GS87], WINSLETT [Wi89], valamint KATSUNO és MENDELZON [KM91] cikkeinek alapján.

5.3 Példa. Tekintsük a következő tudásbázisokat:

$$\varphi := (t_1 \wedge \neg t_2 \wedge \neg t_3 \wedge \neg t_4 \wedge \neg t_5) \vee (t_1 \wedge \neg t_2 \wedge \neg t_3 \wedge t_4 \wedge t_5)$$
$$\mu := (t_1 \wedge t_2 \wedge t_3 \wedge t_4 \wedge t_5) \vee (\neg t_1 \wedge \neg t_2 \wedge \neg t_3 \wedge \neg t_4 \wedge \neg t_5).$$

A megfelelő modellek:

$$\mathrm{Mod}(\varphi) = \{\{t_1\}, \{t_1, t_4, t_5\}\}$$
$$\mathrm{Mod}(\mu) = \{\{\ \}, \{t_1, t_2, t_3, t_4, t_5\}\}.$$

A 3.2 példában szereplő kijavító operátort alkalmazva az alábbi modellhalmazt kapjuk eredményül:

$$\mathrm{Mod}(\varphi \circ \mu) = \{\{\}\}.$$

Az 5.2 példában ismertetett felfrissítő operátorral pedig a következő eredményhez jutunk:

$$\mathrm{Mod}(\varphi \diamond \mu) = \{\ \{t_1, t_2, t_3, t_4, t_5\},\ \{\}\ \}.$$

Tegyük fel, hogy a leírandó világ egy szoba, melyben egy asztal, és azon kívül öt tárgy található. A tárgyak az asztalon vagy a padlón helyezkedhetnek el. A t_i mondat jelentse azt, hogy az i. tárgy az asztalon van. A φ tudásbázis így azt jelenti, hogy vagy az 1. tárgy, vagy az 1., 4., 5. tárgy az asztalon van. A szobába küldenek egy robotot azzal a feladattal, hogy rendezze el úgy a tárgyakat, hogy vagy mindegyik az asztalon, vagy mindegyik a padlón helyezkedjen el. Ez éppen a μ formula jelentése. A kijavítással kapott eredmény szerint minden tárgy a padlón van. Ez az eredmény nem fogadható el, hiszen a robot feladatának végrehajtása után lehetséges, hogy mindegyik tárgy az asztalon lesz. A felfrissítő operátorral éppen ezt az eredmény kapjuk, vagyis ebben az esetben a formulák jelentését figyelembe véve a φ tudásbázist az új ismeretet tükröző φ formula szerint felfrissíteni kell.

Más választás adódhat, ha a formulákat a következőképpen értelmezzük: A t_1, t_2, t_3, t_4, t_5 mondatok egy-egy zajos csatornán átmenő jelet képviselnek, pl. egy-egy bitet. A t_i mondat jelentése: az i. csatornán leolvasott érték 1. A csatornák állapota változatlan, de a zaj miatt két különböző érték olvasható le: az 10000 és az 10011. Ezeket az értékeket írja le a φ tudásbázis. Egy, a leolvasástól független vizsgálat szerint azonban minden bitnek ugyanaz az értéke, ezt az állítást képviseli a μ formula. Ez esetben elfogadható lesz a kijavítással kapott 00000 eredmény is. □

6. Elsőrendű tudásbázisok transzformációi

6.1 Alapfogalmak

Az elsőrendű L_1 nyelv a következő szimbólumokból áll:

Változók:	$X := \{x_i \mid i \in \mathbb{N}\}$
Konstansok:	$C := \{c_i \mid i \in \mathbb{N}\}$
Predikátumok:	$R := \{R_i \mid i \in \mathbb{N}\}$
Zárójelek:	$(\ ,\ )$
Logikai összekötők:	$\wedge;\ \vee;\ \neg$
Kvantor:	$\exists$
Egyenlőség:	$=,$

ahol $\mathbb{N}$ jelöli a természetes számokat.

Az R_i predikátum argumentumainak számát $arg(i)$ jelöli. A változókat és konstansokat együttesen *term*eknek nevezzük. Ha $arg(i) = n$ és $t_1, t_2, t_3, \ldots, t_n$ termek, akkor $R\,(t_1, t_2, t_3, \ldots, t_n)$ és $t_k = t_\ell$ atomok. Ha a $t_1, t_2, t_3, \ldots, t_n$ termek mindegyike konstans, akkor $R\,(t_1, t_2, t_3, \ldots, t_n)$ és $t_k = t_\ell$ alapatomok. A *jólformált formulák* a szokásos módon képezhetők a $\wedge;\vee;\neg$ bázison.

Az *ab adatbázis* relációk véges halmazából áll: $ab := \{r_1, r_2, r_3, \ldots, r_n\}$ ahol $r_i \subseteq C^{arg(i)}$ minden i-re. Az r_i reláció elemeit r_i sorainak nevezzük, és $\langle c_1, c_2, c_3, \ldots, c_{arg(i)}\rangle$-vel jelöljük.

Az ab adatbázis *sémája* $s(ab) := \{R_1, R_2, R_3, \ldots, R_n\}$. A μ (jólformált) formula sémája a μ-ben előforduló predikátumszimbólumok halmaza, jelölése: $s(\mu)$.

A μ formula *interpretációi* mindazok az ab adatbázisok, amelyekre $s(\mu) \subseteq s(ab)$. Az összes interpretáció halmaza $\mathfrak{I}$, az összes adatbázis halmaza AB.

A μ formula *modelljei* a μ azon ab interpretációi, amelyekre a következő tulajdonságok teljesülnek:

Ha μ a következő formulák valamelyike

1. $c_k = c_\ell$ akkor $k = \ell$.
2. $R_i\,(c_1, c_2, c_3, \ldots, c_n)$ akkor $\langle c_1, c_2, c_3, \ldots, c_n\rangle \in r_i$.
3. $\nu \wedge \varphi$, akkor ab modellje a ν-nek is és φ-nek is.
4. $\nu \vee \varphi$, akkor ab vagy ν-nek vagy φ-nek modellje.
5. $\neg\nu$, akkor ab nem modellje ν-nek.
6. $\exists x\nu$, akkor ab modellje a $\nu(x|c)$, $c \in C$, formulának, ahol $\nu(x|c)$ a c konstans helyettesítését jelenti a ν formulában x minden szabad előfordulásába.

Elsőrendű tudásbázison (a továbbiakban tudásbázison) azonos sémájú adatbázisok véges halmazát értjük. Például valamely φ elsőrendű formula modelljei (elsőrendű) tudásbázist alkotnak. A könyebb áttekinthetőség kedvéért ebben a fejezetben a φ formulát a modelljeivel meghatározott tudásbázissal reprezentáljuk, és tb_φ-vel jelöljük. A *tb tudásbázis sémája* egyenlő a benne szereplő adatbázisok sémájával, jelölése: $s(tb)$.

Minden tudásbázisnak megfeleltethető egy nulladrendű formula, ami úgy adható meg, hogy elemi ítéletnek tekintjük a $c_k = c_\ell$ és az $R_i\ (c_1, c_2, c_3, \ldots, c_n)$ [Rei78]. Ezért a 3.1, 3.4, 5.1 tételek közvetlenül alkalmazhatók.

6.2 Elsőrendű tudásbázis kijavítása

3.4 tételnek megfelelően elegendő egy globálisan megbízható, lojális függvényt megadni [BN94]. A 3.2 példában ismertetett távolságfüggvény alapján tudásbázisok különbözőségét és távolságát a következőképpen értelmezhetjük. Az azonos sémájú r_i, r_j relációk különbözőségét

$$\text{kül}(r_i, r_j) := |r_i \oplus r_j|$$

adja meg. Az azonos ab_m, ab_n sémájú adatbázisok különbözőségét

$$\text{kül}(ab_m, ab_n) := \sum \text{kül}(r_i^m, r_i^n), \text{ ahol } r_i^m \in ab_m,\ r_i^n \in ab_n$$

összefüggéssel definiáljuk. A tb tudásbázis és az ab adatbázisok távolsága:

$$\text{táv}(tb_\varphi, ab) := \text{Min}\left\{ \underset{ab_k \in tb_\varphi}{\text{kül}}(ab_k, , ab) \right\}.$$

Így $ab_m \leq_\varphi ab_n$ akkor és csak akkor, ha $\text{táv}(tb_\varphi, ab_m) \leq \text{táv}(tb_\varphi, ab_n)$ teljesül. Legyen g a tudásbázisokon értelmezett függvény, amely minden φ tudásbázishoz a fenti $\leq_\varphi$ előrendezést rendeli hozzá. A $\leq_\varphi$ előrendezés totális, és könnyen látható, hogy a g függvényre a lojalitás és a globális megbízhatóság többi tulajdonságai is teljesülnek. Így a

$$k : TB \times TB \to TB,\ k(tb_\varphi, tb_\mu) := \text{Min}\{tb_\mu, \leq_\varphi\}$$

összefüggéssel megadott operátor kielégíti a [K1]-[K7] axiómákat. Ez esetben is felvetődik az előrendezés „jóságának" kérdése. Ugyanis ha pl. a $\text{táv}(tb, ab)$ távolságot az egyes különbözőségi értékek maximumaként definiáljuk, akkor az ennek megfelelő függvény is globálisan megbízható és lojális, vagyis szintén kielégíti a [K1]-[K7] axiómarendszert.

Ha pedig a relációk különbözőségét pl. az alábbi módokon adjuk meg, akkor jobb jellemzőt kapunk, hiszen ily módon nemcsak a relációkban különböző sorok számát, hanem azoknak a reláció méretéhez viszonyított arányát is figyelembe vesszük:

1.
$$\text{kül}(r_i, r_j) := \begin{cases} \dfrac{|r_i \setminus r_j|}{|r_i|} + \dfrac{|r_j \setminus r_i|}{|r_j|} & \text{ha } |r_i| \neq 0,\ |r_j| \neq 0 \\ \dfrac{|r_i \setminus r_j|}{|r_i|} & \text{ha } |r_i| \neq 0,\ |r_j| = 0 \\ \dfrac{|r_j \setminus r_i|}{|r_i|} & \text{ha } |r_i| = 0,\ |r_j| \neq 0 \\ 0 & \text{ha } |r_i| = 0,\ |r_j| = 0. \end{cases}$$

2. $$\text{kül}(r_i, r_j) := |r_i \oplus r_j| \;/\; |r_i \cup r_j|$$

6.3 Elsőrendű tudásbázis modell-illesztése

Hasonlóan az előző fejezetben követett módszerhez, a 4.1 tétel alkalmazásához elegendő egy lojális függvény megadása [BN94].

Az előző fejezetben megadott elsőrendű kijavító operátorok lojálisak is, ezért egyben modell-illesztő operátorok. A 4. fejezetben ismertetett 4.2 példához hasonlóan ezen operátorok alábbi módosítása is modell-illesztő operátor:

$$\text{táv}(tb_\varphi, ab) := \text{Max}\left\{ \underset{ab_k \in tb_\varphi}{\text{kül}} (ab_k, ab) \right\}.$$

6.4 Elsőrendű tudásbázis felfrissítése

Az 5. fejezetben ismertetett nulladrendű felfrissítő operátort az 5.1 tétel felhasználásával GRAHNE, MENDELZON és RÉVÉSZ elsőrendű nyelvre általánosította. [GMR92]-ben az alábbi példa szerepel elsőrendű tudásbázisok felfrissítésére:

Az ab_m adatbázis közelebb van az ab adatbázishoz mint az ab_n adatbázis, ha

1. $s(ab_m) = s(ab_n)$ és $s(ab) \subseteq s(d_m)$
2. $ab_m \leq_{ab} ab_n$ ha minden $r_i^m \in ab_m, r_i^n \in ab_n, r_i \in ab$
   a.) $r_i^m \oplus r_i \subseteq r_i^n \oplus r_i$ minden olyan relációra, amely az ab_m, ab_n, ab adatbázisok mindegyikében előfordul
   b.) $r_i^m \oplus \emptyset \subseteq r_i^n \oplus \emptyset$ a többi relációra.

Könnyen látható, hogy a $\leq_{ab}$ parciális előrendezés AB-n. Legyen I az adatbázisokon értelmezett függvény, amely minden ab adatbázishoz a fentieknek megfelelő $\leq_{ab}$ előrendezést rendeli hozzá. Az I függvényre teljesülnek a lokális megbízhatóság tulajdonságai. Így az

$$f : TB \times TB \to TB, \quad f(\varphi, \mu) := \bigcup_{ab \in tb} \text{Min}\{tb_\mu, \leq_\varphi\}$$

összefüggéssel megadott f operátor felfrissítő operátor az 5.1 tétel szerint.

7. Súlyozott tudásbázisok

A világról kapott ismeretek, információk nem biztos, hogy egyformán fontosak, lehet, hogy egyik állítás bizonyos szempontból jobban tükrözi a leírandó világot, mint a másik. Pl. ha egy bizonyos kérdéskörről tartott közvéleménykutatásnál a vélemények megoszlanak: a megkérdezettek a%-ának véleményét az A állítás, b%-ának a B, stb. tükrözi, akkor az ennek alapján létrehozott tudásbázisban célszerű tárolni az erre vonatkozó ismereteket. [KM91]-ben ehhez hasonló példa található, minden egyes ítéletváltozóhoz két súly tartozik. Ezektől lényegesen különbözik az

[R93]-ban bevezetett súlyozott tudásbázis fogalma, ahol minden egyes interpretációhoz tartozik egy súly. A továbbiakban a súlyozott adatbázis ezen értelmezésének olyan módosítását adjuk meg, amely alkalmas negáció kifejezésére is.

7.1 Alapfogalmak

A továbbiakban, ha másként nem definiáljuk, a 2. fejezetben ismertetett jelöléseket használjuk.

7.1.1 Definíció. Nulladrendű súlyozott tudásbázisnak nevezzük a

$$\underline{\varphi} : \mathfrak{I} \to [0,1] \text{ függvényt.} \qquad \square$$

A továbbiakban súlyozott tudásbázison mindig a fent definiált nulladrendű súlyozott tudásbázist értjük. A súlyozott tudásbázisok halmazát $\underline{TB}$ jelöli.

7.1.2 Definíció. Súlyozott interpretációnak nevezzük az $(I, \alpha) \in \mathfrak{I} \times [0,1]$ rendezett párokat. A $\underline{\varphi}$ súlyozott tudásbázis modelljei mindazok a súlyozott interpretációk amelyekre $\underline{\varphi}(I) \geq \alpha > 0$:

$$\mathrm{Mod}(\underline{\varphi}) := \{(I,\alpha) \mid I \in \mathfrak{I},\ \underline{\varphi}(I) \geq \alpha > 0\}. \qquad \square$$

A fenti definícióból következik, hogy a $\underline{\varphi}$ súlyozott tudásbázis akkor és csak akkor *kielégíthetetlen*, ha $\underline{\varphi}(I) = 0$ minden $I \in \mathfrak{I}$-re.

C-Mod($\underline{\varphi}$)-vel (**C**lassical Modell) jelöljük azon interpretációk halmazát, amelyekre $\underline{\varphi}(I) > 0$.

A $\underline{\varphi}$ súlyozott tudásbázis *implikálja* a $\underline{\mu}$ súlyozott tudásbázist, ha minden $I \in \mathfrak{I}$ esetén $\underline{\varphi}(I) \leq \underline{\mu}(I)$, jelölése: $\underline{\varphi} \to \underline{\mu}$. Az implikáció értelmezése alapján amennyiben az ekvivalenciát a nulladrendben szokásos módon definiáljuk, akkor a $\underline{\varphi}$ súlyozott tudásbázis akkor és csak akkor *ekvivalens* a $\underline{\mu}$ súlyozott tudásbázissal, ha $\underline{\varphi}(I) = \underline{\mu}(I)$ minden $I \in \mathfrak{I}$ esetén (vagyis ha $\underline{\varphi} \to \underline{\mu}$ és $\underline{\mu} \to \underline{\varphi}$ egyidejűleg fennáll).

Súlyozott tudásbázisok között a következő műveleteket értelmezzük:

7.1.3 Definíció.

$$\underline{\varphi} \vee \underline{\mu}(I) = \mathrm{Max}\{\underline{\varphi}(I), \underline{\mu}(I)\}$$
$$\underline{\varphi} \wedge \underline{\mu}(I) = \mathrm{Min}\{\underline{\varphi}(I), \underline{\mu}(I)\}$$
$$\neg\underline{\varphi} = 1 - \underline{\varphi}(I). \qquad \square$$

A továbbiakban rátérünk a súlyozott tudásbázisokon értelmezhető transzformációkra.

7.2. Súlyozott tudásbázisok kijavítása

A súlyozott tudásbázisok esetében megadott implikáció fogalmából következik, hogy a [K1]-[K6], klasszikus kijavító operátorra megadott axiómák közül a 4.-re nincsen szükség. Így súlyozott esetben a $\underline{\circ} : \underline{TB} \times \underline{TB} \to \underline{TB}$ operátort *kijavító* operátornak nevezzük, ha az alábbi axiómák teljesülnek:

[SK1] $\underline{\varphi} \underline{\circ} \underline{\mu}$ implikálja $\underline{\mu}$-t.

[SK2] Ha $\underline{\varphi} \wedge \underline{\mu}$ kielégíthető, akkor $\underline{\varphi} \underline{\circ} \underline{\mu} \longleftrightarrow \underline{\varphi} \wedge \underline{\mu}$.

[SK3] Ha $\underline{\mu}$ kielégíthető, akkor $\underline{\varphi} \underline{\circ} \underline{\mu}$ is kielégíthető.

[SK4] $(\underline{\varphi} \underline{\circ} \underline{\mu}) \wedge \underline{\nu}$ implikálja $\underline{\varphi} \underline{\circ} (\underline{\mu} \wedge \underline{\nu})$-t.

[SK5] Ha $(\underline{\varphi} \underline{\circ} \underline{\mu}) \wedge \underline{\nu}$ kielégíthető, akkor $\underline{\varphi} \underline{\circ} (\underline{\mu} \wedge \underline{\nu})$ implikálja $(\underline{\varphi} \underline{\circ} \underline{\mu}) \wedge \underline{\nu}$-t.

Súlyozott tudásbázisok transzformációinál a súlyozott interpretációk között értelmezünk előrendezést, vagyis az $\mathfrak{I} \times [0,1]$ rendezett párok halmazán. Legyen ezen előrendezések halmaza $\underline{ER}$.

7.2.1 Definíció. Az $\underline{f} : \underline{TB} \times \underline{TB} \to \underline{ER}$ függvény *globálisan megbízható*, ha minden $\underline{\varphi}$ tudásbázishoz a következő tulajdonságokkal rendelkező $\leq_{\underline{\varphi}}$ előrendezést rendeli:

a.) Az előrendezés az első elemek szerint totális előrendezés.

b.) Ha $I \in \text{C-Mod}(\underline{\varphi})$ és $J \notin \text{C-Mod}(\underline{\varphi})$, akkor $(I,\alpha) <_{\underline{\varphi}} (J,\beta)$.

c.) Ha $(I,\alpha),(J,\beta) \in \text{Mod}(\underline{\varphi})$, akkor $(I,\alpha) \leq_{\underline{\varphi}} (J,\beta)$ és $(J,\beta) \leq_{\underline{\varphi}} (I,\alpha)$.

d.) Minden $\underline{\varphi}$ súlyozott tudásbázishoz és I interpretációhoz megadható egy olyan $\underline{\varphi}$-től függő $\alpha_{\underline{\varphi}}(I) \in]0,1]$ konstans, amelyre $(I, \text{Min}\{\alpha_{\underline{\varphi}}(I), \beta\}) \leq_{\underline{\varphi}} (I,\beta)$ és $\alpha_{\underline{\varphi}}(I) = \underline{\varphi}(I)$, ha $I \in \text{Mod}(\underline{\varphi})$. □

Megjegyzés: A c.) tulajdonság azt jelenti, hogy $I, J \in \text{C-Mod}(\underline{\varphi})$ esetén $I =_{\text{form(C-Mod}(\underline{\varphi}))} J$.

Súlyozott tudásbázisok körében is érvényes a 3.4 tételhez hasonló, 7.2.2 tétel.

7.2.2 TÉTEL. *Az* $\underline{\circ} : \underline{TB} \times \underline{TB} \to \underline{TB}$ *operátor akkor és csak akkor elégíti ki az* [SK1]-[SK5] *axiómákat, ha van olyan* $\underline{f}$ *globálisan megbízható függvény, hogy*

$$\text{Mod}(\underline{\varphi} \underline{\circ} \underline{\mu}) = \text{Min}\{\text{Mod}(\underline{\mu}), \leq_{\underline{\varphi}}\},$$

ahol $\leq_{\underline{\varphi}}$ *az* $\underline{f}$ *által a* $\underline{\varphi}$ *tudásbázishoz rendelt előrendezés.*

Bizonyítás.

I. Tegyük fel, hogy a $\underline{\circ}$ operátor kielégíti az [SK1]-[SK5] axiómákat. Az $\underline{f}$ függvény a következő $\leq_{\underline{\varphi}}$ relációt rendelje a $\underline{\varphi}$ tudásbázishoz:

Amennyiben I és J különbözők, legyen $(I,\alpha) \leq_{\underline{\varphi}} (J,\beta)$ akkor és csak akkor, ha $I \in \text{C-Mod}(\underline{\varphi} \circ ((I,1) \vee (J,1)))$, továbbá tetszőleges $I \in \mathfrak{I}$-re $(I, \text{Min}\{\alpha_{\underline{\varphi}}(I), \beta\}) \leq_{\underline{\varphi}} (I,\beta)$, ahol $\alpha_{\underline{\varphi}}(I) = (\underline{\varphi} \circ (I,1))(I)$.

Azt kell bizonyítani, hogy

1. az $\underline{f}$ függvény globálisan megbízható,
2. $\text{Mod}(\underline{\varphi} \circ \underline{\mu}) = \text{Min}\{\text{Mod}(\underline{\mu}), \leq_{\underline{\varphi}}\}$.

1. Az a.)–c.) tulajdonságok bizonyítása a 3.4 tétel bizonyításának I/1/a.)–c.) pontjaival megegyezik.

A d.) tulajdonságból az $(I, \text{Min}\{\alpha_{\underline{\varphi}}(I), \beta\}) \leq_{\underline{\varphi}} (I,\beta)$ fennállása közvetlenül a definícióból adódik.

Az $\alpha_{\underline{\varphi}}(I) = \underline{\varphi}(I)$, ha $I \in \text{Mod}(\underline{\varphi})$ tulajdonság pedig az [SK2] axióma következménye. Ugyanis, mint azt a 2. pont bizonyításánál igazoljuk, $\underline{\varphi} \circ \underline{\mu}(I) = \text{Min}\{\alpha_{\underline{\varphi}}(I), \underline{\mu}(I)\}$. Mivel most $I \in \text{Mod}(\underline{\varphi})$, és $I \in \text{Mod}(\mu)$, $\underline{\varphi} \circ \underline{\mu}(I) = \underline{\varphi} \wedge \underline{\mu}(I) = \text{Min}\{\underline{\varphi}(I), \underline{\mu}(I)\}$, vagyis $\text{Min}\{\alpha_{\underline{\varphi}}(I), \underline{\mu}(I)\} = \text{Min}\{\underline{\varphi}(I), \underline{\mu}(I)\}$. Mivel $\underline{\mu}(I)$ tetszőleges, az egyenlőség csak úgy teljesülhet, ha $\alpha_{\underline{\varphi}}(I) = \underline{\varphi}(I)$.

A 2. bizonyításához egyrészt bizonyítani kell, hogy a $\text{Mod}(\underline{\varphi} \circ \underline{\mu})$-ben szereplő párok első elemei megegyeznek a $\text{Min}\{\text{Mod}(\mu), \leq_{\underline{\varphi}}\}$ halmazban szereplő párok első elemeivel. Ez a bizonyítás a 3.4 tétel I/2 részével egyezik.

Továbbá azt kell még bizonyítani, hogy

$$\underline{\varphi} \circ \underline{\mu}(I) = \text{Min}\{\alpha_{\underline{\varphi}}(I), \underline{\mu}(I)\}.$$

Az [SK1] és [SK3] axiómák miatt $0 < \alpha_{\underline{\varphi}}(I) \leq 1$. Legyen a $\underline{\mu}$ kijavító tudásbázis egy súlyozott modellje $(I, \underline{\mu}(I))$.

Mivel ekkor $\underline{\mu}(I) > 0$, ezért $(\underline{\varphi} \circ (I,1)) \wedge \underline{\mu}$ kielégíthető, így az [SK4] és [SK5] axiómák miatt

$$((\underline{\varphi} \circ (I,1)) \wedge \underline{\mu})(I) = (\underline{\varphi} \circ ((I,1) \wedge \underline{\mu}))(I).$$

Tegyük fel először, hogy $\alpha_{\underline{\varphi}}(I) \geq \underline{\mu}(I)$, kapjuk hogy:

$$((\underline{\varphi} \circ (I,1)) \wedge \underline{\mu})(I) = \underline{\mu}(I) = \underline{\varphi} \circ ((I,1) \wedge \underline{\mu})(I) = \underline{\varphi} \circ \underline{\mu}(I).$$

Ha pedig $\underline{\mu}(I) > \alpha_{\underline{\varphi}}(I)$, akkor $((\underline{\varphi} \circ (I,1)) \wedge \underline{\mu})(I) = \alpha_{\underline{\varphi}}(I)$.
Másrészt $\underline{\varphi} \circ ((I,1) \wedge \underline{\mu})(I) = \underline{\varphi} \circ \underline{\mu}(I)$, tehát $\underline{\varphi} \circ \underline{\mu}(I) = \alpha_{\underline{\varphi}}(I)$.

Ezzel bebizonyítottuk, hogy $\underline{\varphi} \circ \underline{\mu}(I) = \text{Min}\{\alpha_{\underline{\varphi}}(I), \underline{\mu}(I)\}$, vagyis a $\circ$ operátor valóban a $\underline{\mu}$ tudásbázis $\leq_{\underline{\varphi}}$ előrendezés szerinti minimális elemeit választja ki.

II. Tegyük fel, hogy az $\underline{f}$ globálisan megbízható függvény által a $\underline{\varphi}$ tudásbázishoz rendelt előrendezés $\leq_{\underline{\varphi}}$ és a $\circ$ operátor pedig a $\text{Mod}(\underline{\varphi} \circ \underline{\mu}) = \text{Min}\{\text{Mod}(\underline{\mu}), \leq_{\underline{\varphi}}\}$ összefüggés által definiált. Azt kell bizonyítani, hogy teljesülnek az [SK1]-[SK5] axiómák.

Az [SK1] axióma teljesül, hiszen $\text{Mod}(\underline{\mu})$ bizonyos elemei adják az eredményt.

Az [SK2] axióma teljesülésének bizonyítása egyrészt megegyezik a 3.4 tétel bizonyításának II/[K2] részével. Másrészt a súlyokra vonatkozóan pedig a már

kiválasztott I interpretációkra az állítás a $\mathrm{Min}\{\alpha_{\underline{\varphi}}(I), \underline{\mu}(I)\} = \mathrm{Min}\{\underline{\varphi}(I), \underline{\mu}(I)\} = (\underline{\varphi} \wedge \underline{\mu})(I)$ azonosságból következik.

Az [SK3] axióma teljesülése a $\overset{\circ}{-}$ definíciójából közvetlenül adódik.

Az [SK4] axióma, ha $(\underline{\varphi} \overset{\circ}{-} \underline{\mu}) \wedge \underline{\nu}$ nem kielégíthető, akkor $((\underline{\varphi} \overset{\circ}{-} \underline{\mu}) \wedge \underline{\nu})(I) = 0$ minden I-re, ezért $((\underline{\varphi} \overset{\circ}{-} \underline{\mu}) \wedge \underline{\nu})(I) \leq \underline{\varphi} \overset{\circ}{-} (\underline{\mu} \wedge \underline{\nu})(I)$ egyenlőtlenség mindig teljesül.

Ha $(\underline{\varphi} \overset{\circ}{-} \underline{\mu}) \wedge \underline{\nu}$ kielégíthető, akkor az [SK4] és [SK5] axiómák teljesülése egyrészt azt jelenti, hogy $\mathrm{Mod}((\underline{\varphi} \overset{\circ}{-} \underline{\mu}) \wedge \underline{\nu}) = \mathrm{Mod}(\underline{\varphi} \overset{\circ}{-} (\underline{\mu} \wedge \underline{\nu}))$, másrészt, hogy $((\underline{\varphi} \overset{\circ}{-} \underline{\mu}) \wedge \underline{\nu})(I) = \underline{\varphi} \overset{\circ}{-} (\underline{\mu} \wedge \underline{\nu})(I)$. Az első állítás bizonyítása a 3.4 tétel bizonyításának II/[K5]-[K6] bizonyításával azonos. A második állítás bizonyítása a megfelelő definíciók alapján $((\underline{\varphi} \overset{\circ}{-} \underline{\mu}) \wedge \underline{\nu})(I) = \mathrm{Min}\{\alpha_{\underline{\varphi}}(I), \underline{\mu}(I), \underline{\nu}(I)\} = \underline{\varphi} \overset{\circ}{-} (\underline{\mu} \wedge \underline{\nu})(I)$. □

7.3 Súlyozott tudásbázisok modell-illesztése

A súlyozatlan tudásbázisoknál a 4. fejezetben ismertetett modell-illesztő operátorokhoz hasonlóan a $\underline{\nabla} : \underline{TB} \times \underline{TB} \to \underline{TB}$ operátort *modell-illesztő* operátornak nevezzük, ha kielégíti az alábbi [S1]-[S6] axiómákat:

[SM1] $\underline{\varphi} \underline{\nabla} \underline{\mu}$ implikálja $\underline{\mu}$-t.

[SM2] Ha $\underline{\varphi}$ kielégíthetetlen, akkor $\underline{\varphi} \underline{\nabla} \underline{\mu}$ is az.

[SM3] Ha $\underline{\varphi}$ és $\underline{\mu}$ mindegyike kielégíthető, akkor $\underline{\varphi} \underline{\nabla} \underline{\mu}$ is az.

[SM4] $(\underline{\varphi} \underline{\nabla} \underline{\mu}) \wedge \underline{\nu}$ implikálja $\underline{\varphi} \underline{\nabla} (\underline{\mu} \wedge \underline{\nu})$-t.

[SM5] Ha $(\underline{\varphi} \underline{\nabla} \underline{\mu}) \wedge \underline{\nu}$ kielégíthető, akkor $\underline{\varphi} \underline{\nabla} (\underline{\mu} \wedge \underline{\nu})$ implikálja $(\underline{\varphi} \underline{\nabla} \underline{\mu}) \wedge \underline{\nu}$-t.

[SM6] $(\underline{\varphi}_1 \underline{\nabla} \underline{\mu}) \wedge (\underline{\varphi}_2 \underline{\nabla} \underline{\mu})$ implikálja $(\underline{\varphi}_1 \vee \underline{\varphi}_2) \underline{\nabla} \underline{\mu}$-t.

A lojalitás tulajdonságának súlyozott tudásbázisokra való megfogalmazásával a 7.2.2 tételhez hasonló tétel itt is fennáll.

7.3.1 Definíció. Az $\underline{sl} : \underline{TB} \to \underline{ER}$ függvény *lojális*, ha teljesül, hogy minden $\underline{\varphi}_1, \underline{\varphi}_2 \in D_{\underline{sl}}$-re, amennyiben

a.) Minden $\underline{\varphi}$ súlyozott tudásbázishoz és I interpretációhoz megadható egy olyan, $\underline{\varphi}$-től függő $\alpha_{\underline{\varphi}}(I) \in]0, 1]$ konstans, amelyre $(I, \mathrm{Min}\{\alpha_{\underline{\varphi}}(I), \beta\}) \leq_{\underline{\varphi}} (I, \beta)$.

b.) ha $\underline{sl}(\underline{\varphi}_1) = \leq_{\underline{\varphi}_1}$, $\underline{sl}(\underline{\varphi}_2) = \leq_{\varphi_2}$ és $(I, \alpha) \leq_{\underline{\varphi}_1} (J, \beta), (I, \alpha) \leq_{\underline{\varphi}_2} (J, \beta)$ akkor $(I, \alpha) \leq_{\underline{\varphi}_1 \vee \underline{\varphi}_2} (J, \beta)$, ahol $\underline{sl}(\underline{\varphi}_1 \vee \underline{\varphi}_2) = \leq_{\underline{\varphi}_1 \vee \underline{\varphi}_2}$. □

A következő tétel biztosítja, hogy $\alpha_{\underline{\varphi}}(I)$ speciális választásával és lojális függvény segítségével megadható az [SM1]-[SM6] axiómákat kielégítő $\underline{\nabla}$ operátor.

7.3.2 Tétel. *Ha $\underline{sl}$ olyan lojális függvény, amely a $\underline{\varphi}$ súlyozott tudásbázishoz a $\leq_{\underline{\varphi}}$ előrendezést rendeli, akkor a*

$$\underline{\nabla} : \underline{TB} \times \underline{TB} \to \underline{TB}, \quad \mathrm{Mod}(\underline{\varphi} \underline{\nabla} \underline{\mu}) := \mathrm{Min}\{\mathrm{Mod}\{\underline{\mu}, \leq_{\underline{\varphi}}\}\}$$

összefüggéssel definiált operátor kielégíti az [SM1]–[SM6] *axiómákat az* $\alpha_{\underline{\varphi}}(I) = 1$ *választás mellett.*

Bizonyítás. Az $\alpha_{\underline{\varphi}}(I) = 1$ választás miatt $\mathrm{Min}\{\alpha_{\underline{\varphi}}(I), \beta\} = \beta$, ezért a kiválasztott párokban a súly mindig $\underline{\mu}(I)$.

Az [SM1]-[SM6] axiómák teljesülésének igazolása az eddigiekhez hasonlóan két részből áll, egyrészt be kell látni, hogy a párok első elemei a megfelelő súlyozatlan modellekben szerepelnek, másrészt, hogy a súlyok választása is megfelelő. Az első rész bizonyítása a 4.1 tétel bizonyításának a II. részében megtalálható. A súlyokra vonatkozóan a következők teljesülnek:

Az [SM1], [SM3] axióma fennáll, hiszen a választott modell súlya $\underline{\mu}(I)$.

Az [SM2] axióma igazolása a súlyozatlan esettel megegyezik.

Az [SM4] axióma teljesül, hiszen

$$((\underline{\varphi} \,\underline{\nabla}\, \underline{\mu}) \wedge \underline{\nu})(I) = \mathrm{Min}\{\underline{\mu}(I), \underline{\nu}(I)\} = \underline{\varphi} \,\underline{\nabla}\, (\underline{\mu} \wedge \underline{\nu})(I).$$

Ehhez hasonlóan, ha $(\underline{\varphi} \,\underline{\nabla}\, \underline{\mu}) \wedge \underline{\nu}$ kielégíthető, akkor

$$\underline{\varphi} \,\underline{\nabla}\, (\underline{\mu} \wedge \underline{\nu})(I) = ((\underline{\varphi} \,\underline{\nabla}\, \underline{\mu}) \wedge \underline{\nu})(I).$$

Az [SM6] axiómában szintén igaz az implikáció, a súlyok egyenlősége miatt:

$$((\underline{\varphi}_1 \,\underline{\nabla}\, \underline{\mu}) \wedge (\underline{\varphi}_2 \,\underline{\nabla}\, \underline{\mu}))(I) = \underline{\mu}(I) = ((\underline{\varphi}_1 \vee \underline{\varphi}_2) \,\underline{\nabla}\, \underline{\mu})(I). \qquad \square$$

A szimmetrikus modell-illesztés, amelynek célja a mindkét tudásbázishoz legjobban illeszkedő súlyozott tudásbázis megadása, a következőképpen értelmezhető:

7.3.3 Definíció. A $\underline{\Delta}$ operátor a $\underline{\varphi}$ és a $\underline{\mu}$ súlyozott tudásbázisok szimmetrikus modellillesztését adja, ha

$$\underline{\varphi} \,\underline{\Delta}\, \underline{\mu} := (\underline{\varphi} \vee \underline{\mu}) \,\underline{\nabla}\, \underline{M}$$

ahol $\underline{M}$ jelenti azt a súlyozott tudásbázist, amely minden $I \in \mathfrak{I}$-hez az 1 súlyt rendeli hozzá. $\square$

7.3.4 Példa. A 7.2.1 tétel szerint elegendő egy súlyozottan lojális függvény megadása. Valamely interpretáció távolsága a $\underline{\varphi}$ tudásbázistól legyen a következő:

$$\mathrm{s\_táv}(\underline{\varphi}, (I, \alpha)) := \sum_{\substack{\underline{\varphi}(j) > 0 \\ (J, \underline{\varphi}(J)) \in \mathrm{Mod}(\underline{\varphi})}} \mathrm{kül}(I, J) * \underline{\varphi}(J).$$

Az interpretációk közti előrendezés: $(I, \alpha) \leq_{\underline{\varphi}} (J, \beta)$ akkor és csak akkor, ha $\mathrm{s\_táv}(\underline{\varphi}, (I, \alpha)) < \mathrm{s\_táv}(\underline{\varphi}, (J, \beta))$.

Adott $\underline{\varphi}$ súlyozott tudásbázishoz az *sl* függvény rendelje a fenti módon megadott $\leq_{\underline{\varphi}}$ előrendezést. Az így definiált *sl* függvény súlyozottan lojális.

Ezért a $\mathrm{Mod}(\varphi \underline{\nabla} \underline{\mu}) := \mathrm{Mod}\{\mathrm{Min}\{\underline{\mu}, \leq_{\underline{\varphi}}\}\}$ összefüggéssel megadott $\underline{\nabla}$ operátor súlyozott modell-illesztő operátor. □

A 4.2 b.) példa súlyozott esetre való átfogalmazása pl. a következő: A tanár véleménye legyen változatlan, így $\underline{\mu}(\{D\}) = \underline{\mu}(\{S, D\}) = 1$. A hallgatók azonban nem egyenlő arányban, hanem pl. következő eloszlásban szeretnének tanulni: 10 hallgató csak SQL-t, 20 csak Datalogot, és 5 SQL-t, Datalogot és Query-by-Example-t. A hallgatók kívánsága szerint tehát $\underline{\varphi}(\{S\}) = 10/35$, $\underline{\varphi}(\{D\}) = 20/35$, és $\underline{\varphi}(\{S, D, Q\}) = 5/35$. A megfelelő távolságok kiszámítása után $\{D\}$ lesz a minimális modell. Mivel $\underline{\mu}(\{D\}) = 1$ volt, ez a súly meg is marad, az eredmény a $(\{D\}, 1)$ modell.

8. Problémák

A 3.1 fejezetben szereplő [K1]-[K6] axiómák jól alkalmazhatók, ha az eredeti φ tudásbázis kielégíthető. Ha azonban φ kielégíthetetlen, nincsen gyakorlati útmutatás arra vonatkozóan, mi is legyen a kijavítás eredménye. A 3.1 tétel nem terjeszthető ki erre az esetre, hiszen nem tudjuk definiálni, hogy mely μ modellek azok, amelyek a φ üres modellhalmazához legközelebb esnek. A legegyszerűbb megoldás ilyen esetben, hogy μ legyen az eredmény.

Így azonban minden eredeti információ elvész. Ezért célszerűnek tűnik az eredeti tudásbázis konzisztens részformuláinak megfelelő modellek kijavításával kapott tudásbázisok vizsgálata.

További vizsgálatot igényel a [K1]-[K7] bővített axiómarendszert kielégítő operátorcsalád, ugyanis a [K7] axióma hozzávétele még nem zárja ki az olyan globálisan megbízható és lojális függvényeket, amelyek azonban mégis intuitív szempontból elfogadhatatlan előrendezéseket rendelnek az egyes tudásbázisokhoz.

Érdekes kérdés vetődik fel, ha a [K7] axióma megfordítását szeretnénk az axiómarendszerhez kapcsolni, pontosabban, ha az alábbi tulajdonság teljesülését is megköveteljük:

[K8] Ha $(\varphi_1 \bullet \mu) \wedge (\varphi_2 \bullet \mu)$ kielégíthető, akkor $(\varphi_1 \vee \varphi_2) \bullet \mu$ implikálja $(\varphi_1 \bullet \mu) \wedge (\varphi_2 \bullet \mu)$-t.

[R93]-ban egy másik transzformáció bevezetésekor szerepel mindkét axióma. A cikkből kiderül, hogy valamely operátor akkor és csak akkor elégíti ki a [K7], [K8] axiómákat, ha van olyan f *szigorúan lojális függvény*, amelyre $\mathrm{Mod}(\varphi \bullet \mu) = \mathrm{Min}\{\mathrm{Mod}(\mu), f(\varphi)\}$.

8.1. Definíció. Az f függvény *szigorúan lojális*, ha a következők teljesülnek:

1. Ha $\varphi_1 \longleftrightarrow \varphi_2$, akkor $f(\varphi_1) = f(\varphi_2)$.
2. Ha $I <_{\varphi_1} J$ és $I \leq_{\varphi_2} J$, akkor $I <_{\varphi_1 \vee \varphi_2} J$.
3. Ha $I \leq_{\varphi_1} J$ és $I \leq_{\varphi_2} J$, akkor $I \leq_{\varphi_1 \vee \varphi_2} J$. □

8.1 LEMMA. *Nem létezik globálisan megbízható és szigorúan lojális függvény.*

Bizonyítás. Legyen $\mathrm{Mod}(\varphi_1) = I_1, I_2, \ldots, I_k$, J és $\mathrm{Mod}(\varphi_2) = I_1, I_2, \ldots, I_k$. A globális megbízhatóság miatt $I_{k_\ell} =_{\varphi_1} J$ és $I_\ell <_{\varphi_2} J$ minden $1 \leq \ell \leq k$. Ha a függvény egyben szigorúan lojális is lenne, akkor $I_\ell <_{\varphi_1 \vee \varphi_2} J$ teljesülne, ami ellentmondás, hiszen

$$J \in \mathrm{Mod}_{\{\varphi_1 \vee \varphi_2\}}. \qquad \square$$

Ha f minden tudásbázishoz ugyanazt az előrendezést rendeli, akkor nyilvánvalóan szigorúan lojális. Más, a gyakorlatban alkalmazható lojális függvény megadásáról nincs tudomásunk. Az a kérdés tehát, hogyan konstruálható szigorúan lojális függvény.

A szigorú lojalitás problémája ugyancsak érvényes, ha súlyozott tudásbázisokra fogalmazzuk meg — az 1. tulajdonság mellőzésével — a szigorú lojalitást.

(A felsorolt problémák mind a nulladrendű, mind az elsőrendű operátorokkal kapcsolatban fennállnak.)

További vizsgálatot igényel a 7.3 fejezetben ismertetett modell-illesztő operátorok konkrét megadása, pontosabban az $\alpha_{\underline{\varphi}}(I)$ számok meghatározása oly módon, hogy a megadott axiómarendszer teljesüljön.

Az e cikkben ismertetett operátorokkal kapcsolatban felvetődhet a komplexitás kérdése. T. EITLER és G. GOTTLOB [EG92]-ben részletesen elemzi a nulladrendű kijavító és felfrissítő operátorokat e szempontból. Ezek mintájára súlyozott esetre, illetve a cikkben megadott új, modell-illesztő operátorra hasonló elemzések készítendők.

IRODALOM

[AGM85] ALCHOURRÓN, C. E., GARDENFORS, P., MAKINSON, D., „On the logic of theory change: partial meet contraction and revision functions", *Journal of Symbolic Logic* **50** (1985), 510–530.

[BN94] BENCZÚR, A., NOVÁK, Á. B., „On Knowledgebase Change Operators", *Proceedings of International Scientific Conference, Section Artificial Intelligence* (Herlany, Slovakia, 1994).

[D88a] DALAL, M., „Updates in Propositional Databases", *Technical Report DCS-TR-222* (Department of Computer Science, Rutgers University, New Brunswick, NJ, 1988).

[D88b] DALAL, M., „Investigations into a Theory of Knowledge Base Revision", *Proceedings of AAAI-88* (St. Paul, MN, 1988).

[EG92] EITLER, T., GOTTLOB, G., „On the complexity of propositional knowledgebase revision, updates and counterfactuals", *Artificial Intelligence* **57**, 227–270.

[GMR92] GRAHNE, A. O., MENDELZON, P., RÉVÉSZ, Z., „Knowledge Base Transformations", *Proceedings of the Eleventh ACM-SIGACT-SIGART Symposium on Principles of Database Systems* (1992).

[KM91a] KATSUNO, H., MENDELZON, A. O., „On the Difference between Updating a Knowledge Base and Revising it", *Proceedings of the Second International Conferences on Principles of Knowledge Representation and Reasoning* (1991).

[KM91b] KATSUNO, H., MENDELZON, A. O., „Propositional Knowledge Base Revision and Minimal Change", *Artificial Intelligence* **52** (1991).

[KW85] KELLER, A. M., WINSLETT, M., „On the use of an extended relational model to handle changing incomplete information", *IEEE Trans. on Software Engineering, SE-11:7* (1985), 620–633.

[R93] RÉVÉSZ, P. Z., *On the Semantics of Theory Change: Arbitration between Old and New Information*, ACM - PODS5/93/Washington, D.C.

[R94] RÉVÉSZ, P.Z., *On the Semantics of Arbitration. Report Series: UNLCSE-94-010 University of Nebraska-Lincoln* (1994).

[Rei78] REITER, R., „Towards a Logical Reconstruction of Relational Database Theory.", *Conceptual Modelling,* (Brodie, Mylopoulos, Schmidt, eds.) (New York, 1984).

[S88] SATOH, K., „Nonmonotonic reasoning by minimal belief revision", *Proceedings International Conference on Fifth Generation Computer Systems* (1988), 455–462.

[We86] WEBER, A., „Updating propositional formulas", *Proceedings of First Conference on Expert Database Systems* (1986), 487–500.

[Wi88] WINSLETT, M., „Reasoning about action using a possible model approach", *Proceedings of the Seventh National Conference on Artificial Intelligence* (1988), 89–93.

(Beérkezett: 1995. június 16.)

BENCZÚR ANDRÁS
EÖTVÖS LORÁND TUDOMÁNYEGYETEM
ÁLTALÁNOS SZÁMÍTÁSTUDOMÁNYI TANSZÉK
BUDAPEST, MÚZEUM KRT. 6.–8.
H–1088

B. NOVÁK ÁGNES
BÁNKI DONÁT MÜSZAKI FÖISKOLA
INFORMATIKA TANSZÉK
BUDAPEST, NÉPSZÍNHÁZ U. 8.
H–1081

RÉVÉSZ Z. PÉTER
COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF NEBRASKA-LINCOLN
115 FERGUSON HALL
P.O. BOX 880 115
LINCOLN, NE 685888–0115
U.S.A.

PROPOSITIONAL AND WEIGHTED KNOWLEDGE BASE TRANSFORMATIONS

A. BENCZÚR, Á. B. NOVÁK AND P.Z. RÉVÉSZ

In this paper — after a review of some knowledgebase transformations, namely the revision and update well-known in propositional logic — as well as the light generalization of revision for first-order case there is an extended set of axioms to avoid a certain problem in connection with the revision. Furthermore we give an extension of the propositional knowledgebase to weighted knowledgebase. Finally we deal with the weighted knowledgebase transformations.

Alkalmazott Matematikai Lapok **17** (1993) 373–396

LOKÁLIS PÁRHUZAMOS ALGORITMUS BINÁRIS KÉPEK ZAJSZŰRÉSÉRE

PALÁGYI KÁLMÁN

Szeged

Egy olyan módszert ismertetünk bináris képek zajszűrésére (adott méretnél kisebb, zajnak minősített komponensek eltávolítására), mely a megmaradó képkomponenseket nem változtatja meg. Az összetett eljárás valamennyi részművelete lokális és párhuzamos. A párhuzamosság azt jelenti, hogy a feldolgozás bármely fázisában az éppen meghatározandó bináris kép valamennyi pontja egyidőben számítható ki, a részműveletek lokális voltán pedig azt értjük, hogy valamely képpont új értéke csak az adott pont egy szűk, esetünkben legfeljebb 3 × 3-as környezetétől függ.

1. Bevezetés

Bináris képen — pontosabban: kétdimenziós bináris digitális raszterképen — egy $n \times m$ méretű $A = (a_{ij})$ mátrixot értünk, melynek elemei „0", „1" értékűek. A mátrix elemeit *képpont*oknak vagy a *pixel*eknek nevezzük. Az „1" pixeleket *objektumpont*oknak, a „0"-kat pedig *háttérpont*oknak nevezzük.

Az a_{ij} képpontnak *4-* illetve *8-szomszéd*ja az a_{kl} pixel, ha

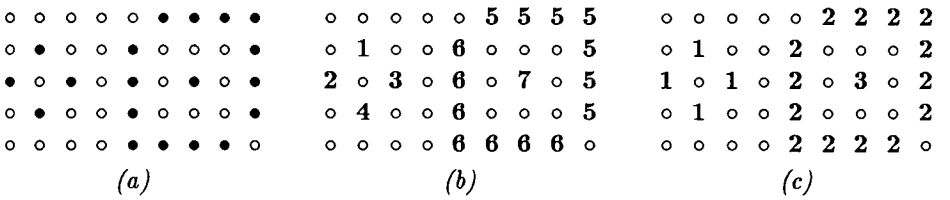
$$|k-i| + |l-j| = 1 \qquad \text{illetve}$$
$$\max(|k-i|, |l-j|) = 1.$$

A fenti módon definiált (irreflexív és szimmetrikus) 4- illetve 8-szomszédsági reláció (1. ábra) reflexív tranzitív lezárásával kapott ekvivalenciarelációt *4-* illetve *8-összefüggőségi reláció*nak nevezzük. A 4- illetve a 8-összefüggőségi reláció a bináris kép objektumpontjain létrehoz egy osztályozást. Ugyanazon ekvivalencia-osztályba eső objektumpontok alkotják a kép egy *4-* illetve *8-komponens*ét. A 4- illetve a 8-komponensekre példát mutató 2. ábrán és a további ábrákon az objektumpontokat a „•", a háttérpontokat pedig a „○" karakter jelöli.

	$p2$			$p4$ $p3$ $p2$	
$p3$	$p0$	$p1$		$p5$ $p0$ $p1$	
	$p4$			$p6$ $p7$ $p8$	
	(a)			*(b)*	

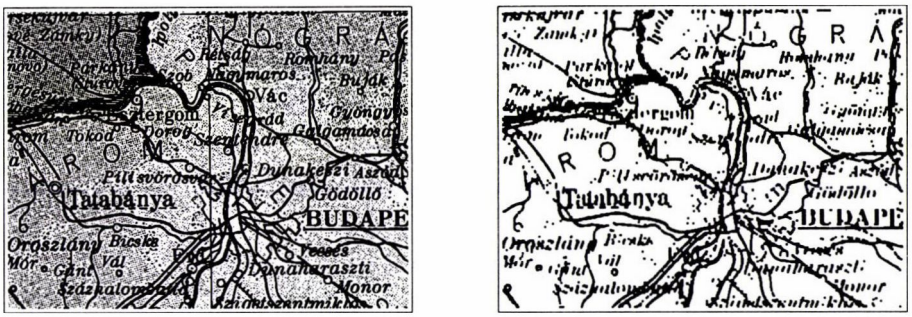
1. ábra. A „p0" pixel 4-szomszédsága (a) és 8-szomszédsága (b).

Bináris képek zajszűrésének (*smoothing, noise filtering*) és zajcsökkentésének (*noise reduction*) célja a „kisméretű", zajnak tekintett komponensek eltávolítása, a komponensek kontúrjának (háttérponttal szomszédos objektumpontjai halmazának) „egyenletesebbé" tétele.



2. ábra. Példa 4- és 8-komponensekre. (a) a példakép, (b) 4-komponensek, (c) 8-komponensek. Az ugyanabba az osztályba tartozó objektumpontokhoz ugyanazt a számot rendeltük.

A többszintű (*multi-level*) digitális képek zajszűrésére, zajcsökkentésére számos módszer ismert, például a Fourier-transzformáción alapuló szűrések, vagy a különféle lokális eljárások (környezeti átlagolások, mediánszűrés, ...) [1]. Bináris (*bi-level*) képekre a fenti módszerek nem bizonyultak célszerűnek. Helyettük leggyakrabban *morfológiai szűrést*, az *open* és a *close* műveleteket alkalmazzák [2], [3]. Az open művelet *erózióval* (*erosion*) eltünteti az adott méretnél kisebb komponenseket, de egyúttal beletöröl a zajnak nem minősülő komponensekbe is. Az eróziót követő *dilatáció* (*dilation*) a megmaradt komponensek „visszahízlalására" szolgál. Az open műveletet követő close dilatációval indul, mely a megmaradó komponenseket tovább hízlalja, feltölt adott méretnél kisebb üregeket, lyukakat. Az eljárás erózióval fejeződik be, célja a „túlhízlalalás" kompenzálása. A morfológiai szűrés tehát általában megváltoztatja a nem-törölt komponenseket (3. ábra). A részletek jobb megőrzése elérhető az eljárás módosításával, például KUOSMANEN és munkatársai [4] is finomítottak az eljáráson. A morfológiai szűrés mellett egyéb módszerek is léteznek, mint például RAY szekvenciális eljárása [5], mely különböző méretű (3 × 3-tól 5 × 5-ig terjedő) lokális környezetet vizsgálva dönt az objektumpontok megtartásáról vagy törléséről. Az eljárás hátránya, hogy a zaj-méret korlátozott és beletöröl a megmaradó komponensekbe is. Érdemes még megemlíteni ALI és PAVLIDIS kontúrkövetéses módszerét [6], mely „gyanús" összeéréseket és szakadásokat szüntet meg.



3. ábra

3. ábra. Példa morfológiai szűrésre. Az (a) — 480 × 640 méretű — térképrészletre az eljárás a (b) képet eredményezte. A nem-törölt komponensek is megváltoztak, számos helységnév olvashatatlanná vált.

Jelen cikkben egy új módszert ismertetünk bináris képek adott méretnél kisebb komponenseinek eltávolítására. Módszerünk két fő erénye, hogy a zaj-méret tetszőlegesen megválasztható, valamint az, hogy a nem-törölt komponensek változtatás nélkül megőrződnek. Az eljárás módosításával adott mérethatárok közé eső komponensek eltávolítása is megoldható, sőt törölhetünk adott méretnél nagyobb, vagy adott méret-intervallumon kívül eső komponenseket is.

Az eljárás összetett, vagyis különböző résztevékenységekből épül fel. A résztevékenységek (pl: a zsugorítás vagy a terjesztés) önmagukban is értelmes, hasznos műveletek.

A részműveletekre adott megoldások lokálisak és párhuzamosak. A párhuzamosság azt jelenti, hogy a feldolgozás adott lépése során meghatározandó bináris kép valamennyi pontja egyidőben számítható ki, vagyis ha rendelkeznénk egy olyan paralell számítógéppel, melyben az adott méretű bináris képünk minden pontjára jutna egy-egy processzor, akkor a teljes kép feldolgozásának időigénye egyetlen képpont új értékének kiszámítási idejével lenne egyenlő.

A részműveletek lokális voltán azt értjük, hogy valamely képpont új értéke csak az adott pont egy szűk, esetünkben legfeljebb 3 × 3-as környezetétől függ. A feladatot, vagyis a zajnak minősülő komponensek eltávolítását tehát lokális műveletek sorozatával oldjuk meg, a zajmérettől függetlenül 3 × 3-as környezetet figyelve.

A dolgozatban leírt módszer alkalmazása hasznos lehet például az úgynevezett digitális raszterképek formájában tárolt tervrajzok, térképek vektoros állománnyá konvertálásakor. A zajnak minősülő komponensek eltávolítása mellett az adott mérethatárok közé eső szimbólumok (pl: számok, térképészeti jelkulcsok) is leválogathatók, külön raszterképre helyezhetők, automatikus vektorizálásukat másképpen paraméterezhetjük, mint a kép maradékáét (pl: a vonalrajzét).

Az ismertetésre kerülő módszer gyenge pontja a komponens-méret önkényes meghatározása. Egy komponens mérete — szóhasználatunk szerint — akkor d, ha azt az alkalmazott zsugorító eljárás pontosan a d-edik lépésben változtatja *izolált*

ponttá. (Egy objektumpont izolált, ha valamennyi szomszédja háttérpont a figyelembe vett szomszédsági reláció mellett.) Módszerünket két változatban dolgoztuk ki: 4-összefüggő és 8-összefüggő komponensekre.

A dolgozat 2. pontjában az igényelt részműveleteket írjuk le, a belőlük felépített összetett eljárást pedig a 3. pontban ismertetjük. A részműveletek helyességére vonatkozó állításaink igazolását a Függelék tartalmazza.

2. Az alkalmazott részműveletek

Jelen pontban az összetett zajszűrő módszerünkhöz felhasznált részműveleteket ismertetjük.

*k-változós bináris képfeldolgozó művelet*en olyan transzformációt értünk, ami k darab bináris képből egy bináris képet képez. A tárgyalásra kerülő műveletek egyrészt *mérettartó*ak (vagyis $n \times m$ méretű bináris képekből ugyancsak $n \times m$-eseket képeznek), másrészt egy- vagy kétváltozósak, továbbá kétváltozósként csak pixel-szintű logikai műveletek fordulnak elő.

Valamely egyváltozós bináris képfeldolgozó $\mathcal{F}$ művelet *p-fázisú* ($p \geq 1$), ha $\mathcal{F}$-et az $\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_p$ egyváltozós műveletek kompoziciójaként adjuk meg, vagyis tetszőleges A bináris képre:

$$\mathcal{F}(A) = \mathcal{F}_p(\ldots(\mathcal{F}_2(\mathcal{F}_1(A)))\ldots).$$

A tárgyalásra kerülő műveleteket a következő csoportokba soroltuk:

- *egyváltozós additív*, ha a művelet az objektumpontokat nem változtatja meg, viszont háttérpontokból objektumpontokat képez, ha 3×3-as környezetük bizonyos feltételeknek eleget tesz;
- *egyváltozós szubtraktív*, ha a művelet csak objektumpontokat változtat meg;
- *kétváltozós logikai* műveletek.

Az egyváltozós bináris képfeldolgozó műveletek fontos jellemzője, hogy megőrzik-e az összefüggőségi viszonyokat vagy sem [7].

Egy additív művelet *megőrzi az összefüggőségi viszonyokat*, ha a komponensek számán nem változtat, továbbá a művelet végrehajtásával kapott kép valamennyi komponense a kiindulási képnek egy és csakis egy komponensét tartalmazza. Más szavakkal: nem hoz létre és nem is olvaszt össze komponenseket.

Egy szubtraktív művelet *megőrzi az összefüggőségi viszonyokat*, ha a kiindulási kép minden komponenséből megtart legalább egy objektumpontot, továbbá bármely kettő, a művelet által nem törölt, a kiindulási képen összefüggő (ugyanazon komponenshez tartozó) objektumpont a művelet végrehajtása után is összefüggő marad. Másképpen: komponenst nem töröl és nem is vág szét.

A (rész)műveleteket, illetve a műveletek elvárásainak eleget tevő megoldásainkat 3×3-es *maszk*okkal, sablonokkal írjuk le. A maszkok elemeinek értéke 3-féle

lehet: „0", „1" és „·" (közömbös, *don't care*). Az $A = (a_{ij})$ $n \times m$ méretű bináris képből az $M = (m_{uv})$ 3×3-as maszkkal adott lokális párhuzamos művelet az $A' = (a'_{ij})$ $n \times m$-es bináris képet hozza létre, ahol:

$$a'_{ij} = a_{ij} \oplus \bigwedge_{u=1}^{3} \bigwedge_{v=1}^{3} a_{i+u-2,j+v-2} \odot m_{uv}$$
$$(i = 1, 2, \ldots, n,\ j = 1, 2, \ldots, m).$$

A „$\oplus$" az antivalenciának (a kizáró vagy műveletnek), a „$\bigwedge$" a konjunkciónak (a logikai „és" műveletnek) felel meg, míg a „$\odot$" műveletet a következőképpen definiáljuk:

$$a \odot m = \begin{cases} 1, & \text{ha } m = \text{„}\cdot\text{"} \text{ vagy } a = m, \\ 0, & \text{különben.} \end{cases}$$

A fentiek szerint egy képpont értéke megváltozik (objektumpontból háttérpont lesz és fordítva), ha az érvényes maszkkal az adott pont 3×3-as környezetét „letakarva" a maszk valamennyi „1" és valamennyi „0" eleme objektumponttal, illetve háttérponttal kerül fedésbe, míg a „közömbös" maszkelemek által lefedett képpontok között objektumpontok és háttérpontok egyaránt lehetnek. („Közömbös" maszkpozíciók használatával egy maszk több 3×3-as környezetre, részképre is illeszkedhet. k darab „közömbös" pozíciót tartalmazó 3×3-as maszk a lehetséges $2^9 = 512$-féle környezetből 2^k esetet fed le $(0 \leq k \leq 9)$.)

Az A' kép szélső pontjainak számításakor feltételezzük, hogy a kiindulási képet egy pixel vastagságú, háttérpontokból álló keret veszi körül, vagyis:

$$a_{0j} = 0, \quad a_{n+1,j} = 0 \qquad (j = 0, 1, \ldots, m+1),$$
$$a_{i0} = 0, \quad a_{i,m+1} = 0 \qquad (i = 0, 1, \ldots, n+1).$$

A lokális párhuzamos műveleteket általában több — „0", „1" és „közömbös" elemeket tartalmazó — maszkkal adjuk meg. Több maszk egyidejű érvényessége esetén egy adott képpont értéke megváltozik, ha a maszkkészlet legalább egy tagja a fenti módon illeszkedik az adott pont 3×3-as környezetére.

Könnyen belátható, hogy több maszkkal megadható tetszőleges lokális bináris képfeldolgozó művelet, mivel a 3×3-as lokális környezettől való függés minden esetben leírható egy 9-változós Boole-függvénnyel. Az a képpont értéke megváltozik, ha az f Boole-függvény „1" értéket vesz fel a pont 3×3-as környezetére:

$$a' = a \oplus f(a_1, \ldots, a_9)$$

(ahol „$\oplus$" az antivalenciát, „$a_1, \ldots, a_9$" pedig az a pixel 3×3-as környezetébe eső pontokat jelöli).

Az f Boole-függvényhez tartozó *diszjunktív normálformula* minden egyes tagjához hozzárendelhetünk egy maszkot. A maszkban egy adott változóhoz tartozó pozíción „1" álljon, ha a vizsgált tagban a változó *ponált* értékkel szerepel, „0" legyen, ha *negálva* van, és válasszuk „közömbös"-nek, ha a változó nem szerepel a tagban. Az f függvény értéke „1", ha a diszjunktív normálformulában legalább egy tag értéke „1", vagyis a tagokhoz rendelt maszkok valamelyike illeszkedik a vizsgált pont környezetére.

2.1. Üregfeltöltés

Az érvényben lévő összefüggőségi reláció a kép háttérpontjait is diszjunkt részhalmazokra bontja. Azon ekvivalencia-osztályt, amely nem tartalmaz háttérpontot az $n \times m$-es kép széléről (első vagy n-edik sorából, első vagy m-edik oszlopából), *üreg*nek (*hole*) nevezzük. Megjegyezzük, hogy általában eltérő összefüggőségi relációt szoktak a komponensekhez és a háttérhez rendelni: 4-összefüggőség a komponensekre, 8-összefüggőség a háttérre, vagy fordítva.

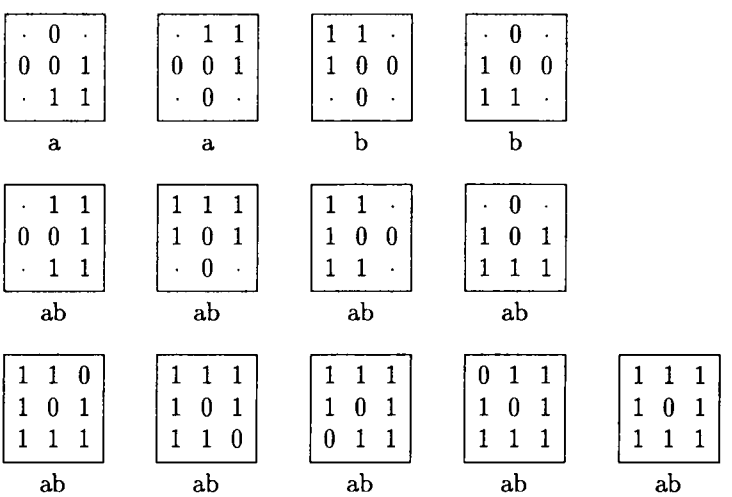
A 2.2. pontban ismertetésre kerülő zsugorító eljárások nem képesek egyetlen izolált ponttá összehúzni olyan komponenseket, melyek üregе(ke)t zárnak magukba. Az üreges komponensek eredményes zsugorítása érdekében a zsugorítás végrehajtása előtt egy, az üregeket feltöltő (az üregpontokat objektumpontokká változtató) transzformációt alkalmazunk.

Az 2.1.1. és a 2.1.2. pontokban ismertetésre kerülő additív műveletek — az üregpontok objektumponttá változtatásán túl — a komponenseket az őket befoglaló tömör téglalapokká igyekeznek hízlalni, ügyelve arra, hogy a 4- illetve a 8-összefüggőség ne sérüljön meg, vagyis a komponensek csak addig növekedhetnek, míg valamely másik komponens hízlalása annak nem állja útját. Az eljárások ezen tulajdonságait a Függelékben bizonyítjuk be.

Jelölje *Fill4(d,A)* illetve *Fill8(d,A)* az „A" képen végrehajtott d-lépéses 4-összefüggő üregfeltöltés (2.1.1. pont) illetve 8-összefüggő üregfeltöltés (2.1.2. pont) eredményét.

2.1.1. Üregfeltöltés 4-összefüggőség mellett

Az eljáráshoz konstruált maszkok:

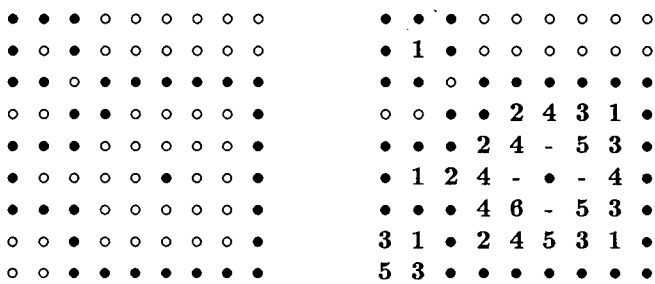


Az üregfeltöltés egy lépését két fázisra bontottuk: az elsőben az „a"-val és az „ab"-vel, a másodikban pedig a „b"-vel és az „ab"-val jelölt maszkok érvényesek.

Valamennyi maszk egyidejű alkalmazásával szeparált komponensek is összeolvadhatnának, tehát az eljárás nem őrizné meg a 4-összefüggőségi viszonyokat (4. ábra).

4. ábra. Példa az egyfázisú üregfeltöltés 4-összefüggőséget sértő hatására. Bal oldalon a kiindulási kép, jobb oldalon az eredmény látható.

Az eljárás a 4-összefüggőségi viszonyokat mindenképpen megőrzi, még annak árán is, hogy bizonyos üregpontokat megtart (lásd 5. ábra).

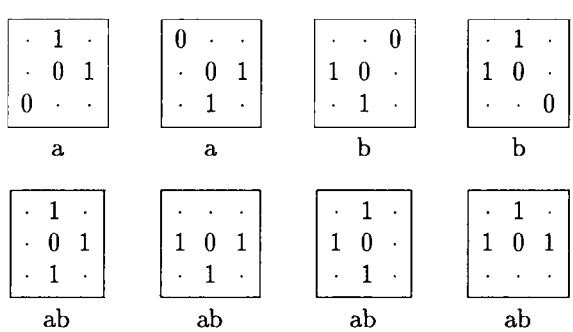


5. ábra. Példa üregpontok megmaradására. A példaként szolgáló bal oldali képre az eljárás a jobb oldali képet eredményezi. Az i-edik fázisban feltöltött pozíciókat az „i" számmal jelöltük ($1 \leq i \leq 6$)*, tehát a kétfázisú üregfeltöltés a harmadik lépésben fejeződött be. A nem feltölthető üregpontokat „-" jelöli. (A háttérre a 8-szomszédság érvényes.)*

A fenti 4-összefüggő eljárásunk és a 2.1.2. pontban tárgyalásra kerülő 8-összefüggő eljárás hatását a 7. ábrán vetjük össze.

2.1.2. Üregfeltöltés 8-összefüggőség mellett

Az eljárás a következő maszkokkal dolgozik:

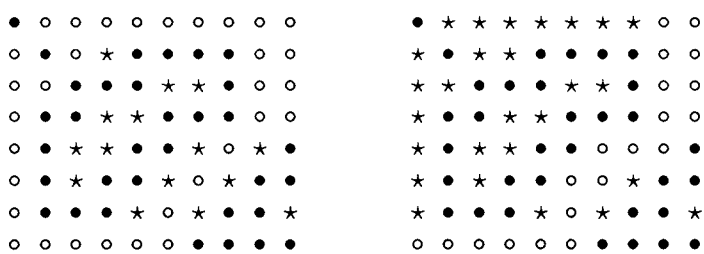


Az üregfeltöltés egy lépését két fázisra bontottuk: az elsőben az „a"-val és az „ab"-vel, a másodikban pedig a „b"-vel és az „ab"-vel jelölt maszkok érvényesek. A fenti maszkok egyetlen fázisban történő alkalmazása nem őrizné meg a 8-összefüggőségi viszonyokat (6. ábra).

6. ábra. Példa az egyfázisú üregfeltöltés 8-összefüggőséget sértő hatására. Bal oldalon a kiindulási kép, jobb oldalon az eredmény látható.

Megjegyzendő, hogy az eljárás — hasonlóan a 4-összefüggő üregfeltöltéshez — a hangsúlyt a 8-összefüggőség megőrzésére és nem az összes üregpont feltöltésére helyezi.

Az 7. ábrán bemutatjuk a 8-összefüggő üregfeltöltés és a 2.1.1. pontban tárgyalt 4-összefüggő eljárás hatását.

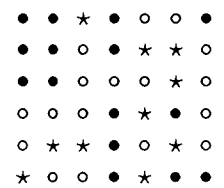


7. ábra. A 4- és a 8-összefüggő üregfeltöltés hatása ugyanazon kiindulási képre. A bal oldali képen a 4-, a jobb oldalin a 8-összefüggő eljárás eredménye látható. A feltöltött pozíciókat a „⋆" karakter jelöli.

2.2. Zsugorítás

Az alábbiakban két — az összefüggőségi viszonyokat megőrző szubtraktív — műveletet ismertetünk 4- és 8-összefüggő komponensek zsugorítására. Az eljárások üregе(ke)t nem tartalmazó komponenst egyetlen izolált pontra húznak össze, míg üreges komponensből *zárt görbét* (vagy egymással összefüggő zárt görbéket) állítanak elő. Zárt görbén olyan komponenst értünk, melynek minden pontja *összekötő pont*. Egy objektumpont összekötő pont, ha a 3×3-as környezetében mint egy 3×3-as bináris képen csak egyetlen komponenst találunk, mely komponens a középpont (a 3×3-as képen a $(2, 2)$ koordinátájú pixel) háttérponttá változtatásával több komponensre esne szét (8. ábra). Az üreges — de üregébe „ágyazott" komponenst nem

tartalmazó — komponensek is izolált ponttá zsugoríthatók, ha előzőleg üregfeltöltést (2.1. pont) alkalmazunk.



8. ábra. Példa 8-összekötő pontra. A „⋆" objektumpontok összekötő pontok, míg a „•" pontok nem azok.

Zsugorításra (*shrinking*) létezik számos más — ugyancsak lokális és párhuzamos — megoldás. LEVIALDI bináris minták számlálására feljesztette ki algoritmusát [8]. Az eljárás csak 2 × 2-es maszkokkal dolgozik, különlegessége még, hogy nem csak töröl, hanem bizonyos esetekben háttérpontot objektumponttá is változtat. Ily módon kibújik A. ROSENFELD klasszikus zsugorító eljárásokra felállított tételének [9] hatálya alól, miszerint egy szekvenciális zsugorítás legalább 3 × 3-as, míg a csupán törlő (objektumpontokból háttérpontokat gyártó) egyfázisú (!) párhuzamos zsugorító eljárás pedig legalább 5 × 5-ös maszkokat igényel.

KAMESWARA RAO és munkatársai 2-fázisú eljárása [10] 3 × 3-as maszkokkal dolgozik, az egyes komponenseket egy-egy izolált pontra húzza össze — az összefüggőségi viszonyok megőrzésével, az izolált pontok megtartása mellett. A módszer szintén alkalmaz kitöltő maszkokat is, melyek kiadják a a 2.1.2. pontbeli üregfeltöltő eljárás maszkjait.

Érdemes megemlíteni GÖKMEN és HALL eljárását is [11], mely a bináris kép pontjait sakktáblaszerűen két almezőre bontja fel. A 3 × 3-as maszkokat alkalmazó eljárás kétfázisú, az első fázisban csak a „világos", a másodikban pedig csak a „sötét" poziciók változhatnak meg. Az eljárásban ugyancsak szerepelnek kitöltő feltételek is.

A 2.2.2. pontban ismertetésre kerülő algoritmusra leginkább PRATT és KABIR [12] módszere hasonlít, mivel mindkettő kétfázisú, tisztán szubtraktív (csak töröl) és 3 × 3-as maszkokkal operál.

Valamennyi hivatkozott eljárást 8-összefüggő komponensek zsugorítására dolgozták ki, a 2.2.1. pontban leírt, 4-összefüggő zsugorító eljárásunkhoz hasonlóra nem bukkantunk.

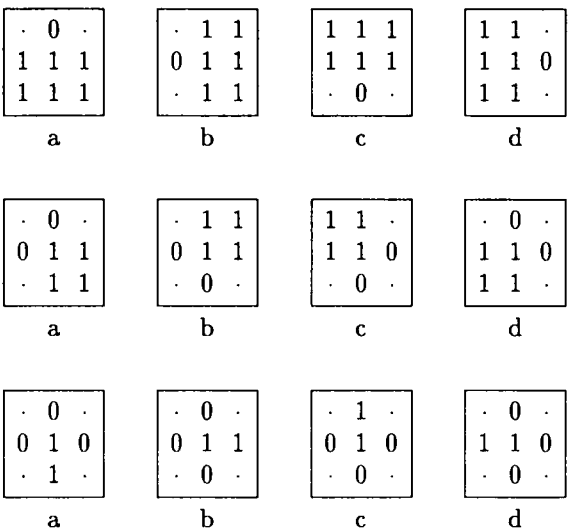
A hivatkozott és a 2.2. pontban található lokális és párhuzamos eljárások közös hátránya, hogy nem képesek izolált ponttá összehúzni olyan üreget tartalmazó komponenst, melynek van más komponenst magába foglaló ürege. Kivétel LEVIALDI módszere, mely viszont az izolált objektumpontokat is törli, így valamiképpen észlelni kell az egy ponttá zsugorodás pillanatát.

Az ismertetésre kerülő zsugorító eljárások tulajdonságait a Függelékben bizonyítjuk be.

A zsugorítást felhasználó zajszűrő módszer ismertetésénél *Shrink4(d,A)* illetve *Shrink8(d,A)* az A képen végrehajtott d-lépéses 4-összefüggő zsugorítás (2.2.1. pont) illetve 8-összefüggő zsugorítás (2.2.2. pont) eredményét fogja jelölni.

2.2.1. Zsugorítás 4-összefüggőség mellett

A műveletet leíró maszkok:



Az eljárás egy lépését négy fázisra bontottuk fel: az elsőben az „a"-val, a másodikban „b"-vel, a harmadikban a „c"-vel, a negyedikben pedig a „d"-vel jelölt maszkok érvényesek.

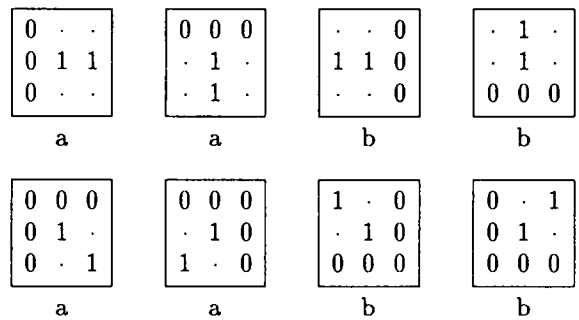
A fenti maszkok négynél kevesebb fázisban történő alkalmazása nem őrizné meg a 4-összefüggőségi viszonyokat, mivel a felső maszksor elemei közül bármely kettő egyidejű alkalmazása bizonyos komponensek „széteséséhez" vezetne (9. ábra).

9. ábra. A felső maszksor első két elemének egyidejű alkalmazása nem őrzi meg a 4-öszefüggőségi viszonyokat, mivel a bal oldali, egy 4-komponenst tartalmazó képből a jobb oldali, két 4-komponensből álló képet kapnánk.

Az eljárás hatását — összevetve a 2.2.2.-beli 8-összefüggő zsugorításával — a 10. ábrán mutatjuk be.

2.2.2. Zsugorítás 8-összefüggőség mellett

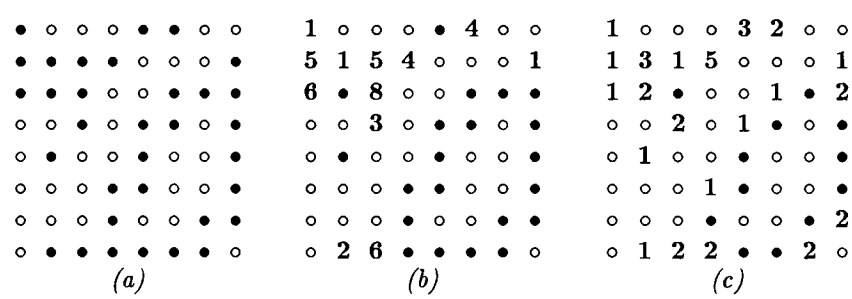
Az eljárás a következő maszkokkal dolgozik:



Az eljárás egy lépése két fázisból áll: az első az „a"-, a második pedig a „b"-jelű maszkokkal dolgozik.

A fenti maszkok egyidejű, egyetlen fázisban történő alkalmazása nem őrizné meg a 8-összefüggőségi viszonyokat. (Egyrészt, mivel A. Rosenfeld [9]-ben bebizonyította, hogy nem konstruálható egyfázisú, 3×3-as (csak törlő) maszkokkal operáló lokális párhuzamos zsugorító eljárás, másrészt látszik, hogy a fenti maszkok egyidejű alkalmazása eltüntetné például az összes 2 pontból álló komponenst vagy akár a 2×2-es négyzetet is.)

Az eljárás hatását — összevetve az előző pontbeli 4-összefüggő zsugorításával — a 10. ábra példájával szemléltetjük.



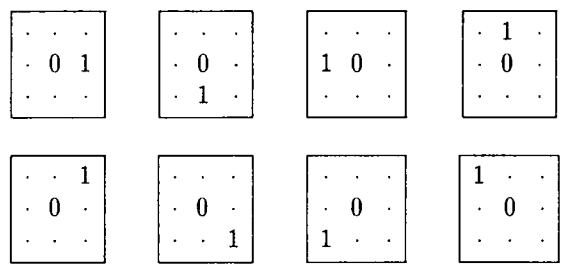
10. ábra. Példa a 4- és a 8-összefüggő zsugorító eljárásra. Az (a) kiindulási képből a 4-összefüggő zsugorítás a (b), a 8-összefüggő pedig a (c) képet állítja elő. Az i-edik fázisban törölt képpontokat „i"-vel jelöltük. (A 4-összefüggő eljárás 4-, a 8-összefüggő pedig 2-fázisú, így az eredmény a 2. illetve a 3. lépésben alakult ki.) Üreget nem tartalmazó komponens izolált pontra, üreges pedig zárt görbévé zsugorodott össze.

2.3. Terjesztés

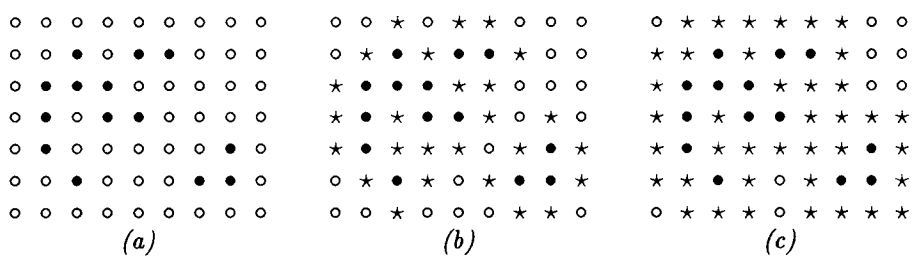
A terjesztés (*propagation* [1]) egy lépése objektumponttá teszi mindazon háttérpontokat, melyeknek van objektumpont szomszédjuk a figyelembe vett szomszédsági reláció mellett. A művelet tehát additív, és könnyen belátható, hogy nem őrzi meg az összefüggőségi viszonyokat.

Megjegyezzük, hogy az irodalomban a terjesztés helyett a dilatáció (*dilation*) elnevezés is előfordul (például [12]-ben), de a dilatáció mint a matematikai morfológia (*mathematical morphology*) művelete [3] a terjesztésnél jóval általánosabb.

A művelet maszkjai:



A felső maszksor elemei érvényesek a 4-szomszédos terjesztés, mind a nyolc maszk pedig a 8-szomszédos terjesztés esetében. A műveletek hatását a 11. ábra szemlélteti.



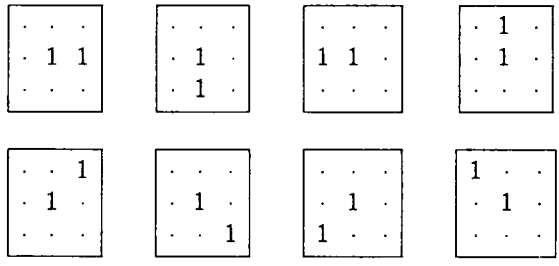
11. ábra. A terjesztés hatása. Az (a) kiindulási képből 4-szomszédos illetve 8-szomszédos terjesztéssel a (b) illetve a (c) kép áll elő, ahol „⋆" jelöli az objektumponttá váló háttérpontokat.

Jelölje *Propagation4(A)* illetve *Propagation8(A)* az A képen végrehajtott 4-szomszédos illetve 8-szomszédos terjesztés eredményét.

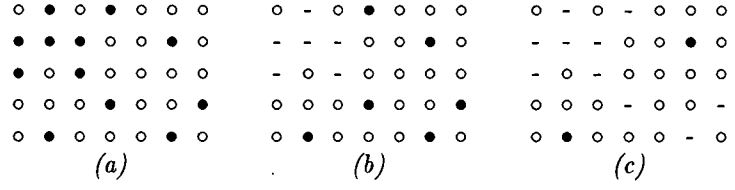
2.4. Izolált pontok detektálása

Izolált pontok detektálásán azt a műveletet értjük, mely minden nem-izolált objektumpontot töröl a képről, tehát alkalmazása után csak az izolált objektumpontok maradnak meg. A művelet szubtraktív és az összefüggőségi viszonyokat nem őrzi meg.

A műveletet leíró maszkok:



A felső maszksor elemei érvényesek 4-izolált pontok detektálásakor, míg mind a nyolc maszk szükséges 8-izolált pontok esetében. A műveletek szemléltetésére a 12. ábra szolgál.



12. ábra. Példa izolált pontok detektálására. Az (a) kiindulási képre a 4-szomszédságot figyelembe véve a (b), a 8-szomszédság mellett pedig a (c) képet kapjuk. Törlésre a „-" pontok kerültek.

Jelölje *IsoDet4(A)* illetve *IsoDet8(A)* az *A* képen végrehajtott 4-izolált illetve 8-izolált pontok detektálásának eredményét.

2.5. Izolált pontok törlése

A szubtraktív művelet az izolált objektumpontok, az egyetlen pontból álló komponensek eltávolítására szolgál.

A 4-izolált illetve a 8-izolált pontok törlése a következő maszkokkal adható meg:

(Az eljárás pontosan azokat az objektumpontokat törli, melyeket a 2.4. pontban tárgyalt, az izolált pontokat detektáló művelet megtart, és viszont, így nem adunk meg az eljárás hatását bemutató ábrát.) Jelölje *IsoRem4(A)* illetve *IsoRem8(A)* az *A* képen végrehajtott 4- illetve 8-izolált pontok eltávolításának eredményét.

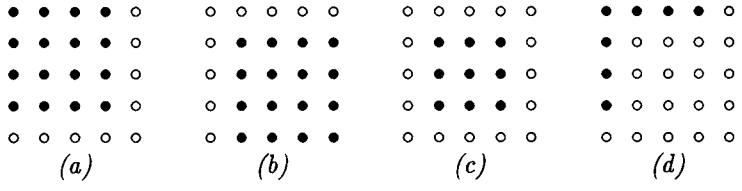
2.6. Logikai műveletek

Összetett zajszűrő eljárásunk 2 kétváltozós, pixelenkénti logikai műveletet, az „és"-t és az „és-nem"-et igényli.

Legyen $A = (a_{ij})$ és $B = (b_{ij})$ két $n \times m$-es bináris kép. Az „és" művelet eredményeként a $A \wedge B = C = (c_{ij})$, az „és-nem" végrehajtásával pedig a $A \wedge \overline{B} = D = (d_{ij})$ — ugyancsak $n \times m$-es — képeket kapjuk, ahol:

$$c_{ij} = a_{ij} \wedge b_{ij} \qquad \text{és}$$
$$d_{ij} = a_{ij} \wedge \overline{b_{ij}} \qquad (i = 1, 2, \ldots, n, \quad j = 1, 2, \ldots, m) .$$

A műveletre a 13. ábra mutat példát.



13. ábra. Példa az „és" és az „és-nem" műveletekre. A műveletek első operandusza az (a), a második a (b) kép. Az „és" művelet a (c), az „és-nem" pedig a (d) képet eredményezi.

Jelölje *And(A,B)* illetve *AndNot(A,B)* az A és a B képen végrehajtott pontonkénti „és" illetve „és-nem" műveletek eredményét.

3. Az összetett zajszűrő eljárás

Zajszűrő módszerünk erénye, hogy nem változtatja meg a megmaradó komponenseket, továbbá a törlendő komponensek, zajok mérete sem korlátozott. Hátránya viszont hogy, a komponens-méret fogalma nem általános (matematikailag nehezen leírható), továbbá az eljárás még kettő, a kiindulási és az eredményül kapott képpel megegyező méretű kép tárolását is igényli.

Módszerünknél a komponens-méretet az alkalmazott zsugorító eljárás határozza meg: egy komponens mérete d, ha pontosan a d-edik zsugorító lépésben húzódik rá egyetlen izolált pontra ($d \geq 0$, az izolált pont mérete 0, az izolált ponttá nem zsugorítható komponensek mérete pedig végtelen). A helyzetet bonyolítja, ha üreget tartalmazó komponenseket is törölni akarunk, ugyanis a javasolt üregfeltöltő eljárások (2.1. pont) megváltoztathatják még az üreget nem tartalmazó komponensek méretét is (14. ábra).

14. ábra. Példa az üregfeltöltés méretváltoztató hatására. Az bal oldali kép 8-összefüggő komponensének mérete a 2.2.2. pontbeli zsugorítás mellett 6 (mivel az a 6. lépésben húzza össze a „⋆" pontra). Ugyanazon komponens az 2.1.2.-beli üregfeltöltéssel a jobb oldali képen látható 5 × 5-ös tömör négyzetté alakul át, melynek mérete 2 (ugyancsak a 2.2.2.-beli zsugorító eljárás mellett).

Az alábbiakban rátérünk zajszűrő módszerünk ismertetésére.

Legyen A a kiindulási $n \times m$-es bináris kép, d a törlendő komponensek maximális mérete, továbbá legyen B és C két — ugyancsak $n \times m$-es — bináris kép (kezdetben közömbös tartalommal).

Az eljárást a közismert ALGOL metanyelv segítségével (lásd pl: [13]-ban) írjuk le 8-összefüggő komponensekre. (4-összefüggő komponensek esetén a programban a 8-szomszédságon alapuló műveleteket a 4-szomszédság szerinti párjukra kell cserélni.)

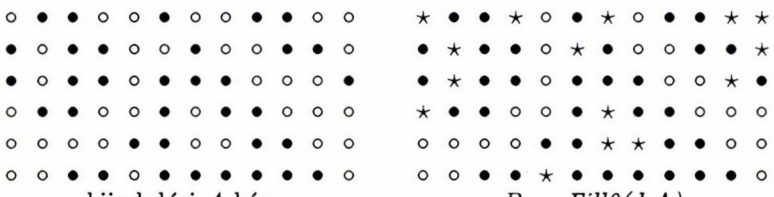
```
1.  if  d = 0  then
2.      A ← IsoRem8(A)
3.  else
        begin
4.          B ← Fill8(d,A)
5.          C ← Shrink8(d,B)
6.          C ← IsoDet8(C)
7.          for i ← 1 until d do
                begin
8.                  C ← Propagation8(C)
9.                  C ← And(B,C)
                end
10.         A ← AndNot(A,C)
        end
```

Megjegyezzük, hogy a módszer nem kötődik szorosan az ismertetett üregfeltöltő és zsugorító eljárásokhoz. Lecserélhetők más — hasonló tulajdonságokkal bíró — gyorsabb, vagy könnyebben programozható eljárásokra.

A fenti program 1. sorában d értékét vizsgáljuk. $d = 0$ esetén csak a 0-méretű izolált pontokat kell törölni (2. sor). Ha $d > 0$, akkor az eljárás egy d-lépéses üregfeltöltéssel indul (4. sor), biztosítandó az üreges komponensek zsugoríthatóságát. Ezt követően a d-lépéses zsugorítás egy-egy izolált pontra húzza össze a törlendő komponenseket (5. sor), majd eltávolítjuk a nem-izolált pontokat (6. sor). (E pillanatban A továbbra is a kiindulási képet tartalmazza, B-ben található az üregfeltöltés eredménye, C-ben pedig a törlendő komponensekre utaló izolált pontok vannak.) Az

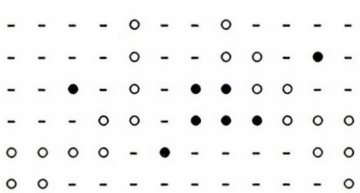
eljárás a terjesztés és az „és" műveletpár ciklikus ismétlésével (7.–9. sor) előállítja a C képen a törlendő komponensek üregfeltöltöttjeit. (A terjesztés a kezdetben izolált pontokat hízlalja, míg az „és" művelettel megakadályozzuk, hogy a hízlalás túlnőjön az üregfeltöltött komponenseken.) Végezetül a kiindulási képre és a törlendő komponensek üregfeltöltötteit (a törlendő komponenseket magukba foglaló, de a nem-törlendőkbe át nem nyúló komponenseket) tartalmazó C képre az „és-nem" műveletet alkalmazzuk (10. sor).

Kövessük végig az eljárás menetét ($d = 2$ és a 8-összefüggőség esetében) a következő példán. (A feldolgozás adott lépésében törölt objektumpontokat „-", a feltöltött háttérpontokat „⋆" jelöli.)

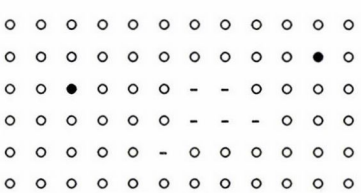


kiindulási A kép

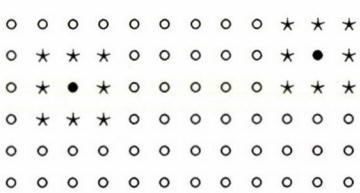
$B \leftarrow Fill8(d,A)$
(az üregfeltöltött kép)



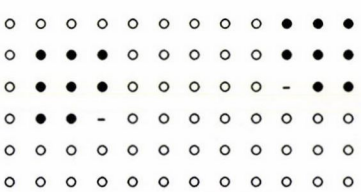
$C \leftarrow Shrink8(d,B)$
(a zsugorított kép)



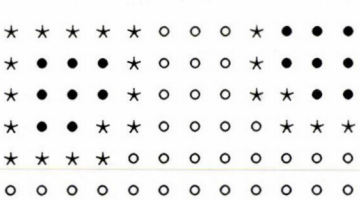
$C \leftarrow IsoDet8(C)$
(izolált pontok detektálása)



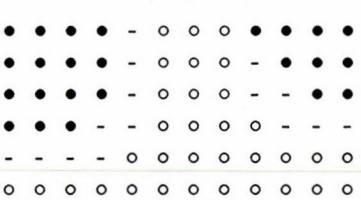
$C \leftarrow Propagation8(C)$
(az első terjesztés)



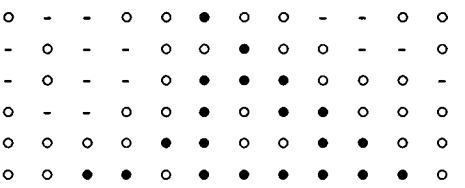
$C \leftarrow And(B,C)$
(az első „és")



$C \leftarrow Propagation8(C)$
(a d-edik terjesztés)



$C \leftarrow And(B,C)$
(a d-edik „és")



$A \leftarrow AndNot(A,C)$
(a zajszűrés befejezése)

Könnyen belátható, hogy az A kép d-nél (határozottan) nagyobb méretű komponenseit is törölhetjük, ha az eljárás végén az „és–nem" művelet helyett az „és"-t alkalmazzuk, vagyis a program 10. sorát a következőre cseréljük:

$$A \leftarrow And(A,C) .$$

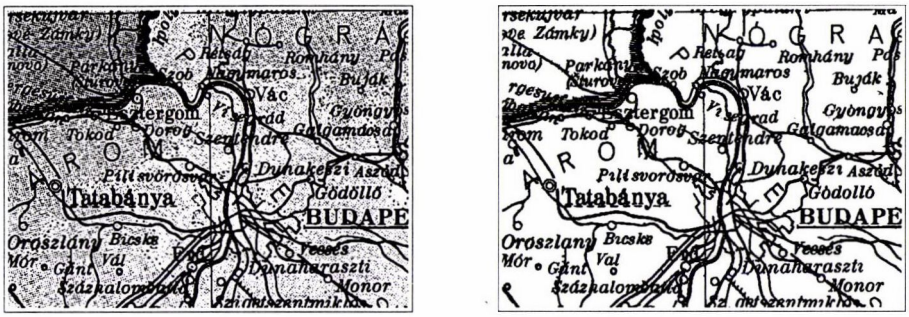
Az eljárás módosításával adott $[\, p+1,\ d\,]$ $(0 \leq p < d)$ méret-intervallumba eső komponensek is törölhetők. Ekkor az eljárás 5. sora helyett a következők szerepelnek:

$$C \leftarrow Shrink8(p,B)$$
$$C \leftarrow IsoRem8(C)$$
$$C \leftarrow Shrink8(d\text{-}p,C) .$$

A d-lépéses zsugorítást egy p- és egy $(d-p)$-lépéses részre osztottuk. A p-lépéses zsugorítás után a p-nél kisebb vagy egyenlő méretű komponensek húzodnak össze egy-egy izolált pontra. Ezen izolált pontokat töröljük, így a továbbiakban a p-nél kisebb vagy egyenlő méretű komponenseknek a C képen nem marad nyomuk, így A-ról nem fognak eltűnni. A maradék $(d-p)$-lépéses zsugorítás után pontosan azon komponensekből keletkezik izolált pont, melyeknek mérete az adott intervallumba esik, így az eljárás végére azok fognak A-ról törlődni.

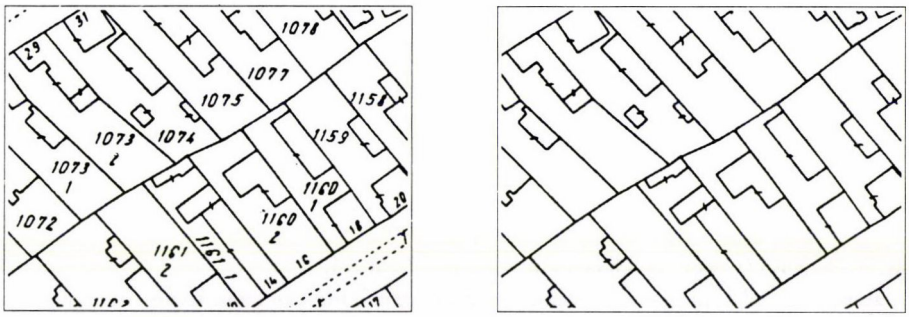
A $[\, p+1,\ d\,]$ méret-intervallumon kívül eső komponensek törlése úgy oldható meg, hogy a fenti intervallumos eljárás végén az „és–nem" helyett az „és" műveletet alkalmazzuk.

Zajszűrő módszerünk alkalmazhatóságát a 15. és a 16. ábrákkal szemléltetjük.



15. ábra

15. ábra. Példa a zajszűrő módszerünkre. Eljárásunk az (a) — 480 × 640 méretű — térképrészletre a (b) képet eredményezte. Az eltávolítandó 8-komponensek, zajok maximális méretének d = 4-et adtunk meg. (Az eljárást ugyanarra a képre hajtottuk végre, mint a 3. ábrán szemléltetett morfológiai szűrést. Ezúttal a nemtörölt komponensek változtatás nélkül megőrződtek, valamennyi helységnév olvasható maradt.)



16. ábra

16. ábra. Példa a zajszűrő módszerünk alkalmazására. Az (a) — 480 × 640-es, egy kataszteri térkép részletét tartalmazó — képre a d = 10-nél kisebb vagy egyenlő méretű 8-komponenseket eltávolító eljárás a (b) képet eredményezte, melyen a helyrajzi számok és a szaggatott vonalak már nem szerepelnek.

FÜGGELÉK

Az alábbiakban a 2.1. és a 2.2. pontokban ismertetett üregfeltöltő és zsugorító eljárások tulajdonságainak bizonyítását közöljük.

Definíció. Egy háttérpont *szigetelő pont*, ha objektumponttá változtatásával összekötő pont (lásd 2.2. pont) lesz belőle.

1. LEMMA. *Egy háttérpont a 256 lehetséges* 3×3*-as környezetéből 123 esetben bizonyul szigetelő pontnak (mind a 4-, mind a 8-szomszédság esetében).*

Bizonyítás. Vizsgáljuk a 4-szomszédságot. Nyilvánvaló, hogy egy háttérpont nem lehet szigetelő pont, ha nincs objektumpont 4-szomszédja, vagy csak egy darab objektumpont szerepel a 4-szomszédjai között. Kettő darab objektumpont 4-szomszédnál a következő esetek lehetségesek:

0	1	·
1	0	0
·	0	·

8 eset, a 3 elforgatottal együtt $4 * 8 = 32$ eset

·	0	·
1	0	1
·	0	·

16 eset, elforgatottjával együtt $2 * 16 = 32$ eset

Nézzük a szigetelő pontokat, ha három darab objektumpont szerepel a 4-szomszédok között:

X	1	Y
1	0	1
·	0	·

$X + Y < 2$ (3 lehetséges érték)
$3 * 4$ eset, a 3 elforgatottal együtt $4 * (3 * 4) = 48$ eset

Végezetül, ha valamennyi 4-szomszéd objektumpont:

X	1	Y
1	0	1
V	1	Z

$X + Y + V + Z \leq 2$, ami 11 esetben teljesül

A fentieket összeadva 123 lehetséges esetet kapunk. A 8-szomszédságra vonatkozó állítást hasonlóképpen láthatjuk be.

2. LEMMA. *Egy objektumpont 256 féle* 3×3*-as környezetéből 123 esetben bizonyul összekötő pontnak (4- és 8-szomszédságra egyaránt).*

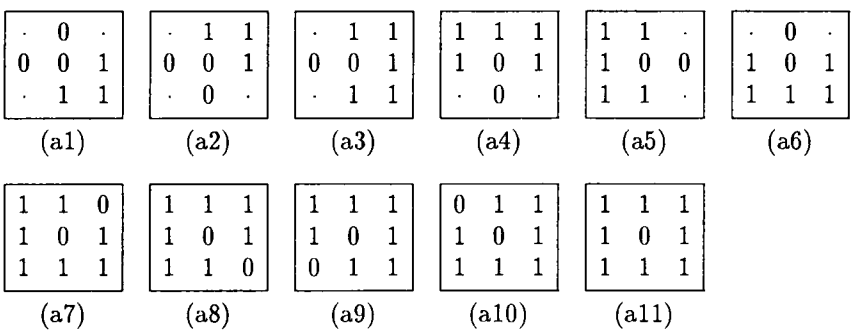
Bizonyítás. Az összekötő pontok és a szigetelő pontok között definíció szerint létezik egy-egyértelmű megfeleltetés (bijekció), így ugyanannyi a szigetelő és az összekötő pontok eseteinek száma, vagyis az 1. Lemma szerint 123.

1. TÉTEL. *A 2.1.1. pontban ismertetett 4-összefüggő üregfeltöltő eljárás párhuzamos végrehajtása megőrzi az összefüggőségi viszonyokat.*

Bizonyítás. Megállapíthatjuk, hogy az eljárás maszkjai csak nem 4-szigetelő pontokra illeszkedhetnek és páronként idegenek, vagyis egy háttérpont legfeljebb csak egyikük által töltődhet fel.

A maszkkészlet bármelyik maszkjáról belátható, hogy illeszkedésekor a „0" vagy közömbös pozicióival lefedett háttérpontokra az adott maszkkal azonos fázisban érvényes maszkok csak úgy illeszkedhetnek, hogy szeparált komponensek nem olvadnak össze.

Tekintsük a kétfázisú eljárás első fázisának maszkjait:



Vizsgáljuk meg egy, az (a1) maszkkal feltölthető háttérpont bal 4-szomszédjának feltölthetőségét. Ezen bal szomszédra kizárólag az (a5) maszk illeszkedhet, ha a vizsgált pont-pár környezete a következő:

$$\begin{array}{cccc} 1 & 1 & 0 & \cdot \\ 1 & \circledast & \odot & 1 \\ 1 & 1 & 1 & 1 \end{array}$$

„$\circledast$" a bal szomszéd,
„$\odot$" a vizsgált pont,

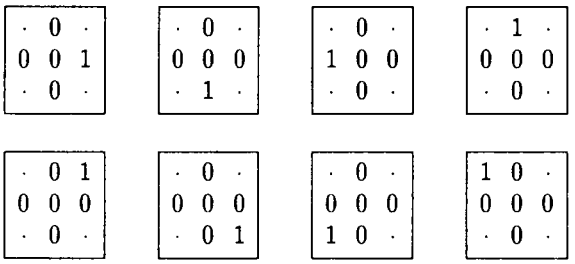
Látható, hogy a „$\circledast$" és a „$\odot$" pontok feltöltése nem változtat a 4-összefüggőségi viszonyokon.

Hasonlóképpen látható be valamennyi maszk összes olyan poziciójára, mely feltölthető háttérponttal kerülhet fedésbe, hogy a lefedett pontok objektumponttá válásával a 4-összefüggési viszonyok nem változnak meg. A második fázisra is is teljesülnek a fentiek, mivel annak maszkjai elforgatottjai az első fázis maszkjainak.

2. TÉTEL. *A 2.1.1. pont 4-összefüggő üregfeltöltő eljárása egy 4-komponenst a befoglaló tömör téglalapjává növeszt, ha a hízlalás más komponens(ek) feltöltésébe nem ütközik.*

Bizonyítás. Azon 4-szigetelő (háttér)pontokra, melyeknek van olyan objektumpont 4-szomszédjuk, mely álló, tömör téglalaphoz tartozik, a következő maszkok

illeszkedhetnek:



Az első sor páronként idegen maszkjai 4–4 közömbös poziciót tartalmaznak, vagyis $4 * 16 = 64$ esetet fednek le. A 2. sor maszkjai 15 lehetséges esetre illeszkednek (a négy sarokelem 16-féle kitöltéséből csak a csupa „0" marad ki). Így a nem 4-szigetelő, tömör téglalappal szomszédos háttérpontok 3×3-as környezete 79-féle lehet.

A fenti maszkok nyilvánvalóan ütköznek az üregfeltöltés maszkjaival, mivel azokban a középpontoknak legalább kettő 4-szomszédja „1" értékű, míg a fentieknél csak egy.

Vegyük számba az üregfeltöltő maszkok által lefedett (szintén nem 4-szigetelő) eseteket:

Az első sor 4 maszkja egyenként 8–8, közös esetet nem tartalmazó környezetet fed le, vagyis 32 esetre illeszkedik.

A második maszksor elemei $4 * 4 = 16$ esetet írnak le, egymással és az első sor maszkjaival nincsenek átfedésben.

A harmadik maszksorban 5 újabb esetre bukkanhatunk.

A tömör téglalappal 4-szomszédos, nem 4-szigetelő pontok 79 esete, valamint az üregfeltöltő eljárás maszkkészletének $32 + 16 + 5 = 53$ esete és a csupa „0" környezet maradéktalanul kiadja az 1. Lemma szerinti $256 - 123 = 133$ lehetséges nem 4-szigetelő pontot.

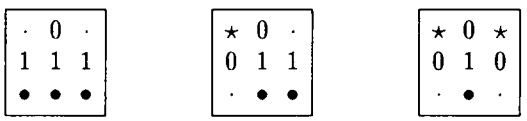
A fentiekből következik, hogy üregfeltöltő eljárásunk tömör téglalapon nem változtat és minden más komponens esetében, ha talál nem 4-szigetelő, a komponenssel 4-szomszédos háttérpontot, akkor azt feltölti.

Az 1. Tétel és a 2. Tétel állításai a 2.1.2-beli 8-összefüggő üregfeltöltő eljárásra is megfogalmazhatók. Bizonyításuk követheti a fentiek gondolatmenetét. (Tömör téglalappal határos nem 8-szigetelő pont 20-féle lehet, a 8-összefüggő üregfeltöltés maszkkészlete pedig 112 esetet fed le, így a csupa „0" környezettel maradéktalanul kiadják az 1. Lemma 133 lehetséges nem 8-szigetelő pontját.

A 2.1.-ben leírt üregfeltöltő eljárások után foglalkozzunk a 2.2. pontbeli zsugorítással. Az alábbiakban a 2.2.1.-beli 4-összefüggő zsugorítás tulajdonságait bizonyítjuk be. (A 8-összefüggő eljárás (2.2.2. pont) elemzése analóg módon történhet.)

3. TÉTEL. *A 2.2.1. pontban közölt 4-összefüggő üregfeltöltő eljárás párhuzamos végrehajtása megőrzi a 4-összefüggőségi viszonyokat.*

Bizonyítás. Az eljárás maszkjai nem illeszkednek 4-összekötő pontokra, továbbá bármelyik maszk az illeszkedésekor lefed legalább egy, az adott fázisban nem-törölhető objektumpontot. Pl: a zsugorító eljárás első fázisának maszkjai csak úgy illeszkedhetnek egy objektumpont környezetére, hogy az illető maszkok „•" szimbólummal jelölt „1" poziciói és a „⋆"-jelű közömbös értéket tartalmazó elemei nem-törölhető objektumpontokkal kerülhetnek csak fedésbe:



(a1) (a2) (a3)

(Az (a1), (a2) és (a3) maszkokkal csakis olyan objektumpontok törölhetők, melyeknek alsó 4-szomszédjuk objektumpont és a felső 4-szomszédjuk háttérpont.)

Vizsgáljuk meg valamely, az (a1) maszkkal törölhető objektumpont bal 4-szomszédjának törölhetőségét, ha az illető pixel is objektumpont. Ez a pont elvileg törölhető az (a1) vagy az (a2) maszkkal (az (a3) maszk nem illeszkedhet rá). Vizsgáljuk meg egy ilyen objektumpont-pár környezetét (jelölje „•" a nem-törölhető objektumpontokat):

· 0 0 · 1 ⊛ ⊙ 1 • • • •	„⊛" a bal szomszéd, „⊙" a vizsgált pont, mindkettőjükre (a1) illeszkedett
· 0 0 · 0 ⊛ ⊙ 1 · • • •	„⊛" az (a2) által törölt bal szomszéd, „⊙" az (a1) által törölt vizsgált pont

Látható, hogy mindkét esetben a „⊛" és a „⊙" pont törlése nem változtat a 4-összefüggőségi viszonyokon.

Hasonlóképpen, mindhárom maszk valamennyi, elvileg törölhető pontot lefedő poziciójára belátható, hogy a lefedett pontok törlése nem változtat a 4-összefüggőségi viszonyokon. A 4-fázisú zsugorító eljárás 2., 3. és 4. fázisaira is teljesülnek a fentiek, mivel azok maszkjai az első fázis maszkjainak elforgatottjai.

4. TÉTEL. *A 2.2.1. pont 4-összefüggő zsugorító eljárása a komponenseket izolált ponttá, vagy — üreges komponens esetében — csupa 4-összekötő pontból álló komponenssé változtatja.*

Bizonyítás. Az eljárás maszkjait vizsgálva megállapíthatjuk, hogy izolált pontot nem törölhetnek (minden maszk-középpontnak van legalább egy „1" 4-szomszédja), továbbá 4-összekötő pontra egyik maszk sem illeszkedhet. Ily módon 4-izolált pontra és zárt görbére (vagy 4-összekötő pontokkal összeláncolt zárt görbékre) az eljárás invariáns.

Az alábbiakban megmutatjuk, hogy minden más 4-komponensen talál az eljárás törölhető ponto(ka)t.

Egy objektumpont 16-féle 3 × 3-as környezetben lehet 4-izolált pont (a 4-szomszéd mindegyike háttérpont, a 4 sarokelem tetszőleges). A 2.Lemma szerint 123 a 4-összekötő pontok esetszáma, tehát a 256 lehetséges 3 × 3-as környezetből nem 4-izolált és nem 4-összekötő objektumpontból 117-féle lehet.

Könnyen belátható, hogy az eljárás maszkjai 112 esetre illeszkednek, így 5 eset maradt ki, mégpedig a következők:

$$\begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix} \quad \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

Nyilvánvaló, hogy egy 4-komponens (ami nem egyetlen izolált pont és nem csak összekötő pontokból áll), nem tartalmazhat csakis olyan pixeleket, melyek mindegyikére illeszkedik a fenti 5 maszk valamelyike. Tehát kell, hogy legyen olyan pontja is, mely törölhető a zsugorító eljárásunk valamelyik maszkjával.

Végezetül az üreges komponensekről: Az eljárás maszkjai csak törölnek, így az üregpontok nem töltődhetnek fel, sőt üregponttal 4-szomszédos törölhető objektumpont maga is üregponttá válik. Ennélfogva üreges komponensek nem húzódhatnak rá egyetlen 4-izolált pontra.

IRODALOM

[1] A. ROSENFELD, A. C. KAK, *Digital picture processing* (Academic Press, New York, 1976).

[2] R. C. GONZALES, R. E. WOODS, *Digital image processing* (Addison-Wesley, Reading, Massachusetts, 1992).

[3] J. SERRA, *Image analysis and mathematical morphology* (Academic Press, London, 1988).

[4] P. KUOSMANEN, L. KOSKINEN, J. ASTOLA, „Detail preserving morphological filtering", in Proc. of the 11th IAPR, Vol. III (IEEE Computer Society Press, Los Vaqueros Circle, 1992), 236–239.

[5] S. RAY, „A heuristic noise reduction algorithm applied to handwritten numeric characters", *Pattern Recognition Letters* **7** (1988), 9–12.

[6] F. ALI, T. PAVLIDIS, „Noise filtering in binary pictures by combinatorial techniques", *Pattern Recognition* **15. No. 3** (1982), 131–135.

[7] C. V. KAMESWARA RAO, P. E. DANIELSSON, B. KRUSE, „Checking connectivity preservation properties of some types of picture processing operations", *CGIP* **8** (1978), 299–309.

[8] S. LEVIALDI, „On shrinking binary patterns", *Communications of ACM* **15** (1972), 7–10.

[9] A. ROSENFELD, „Connectivity in Digital Pictures", *J. ACM* **17** (1970), 146–160.

[10] C. V. KAMESWARA RAO, B. PRASADA, K. R. SARMA, „A parallel shrinking algorithm for binary patterns", *CGIP* **5** (1976), 265–270.

[11] M. GÖKMEN, R.W. HALL, „Parallel shrinking algorithms using 2-subfields Approaches", *CVGIP* **52** (1990), 191–209.

[12] W. K. PRATT, I. KABIR, „Morphological binary image processing with a local neighbourhood pipeline processor", in Frontiers in computer graphics (Proc. of Computer Graphics Tokyo '84) (Springer-Verlag, Tokyo Berlin Heidelberg New York, 1985), 321–343.

[13] A. V. AHO, J. E. HOPCROFT, J. D. ULLMAN, *The design and analysis of computer algorithms* (Addison-Wesley, Reading, Massachusetts, 1974).

(Beérkezett: Beérkezett: 1994. augusztus 4.)

JÓZSEF ATTILA TUDOMÁNYEGYETEM
ALKALMAZOTT INFORMATIKAI TANSZÉK
6701 SZEGED, ÁRPÁD TÉR 2. PF: 652,
E-mail: K.Palagyi@inf.jate.u-szeged.hu

A LOCAL PARALLEL NOISE REDUCTION ALGORITHM FOR BINARY IMAGES

KÁLMÁN PALÁGYI

A noise reduction algorithm for parallel processing of binary images is proposed. The complex algorithm is decomposed a sequence of local parallel operations based on window size 3×3. Emphasis is to be put, that the proposed algorithm is capable of preserving perfectly the all image details and the size of the components to be removed is not restricted.

Alkalmazott Matematikai Lapok **17** (1993) 397–411

KIS RENDŰ PROJEKTÍV SÍKOK METSZÉSSZÁMÁNAK SZÁMÍTÓGÉPES VIZSGÁLATA

BÉRES LÁSZLÓ ÉS ILLÉS TIBOR

Budapest

Cikkünkben Erdős Pál véges projektív síkok blokkoló halmazainak metszésszámával kapcsolatos sejtésével foglalkozunk. Megadjuk a probléma egy egészértékű lineáris programozási (ELP) modelljét adott rendű projektív sík esetén. Figyelembe véve az egzakt módszerek műveletigényét és számítógépes lehetőségeinket, mohó módszer alkalmazása mellett döntöttünk. A probléma matematikai programozási megközelítése a véges projektív síkok elméletében alapvetően új. Mohó algoritmusunkkal előállítottuk az ELP szuboptimális megoldását (jelöljük $\alpha_G(q)$-val) a $7 \leq q \leq 89$ intervallumba eső prímrendekre, valamint a $q = 8, 9, 16$ prímhatványrendekre is. Mohó algoritmusunk egyben olyan eljárás, amellyel egy adott intervallumba (esetünkben $[\alpha_G(q), q]$, $\alpha_G(q) < q$) tartozó bármely α egészértékhez, α metszésszámú blokkoló halmaz konstruálható. Az $\alpha_G(q)$ értéke számítógépes tapasztalataink szerint $c \log q$, ahol $1 \leq c \leq 2$.

Ha $q = 89$ akkor az ELP modell 8011 bináris és egy korlátos egészértékű változót, továbbá 8011 alsó és felső korláttal rendelkező egyenlőtlenséget tartalmaz. A dolgozat végén közöljük számítógépes eredményeinket, és azok összehasonlítását az elméleti eredményekből a blokkoló halmazok metszésszámára nyert korlátokkal.

Kulcs szavak: véges projektív sík, blokkoló halmaz, 0-1 programozás.

1. Bevezetés

Hipergráfokkal kapcsolatban számos érdekes szélsőérték feladat fogalmazható meg. A problémák gyakran 0-1 (vagy egészértékű) lineáris programozási feladatként (is) megadhatók. FÜREDI [10] összefoglaló munkájában említettek közül csupán eggyel foglalkozunk, mégpedig a véges projektív síkok blokkoló halmazainak metszésszámával kapcsolatos Erdős problémával. Dolgozatunk elkészítését az motiválta, hogy minden rendre egyaránt működő konstrukció, sőt elméleti eredmény is csak kevés ismert (ERDŐS, SILVERMANN és STEIN [9]). Igazán látványos eredményeket [2,5,11] csak prímhatványrendekre találunk. Ez ösztönzött bennünket arra, hogy számítógépes vizsgálatainkat prímrendű projektív síkokon végezzük el. PASCAL nyelven írt programunkkal, PC-AT 386/25 Mhz-es, 2 Mbyte RAM-mal rendelkező személyi számítógépen állítottuk elő szuboptimális megoldásainkat, a $7 \leq q \leq 89$ intervallumba eső prímrendekre, valamint a $q = 8, 9, 16$ prímhatványrendekre is. A szuboptimális értékhez tartozó blokkoló halmazaink néhány jellemző adatát (mérete, fedőszáma, metszésszáma) a 2. táblázatban foglaljuk össze, míg az 1. táblázat a szakirodalomból ismert eredményekkel való összehasonlítást illusztrálja. A 3. táblázatunkban karakterizáljuk megoldásainkat.

A 2. fejezetben az Erdős-probléma megértéséhez szükséges fogalmakat adjuk meg. Összefoglaljuk a témakörben ismert eredményeket és megfogalmazzuk a szükséges feltételét annak, hogy egy minimális blokkoló halmaz $B(\gamma+1)$ tulajdonságú legyen. Szükséges feltételünkből $\gamma = 3$ esetén azt nyerjük, hogy $B(4)$ tulajdonsággal rendelkező minimális blokkoló halmaz csak a 3-ad illetve 4-edrendű projektív síkon létezik. A 3. fejezetben az adott rendű projektív sík esetére megfogalmazunk egy egészértékű lineáris programozási (ELP) modellt, amely az Erdős-probléma egy lehetséges modellje. Definiálunk egy mohó algoritmust, amellyel az ELP feladat szuboptimális megoldását tudjuk előállítani. A $q = 7$ rend esetén szuboptimális megoldásunk egyben optimális is (figyelembe véve a 2.1. Állítást). Megadjuk az algoritmus megállási kritériumát és elemezzük általános lépését is. A 4. fejezetben ismertetjük számítógépes eredményeinket. A mohó módszerrel nyert eredményeinket megkíséreltük a CPLEX [8] programcsomag segítségével megjavítani. A CPLEX programmal *Convex C3820* miniszuperszámítógépen illetve *HP 9000/720* számítógépen teszteltük modellünket. A rendelkezésünkre álló, korlátozott gépidő és memória kapacitás mellett egyetlenegy szuboptimális megoldást sem sikerült megjavítani.

2. Az Erdős-probléma

*Véges projektív sík*nak nevezünk egy (Π, Λ) párt, ahol $\Pi \neq \emptyset$ halmaz (elemei *pont*ok) és Λ elemei pedig Π bizonyos részhalmazai (*egyenes*ek), melyek eleget tesznek az alábbi feltételeknek

(i) Bármely két különböző ponthoz létezik pontosan egy, mindkét pontot tartalmazó egyenes.

(ii) Bármely két különböző egyenesnek pontosan egy közös pontja van.

(iii) Létezik olyan négy pont, amelyek közül semelyik hármat nem tartalmazza egy egyenes.

(iv) Létezik olyan egyenes, amely pontosan $q+1$ pontot tartalmaz (ahol $q \geq 2$, egész szám).

Az (i) és (ii) feltételek a klasszikus projektív sík illeszkedési axiómáihoz tartoznak és a pontok illetve egyenesek közötti illeszkedési kapcsolatot adják meg. Az (iii) feltétel biztosítja az alakzatok síkbeliségét és így kizárja az elfajuló eseteket. Az (iv) feltételt szokás még *végességi megkötés*nek is nevezni. Az (i)–(iv) feltételekkel definiált projektív síkot *q-adrendű*nek nevezzük és $\Pi(q)$-val jelöljük. A $GF(q)$ test[1] felett koordinátázható véges projektív síkot *q-adrendű Galois-sík*nak nevezzük és $PG(2,q)$-val jelöljük. A számítógépes konstrukciók során Galois-síkokkal dolgoztunk.

Legyen adott egy $A \subset \Pi$ ponthalmaz a q-adrendű projektív síkon. A sík adott egyenese *kitérő* illetve *metsző* az A halmazhoz viszonyítva, ha 0 vagy legalább 1

[1] $GF(q)$-val jelöljük a q elemű véges (Galois) testet.

közös pontja van A-val. Az egy pontban metsző egyeneseket *érintő*knek nevezzük. A $\Pi(q)$ projektív sík $B \subset \Pi$ ponthalmazát *blokkoló halmaz*nak (b.h.) mondjuk, ha bármely egyenes metszi a B halmazt, de az nem tartalmaz teljes egyenest. A B *minimális* blokkoló halmaz (m.b.h.), ha nem tartalmaz valódi részhalmazként blokkoló halmazt. Ekkor a blokkoló halmaz bármely pontjában létezik érintő. A m.b.h.-ok méretére BRUEN és THAS [5] a következő alsó és felsőkorlátot adta $q + \sqrt{q}+1 \leq |B| \leq q\sqrt{q}+1$. Egy B b.h.-t *t-lefedhető*nek nevezzük, ha pontjai lefedéséhez t különböző egyenes szükséges, [4].

A $\Pi(q)$ projektív sík B blokkoló halmazát $B(\gamma)$ *tulajdonságú*nak nevezzük, ha bármely l egyenes esetén $|l \cap B| < \gamma$, $(\gamma \in \mathbb{N})$. Nyilvánvaló, hogy $2 \leq \gamma \leq q$ teljesül. Hasonlóan, a $\Pi(q)$ projektív síkot $B(\gamma)$ *tulajdonságú*nak nevezzük, ha létezik $B(\gamma)$ tulajdonságú blokkoló halmaza. A $\gamma - 1$ értéket a blokkoló halmaz (projektív sík) *metszésszám*ának hívjuk.

A dolgozatunkban, kis prímrendű Galois-síkokon számítógéppel megvizsgált kérdés Erdős nevéhez fűződik.

Erdős-probléma: Létezik-e γ^* abszolút konstans, amelyre igaz, hogy bármely $q \geq q_0$ rend esetén a $\Pi(q)$ projektív sík $B(\gamma^*)$ tulajdonságú.

A fejezet további részében összefoglaljuk az Erdős-problémával kapcsolatos eredményeket, továbbá szükséges feltételt adunk a $B(\gamma)$ tulajdonságú b.h. létezésére.

ERDŐS, SILVERMANN és STEIN [9] dolgozatukban kimutatták, ha q elég nagy és $c \geq 2e$, ahol e a természetes alapú logaritmus alapszáma, akkor a $\Pi(q)$ projektív sík $B(c \log q)$ tulajdonságú, ESS1. Eredményüket valószínűségi módszerrel igazolták. A már említett dolgozatukban a $\Pi(q)$ projektív síkon $B(q - \sqrt{q})$ tulajdonságú B b.h. konstrukcióját is megadják, ESS2. Erdősék eredményét, ESS1-et, ABBOTT és LIU [1] $q = p^k, p > 2$ prím és $k > 0$ páratlan egész esetben tovább finomították, náluk már $c > \frac{2}{\log 2}$, AL1. Ezenkívül Abbott és Liu p prímszám esetén, p-edrendű projektív síkra $B(2h(p) + 3)$ tulajdonságú b.h. létezését igazolják, AL2. Itt $h(p)$ a p prímszám egymást követő kvadratikus nemmaradékai[2] által alkotott leghosszabb blokk hosszát jelöli. BURGESS [7] eredménye szerint $h(p) = O(p^{\frac{1}{4}+\varepsilon})$, ha azonban igaz az általánosított Riemann-hipotézis, akkor $h(p) = O(\log p^2)$. SZŐNYI [15], Abbott és Liu eredményéhez hasonlót igazolt a $PG(2, q), q \equiv 1$ vagy 3 (mod 4) síkok esetén, k darab parabola[3] uniójából $B = \bigcup_{i=1}^{k} P_{a_i}$ minimális blokkoló halmazt konstruált. A $q \equiv 1$ (mod 4) esetben a parabolák $a_1, a_2, \ldots, a_k$ paraméterei kielégítették a következő tulajdonságot: $a_i - a_j$ nem négyzeteleme a $GF(q)$ Galois testnek, bármely $i \neq j$, $i, j = 1, 2, \ldots, k$ esetén és az $\{a_1, \ldots, a_k\}$ maximális erre a tulajdonságra nézve. ABBOTT és LIU [1] illetve SZŐNYI [15, 16] a blokkoló halmazokat parabolák uniójaként adta meg. A parabolák kiválasztásához a mod p maradékosztálytest kvadratikus maradékait illetve nem-maradékait használták. Ekkor a B m.b.h. $B(2k + 1)$ tulajdonságú, SZ, [15].

[2]Ha $(a, m) = 1$ és az $x^2 \equiv a$ (mod m) kongruenciának nincs megoldása, akkor azt mondjuk, hogy az *a kvadratikus nemmaradék* modulo m.

[3]A $P_a := \{(x, y) \in GF(q)^2 \mid y = x^2 + a, a \in GF(q)\} \cup \{Y_\infty\} \subset \Pi$ ponthalmaz *parabola*, ahol Y_∞ az y-tengely ideális pontját jelöli.

SZŐNYI [16] cikkében egy olyan eredményt közöl, amely szerint egy parabolák uniójából álló m.b.h. maximális metszésszáma legalább $\log q$ nagyságrendű.

A következő eredmények prímhatványrendű Galois-síkok esetén biztosítják meglepően kis γ értékkel $B(\gamma)$ tulajdonságú b.h. létezését.

BRUEN és FISHER [6] a $PG(2, 3^s)$, $s \geq 2$ síkok esetén igazolták a $B(5)$ tulajdonságot. Eredményüket BOROS [2] általánosította $PG(2, p^s)$, $p > 2$ és $s \geq 2$ síkra, $B(p+2)$ tulajdonságú b.h.-t konstruálva. ILLÉS, SZŐNYI és WETTL [12] a következő eredményt igazolta a $p = 2$ esetre: ha r páros a $PG(2, 2^r)$ projektív sík $B(6)$ tulajdonságú, és ha r páratlan akkor $B(7)$ tulajdonságú, ISZW, [12].

Egyszerűen megfogalmazható annak szükséges feltétele, hogy egy B m.b.h. $B(\gamma+1)$ tulajdonságú legyen, [3]. Jelölje $y_1, y_2, \ldots, y_\gamma$ azon egyenesek számát, amelyek a B halmazból pontosan $1, 2, \ldots, \gamma$ pontot tartalmaznak és legyen $|B| = m$. Ekkor

$$\sum_{i=1}^{\gamma} y_i = q^2 + q + 1, \tag{2.1.}$$

$$\sum_{i=1}^{\gamma} i y_i = m(q+1), \tag{2.2.}$$

$$\sum_{i=2}^{\gamma} \binom{i}{2} y_i = \binom{m}{2}, \tag{2.3.}$$

$$q + \sqrt{q} + 1 \leq m \leq q\sqrt{q} + 1, \tag{2.4.}$$

$$y_1 \geq m, \tag{2.5.}$$

$$y_\gamma \geq 1, \text{ és} \tag{2.6.}$$

$$y_i \geq 0, \quad i = 2, 3, \ldots, \gamma - 1, \tag{2.7.}$$

ahol y_i, m, γ egész számok és q jelöli a projektív sík rendjét.

A (2.1.) egyenlet nem egyéb, mint a sík egyenesei darabszámára vonatkozó megkötés, (2.2.) az illeszkedések, míg (2.3.) a pontpárok-egyenesek illeszkedését számlálja le kétféleképpen. A (2.4.) egyenlőtlenség a Bruen és Fisher által adott korlát, (2.5.) pedig előírja legalább m érintő létezését (ami szükséges feltétele annak, hogy m.b.h.-t kapjunk, hiszen minden pontban kell, hogy legyen legalább egy érintő). A (2.6.) biztosítja legalább egy olyan egyenes létezését, amely γ pontot tartalmaz a B halmazból és adott γ és q esetén (2.7.) természetes megkötés a $2 \leq i \leq \gamma - 1$ pontban metszők számára. Legyen

$$F(q, \gamma) := \{(y_1, y_2, \ldots, y_\gamma, m) \mid (y_1, y_2, \ldots, y_\gamma, m) \text{kielégíti a (2.1.)–(2.7.) feltételeket}\},$$

a (2.1.)–(2.7.) feltételrendszer megengedett megoldásainak halmaza. Ekkor nyilvánvaló, hogy ha $F(q,\gamma) = \emptyset$ akkor a $\Pi(q)$ projektív síkon nem létezik $B(\gamma+1)$ tulajdonságú minimális blokkoló halmaz.

Ez a $\gamma = 3$ esetben a következő eredményre vezet

$$y_1 = \frac{m^2}{2} - m\left(2q + \frac{5}{2}\right) + 3(q^2+q+1),$$

$$y_2 = -m^2 + m(3q+4) - 3(q^2+q+1),$$

$$y_3 = \frac{m^2}{2} - m\left(q + \frac{3}{2}\right) + (q^2+q+1).$$

Könnyen megmutatható, hogy ha $q \geq 5$ akkor $y_2 < 0$ bármely m-re. A fennmaradó esetekben ($q = 3, 4$) a $\Pi(q)$ projektív sík $B(4)$ tulajdonságú [11]. Eredményünket a következő állításban foglalhatjuk össze:

2.1. ÁLLÍTÁS. *A $\Pi(q)$ projektív sík pontosan akkor $B(4)$ tulajdonságú, ha $q = 3$ vagy $q = 4$.*

A $\gamma = 4$ esetben két olyan bázisa lehet a (2.1.)–(2.3.) lineáris egyenletrendszernek, amelyek esetén y_1 és y_4 bázisváltozók. B' bázis esetén legyenek az $1, 2, 4$, míg a B'' bázis esetén az $1, 3, 4$ indexű változók a bázisváltozók. A B' bázist használva

$$y_1' = \frac{1}{3}m^2 - m\left(\frac{5}{3}q + 2\right) + \frac{8}{3}(q^2+q+1),$$

$$y_2' = -\frac{1}{2}m^2 + m\left(2q + \frac{3}{2}\right) - 2(q^2+q+1),$$

$$y_3' = 0,$$

$$y_4' = \frac{1}{6}m^2 - m\left(\frac{q}{3} + \frac{1}{2}\right) + \frac{1}{3}(q^2+q+1)$$

adódik. Elemi számolással megmutatható, hogy $y_1' \geq m$ teljesül bármely m-re, ha $q \geq 9$. Továbbá $y_2' \geq 0$ csak akkor teljesül, ha bármely q rend esetén

$$2q + \frac{5}{2} - \sqrt{6q + \frac{9}{4}} \leq m \leq 2q + \frac{5}{2} + \sqrt{6q + \frac{9}{4}}\,.$$

Végül $y_4' \geq 1$ igaz bármely m-re, ha $q \geq 3$. A B'' bázis esetén a (2.1.)–(2.3.) egyenletrendszer bázismegoldása

$$y_1'' = \frac{1}{6}m^2 - m\left(q + \frac{7}{6} + 2(q^2+q+1)\right),$$

$$y_2'' = 0,$$

$$y_3'' = -\frac{1}{2}m^2 + m\left(2q + \frac{3}{2}\right) - 2(q^2 + q + 1),$$

$$y_4'' = \frac{1}{3}m^2 - m\left(q + \frac{4}{3}\right) + (q^2 + q + 1)$$

adódik. Hasonlóan, mint a B' bázisnál $y_1'' \geq m$ bármely m-re, ha $q \geq 8$ teljesül. Továbbá $y_4'' \geq 1$, ha $q \geq 6$, míg az $y_3'' \geq 0$ teljesülésének ugyanaz a feltétele, mint az $y_2' \geq 0$ feltételnek. Az előzőekből a következő állítást nyerjük:

2.2. ÁLLÍTÁS. *Ha valamely B m.b.h. a $\Pi(q)$ projektív síkon $B(5)$ tulajdonságú, akkor*

$$2q + \frac{5}{2} - \sqrt{6q + \frac{9}{4}} \leq m \leq 2q + \frac{5}{2} + \sqrt{6q + \frac{9}{4}},$$

ahol $|\ B\ | = m$ és $q \geq 9$.

A 2.2. Állítás bizonyítása során használt módszerrel könnyen belátható a

2.3. ÁLLÍTÁS. *Ha valamely B m.b.h. a $\Pi(q)$ projektív síkon $B(\gamma + 1)$ tulajdonságú, akkor*

$$\frac{1}{2}\left(1 + \gamma(q+1) - \sqrt{f(\gamma, q)}\right) \leq m \leq \frac{1}{2}\left(1 + \gamma(q+1) + \sqrt{f(\gamma, q)}\right),$$

ahol $q \geq 8$, $\gamma > 3$, $|B| = m$ és $f(\gamma, q) = (\gamma(q+1) - 1)^2 - 4\gamma q^2$.

Könnyen belátható, hogy $\gamma < 4$ és $q \geq 8$ esetén, $f(\gamma, q) < 0$, azaz $m \in \mathbb{N}$ miatt 2.3. Állítás nem alkalmazható.

3. Az Erdős-probléma matematikai programozási modellje

Ebben a fejezetben ismertetjük az Erdős-problémának megfelelő egészértékű lineáris programozási feladatot (ELP), amely lehetővé teszi a probléma egy számítógépes megközelítését is.

Adott rendű projektív síkon, az Erdős-probléma a következő egészértékű lineáris programozási feladatként irható fel:

$$(ELP) \quad \begin{cases} \min \alpha \\ e \leq Lx \leq \alpha e \\ 2 \leq \alpha \leq q \\ x \in \{0,1\}^n \end{cases}$$

$e = (1, 1, \ldots, 1)$ n dimenziós vektor, α egész szám, $n = q^2 + q + 1$, ahol q a projektív sík rendje és L jelöli a sík egyenes-pont illeszkedési mátrixát. Az L mátrix

elkészítésének módja megtalálható Kárteszi könyvében [13] illetve Rényi cikkében [14] is.

Ha $\alpha = q$ akkor a feltételrendszert kielégítő vektorok pontosan a blokkoló halmazoknak megfelelő karakterisztikus vektorok. A q-tól különböző α-ra pedig azon blokkoló halmazoknak megfelelő karakterisztikus vektorokat kapjuk, amelyek $B(\alpha+1)$ tulajdonságúak. Az ELP feladat minden optimális megoldása egy legkisebb metszésszámú blokkoló halmaz a q-adrendű projektív síkon és az optimális α a sík maximális metszésszáma. Az Erdős-problémát modellező ELP egy NP-teljes problémára vezet. A feladat LP-relaxáltjának triviálisan adódó megoldása az $\alpha = 2$ és $x = (\frac{2}{q+1}, \frac{2}{q+1}, \ldots, \frac{2}{q+1})$ vektor.[4] Mivel az LP-relaxáció eredményéből kiindulva az ELP feladat optimális 0-1 megoldását előállítani egzakt (vágás illetve korlátozás és szétválasztás típusú [17]) módszerekkel még egészen kis rendek esetén is reménytelen vállalkozás lenne, ezért a feladat struktúráját kihasználva mohó algoritmust fogalmaztunk meg, amely az ELP szuboptimális megoldását adja. Algoritmusunkkal egyben egy új eljárást adunk számos γ értékhez tartozó $B(\gamma)$ tulajdonságú (minimális) blokkoló halmazok konstrukciójára is.

A továbbiakban célunk, egy olyan algoritmus megfogalmazása, amely egy adott blokkoló halmazból (az ELP egy megengedett megoldása, az $\alpha = q-1$ esetén) kiindulva, javító lépések sorozatával, olyan blokkoló halmazt állít elő, amelynél $\alpha < q-1$ és algoritmusunkkal tovább nem csökkenthető az α értéke. Az algoritmus általános lépése két fázisból áll: *törlés és bővítés*. Először a blokkoló halmaz néhány pontját töröljük azzal a céllal, hogy az α értéke csökkenjen. A törlés után előálló ponthalmaz általában nem blokkoló halmaz, azaz létezik olyan egyenes, amelynek a törléssel kapott halmazzal nincsen közös pontja (kitérő egyenesek). Tehát sérül az ELP modell feltételeiben adott alsókorlát megkötés. A bővítés fázisban a kitérő egyeneseket igyekszünk pontokkal lefogni, úgy, hogy α értéke (maximális metszésszám) ne növekedjen.

Az algoritmus leírásakor szükségünk lesz a következő jelölésekre:

- B_0 az induláskor adott blokkoló halmaz;
- B_k jelöli a k. lépés kezdetén a blokkoló halmazt, míg a $\hat{B}_k$ a k. lépés során a B_k-ból törléssel illetve bővítéssel nyert ponthalmaz;
- $\alpha_k = \max_i |l_i \cap B_k|$, a B_k b.h. maximális metszésszáma;
- $S_j^k := \{i : j = |l_i \cap \hat{B}_k|\}$; ekkor az S_1^k jelöli a k. lépésben az érintő, míg az S_0^k a kitérő egyenesek és S_{α_k} a maximális számú pontot tartalmazó egyenesek indexhalmazát;

[4] Alkalmunk volt a delfti Műszaki Egyetem *Convex C3820* miniszuperszámítógépén (a műveleti sebessége 480 Mflops) a CPLEX [8] programcsomaggal tesztelni modellünket. A $q = 7$ esetén a CPLEX által felhasznált CPU idő 896.49 másodperc volt. A bináris fában levő élek száma elérte az alapértelmezésben adott 10000-es felsőkorlátot, míg az összes LP-relaxált megoldása során elvégzett iterációk száma 297655 volt. A CPLEX programcsomaggal négy volt a legkisebb célfüggvényérték, amelyre megengedett megoldást előállítottunk a $PG(2,7)$ projektív síkon, azaz a sík $B(5)$ tulajdonságát sikerült megmutatni. A CPLEX programcsomag alapértelmezésben adott paraméterei mellett már a $q = 7$ esetén sem adott optimális megoldást.

- $M := \{P \in \hat{B}_k : P \in l_i \text{ és } i \in S_{\alpha_k}\}$, azon pontok halmaza $\hat{B}_k$-ból, amelyek törlése esetén valamely egyenesen a $\hat{B}_k$ pontjainak a száma a maximumról eggyel csökken;
- $w : M \to \mathbb{N}$, $w(P) := |\{j \in S_{\alpha_k} : P \in l_j\}|$ függvény méri azt, hogy egy pont törlése esetén hány egyenes metszésszáma csökken le eggyel a maximumról;
- $v : M \to \mathbb{N}$, $v(P) := |\{j \in S_1 : P \in l_j\}|$ függvény azt mutatja meg, hogy hány érintő válik kitérővé a P pont törlése esetén;
- $T := \{P \in M : v(P) = 0\}$ az M halmaz olyan részhalmaza, amely tetszőleges pontját törölve egyetlen érintő sem válik kitérővé;
- $L := \{P \in M : v(P) = \min\limits_{R \in M} v(R)\}$ az M halmaz olyan részhalmaza, amely tetszőleges pontját törölve a legkevesebb érintő válik kitérővé;
- $N := \Pi \setminus (\hat{B}_k \bigcup\limits_{i \in S_{\alpha_k}} \bigcup\limits_{P \in l_i} P)$ halmaz a sík azon pontjaiból áll, amelyekkel a kitérő egyenesek lefoghatók, anélkül, hogy bármelyik egyenes metszésszáma újra α_k-ra növekedne;
- $u : N \to \mathbb{N}, u(P) := |\{j \in S_0 : P \in l_j\}|$ függvény méri az N halmazba tartozó P pont lefogó képességét.

Induló blokkoló halmaz konstrukciójára számos példa található [11]-ben. Feltehetjük, hogy a B_0 blokkoló halmazt három egyenes pontjaiból a következő módon állítottuk elő: $B_0 := \{P \in \Pi \mid P \in l_i,\ i = 1,2,3\} \setminus \{P_{12}, P_{13}, P_{23}\}$, ahol $P_{12} = l_1 \cap l_2$, $P_{13} = l_1 \cap l_3$ és $P_{23} = l_2 \cap l_3$. Könnyen belátható, hogy B_0 m.b.h.

3.1. ALGORITMUS. Legyen adott egy induló B_0 blokkoló halmaz a hozzátartozó maximális metszésszámmal.

***k*. lépés:** TÖRLÉS

Legyen $\hat{B}_k = B_k$ határozzuk meg az α_k értéket és az S_{α_k} halmazt.

(1) Ha $S_{\alpha_k} = \emptyset$

akkor menjünk a BŐVÍTÉSre

különben határozzuk meg az S_1 és M halmazokat, számoljuk ki $\forall P \in M$ pontra a $w(P)$ és $v(P)$ értékeket, adjuk meg a T halmazt.

Ha $T \neq \emptyset$

akkor válasszuk ki a $Q \in T$ pontot, amelyre $w(Q) = \max\limits_{P \in T} w(P)$.

különben határozzuk meg az L halmazt és a $Q \in L$ pontot, amelyre $w(Q) = \max\limits_{P \in L} w(P)$.

Legyen $\hat{B}_k = \hat{B}_k \setminus \{Q\}$, módosítsuk az S_{α_k} halmazt és menjünk (1)-re.

BŐVÍTÉS

Határozzuk meg az S_0 és S_{α_k-1} halmazokat.

(2) Ha $S_0 = \emptyset$

akkor $B_{k+1} := \hat{B}_k, \alpha_{k+1} := \alpha_k - 1$ és $k := k+1$.

különben határozzuk meg az N halmazt.

Ha $N = \emptyset$

akkor STOP. B_k és α_k a megoldás.

különben számoljuk ki $\forall P \in N$ pontra $u(P)$ értékeket.

Ha $|\bigcup_{P\in N}\{i \in S_0 : P \in l_i\}| < |S_0|$

akkor STOP. B_k és α_k a megoldás.

különben válasszunk ki egy $Q \in N$ pontot, amelyre

$$u(Q) = \max_{P\in N} u(P) \quad \text{és} \quad \hat{B}_k = \hat{B}_k \cup \{Q\}.$$

Határozzuk meg az S_0 és S_{α_k-1} halmazokat és menjünk a (2)-re.

Az algoritmus *megállási kritériuma* jogosságát a következő lemma igazolja.

3.2. LEMMA. *Tegyük fel, hogy adott $\hat{B}_k$, a hozzá tartozó N, S_0, S_{α_k} halmazokkal. Ha $|\bigcup_{P\in N}\{i \in S_0 : P \in l_i\}| < |S_0|$ akkor $\hat{B}_k$ nem egészíthető ki blokkoló halmazzá a 3.1. algoritmussal.*

Bizonyítás. Az N halmaz lefogóképességét az $r := |\bigcup_{P\in N}\{i \in S_0 : P \in l_i\}|$ szám adja meg Ezért ha az $r < |S_0|$ igaz, akkor ez azt jelenti, hogy a teljes N halmazzal kibővített $\hat{B}_k$, sem lenne blokkoló halmaz vagyis maradna kitérő egyenes.

A 3.1. algoritmussal a B_0 induló blokkoló halmazból nyert minimális α értéket a q-adrendű projektív sík esetén jelöljük $\alpha_G(q)$-val.

3.3. ÁLLÍTÁS. *Bármely $\alpha \in [\alpha_G(q), q]$ egész értékhez a q-adrendű projektív síkon konstruálható $B(\alpha+1)$ tulajdonságú blokkoló halmaz.*

Bizonyítás. Az $\alpha = q$ esetén ismert a következő m.b.h. [11]

$$B := \{P \in \Pi \mid P \in l_i,\ i = 1,2\} \setminus \{Q, R\} \cup \{U\},$$

ahol $Q \in l_1$, de $Q \notin l_2$ és $R \notin l_1$, de $R \in l_2$, továbbá $U \in l_{QR} \setminus \{Q, R\}$. Ekkor $|l_1 \cap B| = |l_2 \cap B| = q$.

Az $\alpha = q - 1$ esetre megfelel a B_0 m.b.h.

Az $\alpha_G(q)$ meghatározásából látható, ez az a minimális érték, amelyre a 3.1. algoritmussal a B_0 blokkoló halmazból kiindulva $\alpha_G(q)$ maximális metszésszámú b.h. konstruálható. Azt kell belátnunk, hogy bármely BŐVÍTÉS fázis végén b.h.-t nyerünk mindaddig, amíg $\alpha \geq \alpha_G(q)$ és a 3.1. algoritmus bármely lépésében α értéke pontosan eggyel csökken. Az előző állítás első fele nyilvánvaló. Mivel az S_α halmazt a TÖRLÉS fázisban minden pont elhagyása után újra számoljuk, ezért az

utolsó törlendő Q pontot, olyan egyenesről választottuk ki, amely metszésszáma α volt. Így a Q pont törlése után legalább egy egyenesnek $\alpha - 1$ lesz a metszésszáma. A BŐVÍTÉS fázis a maximális metszésszámot nem módosítja.

A 3.1. algoritmus TÖRLÉS részében ki kell jelölnünk egy $l_i, i \in S_{\alpha_k}$ egyenesről egy $Q \in \hat{B}_k$ pontot törlésre. Ehhez határozzuk meg az M halmazt. A pont kiválasztása az alábbi stratégiák valamelyikével történhet.

1. határozzuk meg az $L := \{P \in M \mid v(P) = \min_{R \in M} v(R)\}$ halmazt és legyen $Q \in L$, olyan pont, amelyre $w(Q) = \max_{R \in L} w(R)$ igaz;
2. határozzuk meg az $L := \{P \in M \mid v(P) = \min_{R \in M} v(R)\}$ halmazt és legyen $Q \in L$, olyan pont, amelyre $w(Q) = \min_{R \in L} w(R)$ igaz;
3. határozzuk meg az $L := \{P \in M \mid v(P) = \max_{R \in M} v(R)\}$ halmazt és legyen $Q \in L$, olyan pont, amelyre $w(Q) = \min_{R \in L} w(R)$ igaz;
4. határozzuk meg az $L := \{P \in M \mid v(P) = \max_{R \in M} v(R)\}$ halmazt és legyen $Q \in L$, olyan pont, amelyre $w(Q) = \max_{R \in L} w(R)$ igaz.

Mind a négy stratégiát különböző induló blokkoló halmazok esetén teszteltük. Tapasztalataink szerint mindegyikkel előállítható $B(c \log q)$ tulajdonságú b.h. A legjobb eredményeket akkor kaptuk, ha az első stratégiát használtuk és az induló blokkoló halmazunkat a három egyenesből konstruáltuk. Ezekben az esetekben $\alpha_G(q) \in (\log q, 2 \log q)$ adódott.

4. Számítógépes eredmények elemzése

A Galois-sík illeszkedés tábláját generáló és az ELP feladatot a mohó algoritmussal szuboptimálisan megoldó programot Turbo Pascalban kódoltuk PC-AT 386/25 Mhz-es, 2 Mbyte RAM memóriával rendelkező gépen. Programunk futási ideje a rend növekedésével jelentősen növekedett. Így nagyobb rendű geometriákra a futási idő több napot vett igénybe. A $q = 89$ renddel bezárólag minden prímre sikerült lefuttatnunk a programunkat. Ezenkívül az algoritmusunkkal, még a $q = 8, 9, 16$ prímhatványokra is előállítottuk az ELP feladat szuboptimális megoldását.

Az 1. táblázatban különböző szerzők becsléseit láthatjuk a $B(\gamma)$ tulajdonság γ értékére. A rövidítések azonosak az eredmények ismertetése után használtakkal, míg BI a programunk eredményeit jelöli.

A BF, B, ISZW, AL2 oszlopokban található eredmények bizonyításában megtalálhatjuk a kívánt tulajdonságú b.h. konstrukcióját is. Mivel BOROS [2] illetve ILLÉS, SZŐNYI és WETTL [12] eredményei prímhatványokra vonatkoznak, ezért algoritmusunk eredményeivel közvetlenül nem tudjuk összehasonlítani azokat (kivétel a 8, 9 és 16-odrendű geometria). Figyelembe véve eredményeinket, várhatóan algoritmusunk rosszabb értékeket adna, mint az említett becsléseknél lévő, hiszen azok a q-adrendű ($q = p^k$) Galois-testek algebrai tulajdonságait jelentősen kihasználják.

q	B	ISZW	AL2	SZ	AL1	ESS1	ESS2	BI
7	–	–	7	–	–	10	–	5
8	–	7	–	–	–	11	6	5
9	5	–	–	–	–	11	6	6
11	–	–	9	–	7	13	8	6
13	–	–	11	7	8	14	10	6
16	–	6	–	–	–	15	12	6
17	–	–	9	7	9	15	13	7
19	–	–	11	–	9	16	15	7
23	–	–	11	–	9	17	19	7
25	7	–	–	–	–	17	20	–
27	5	–	–	–	10	17	22	–
29	–	–	9	9	10	18	24	8
31	–	–	11	–	10	18	26	8
32	–	7	–	–	–	18	27	–
37	–	–	11	9	11	19	31	8
41	–	–	13	11	11	20	35	8
43	–	–	13	–	11	20	37	8
47	–	–	11	–	11	21	41	8
49	9	–	–	–	–	21	42	–
53	–	–	15	11	11	21	45	9
59	–	–	13	–	12	22	51	9
61	–	–	15	11	12	22	53	9
64	–	6	–	–	–	22	56	–
67	–	–	15	–	12	22	58	9
71	–	–	15	–	12	23	62	9
73	–	–	11	11	12	23	64	9
79	–	–	15	–	13	23	70	9
83	–	–	17	–	13	24	74	9
89	–	–	15	11	13	24	80	9

1. Táblázat: A γ értékei különböző szerzők nyomán.

(Ez alól kivételt képez a $q = 8$ eset, amelynél mohó módszerünkkel a legkisebb ismert metszésszámot nyertünk.) Az AL2 eredménynél kiszámítottuk pontosan a $h(p)$ függvény értékét.

Az AL1, ESS1, ESS2 oszlopokban levő eredményeknél algoritmusunk jobbat szolgáltatott a megvizsgált rendeken. A γ értékek kiszámításánál a c értékre a szerzők által adott alsó korlátot használtuk (ami egyértelműen csökkentette az értékeket). Az eredményül kapott valós számból felsőegészrész képzéssel nyertük a γ értéket.

Az 1. táblázat elemzésekor, prímrendek esetén, érdekes információkat az AL2, SZ, AL1 és BI oszlopok hordoznak. ABBOTT és LIU [1] illetve SZŐNYI [15,16] eredményeinél a m.b.h. parabolák egyesítéséből adódnak, míg a mi számítógépes

eredményeink esetén nem (feltétlenül). Mégis mindkét esetben a γ értéke nagyságrendileg $\log q$ és az m.b.h. mérete $q \log q$.

Algoritmusunkkal néhány rend esetén, az ismertektől eltérő $q \log q$ méretű m.b.h.-t állítottunk elő.

A véges projektív síkok elméletének érdekes megoldatlan kérdése az is, hogy létezik-e $q \log q$-nál nagyobb méretű m.b.h., ha a q prímszám.

Észrevehetjük, hogy a blokkoló halmazaink számossága a rend növekedésével Bruen alsó korlátjához közelít. A blokkoló halmazt lefedő egyenesek száma a geometria rendjével nő. Ugyanezt állapíthatjuk meg a b.h. és a sík metszési számáról is.

Ha azonban a maximális metszésszám növekszik a rend növelésével, akkor a maximális metszésszámmal rendelkező egyenesek száma csökken. Mindez a 2. táblázatban látható.

q	változók sz.	alsókorlát	$\lvert B\rvert$	felsőkorlát	α	α-egyenesek sz.	t
7	58	11	15	19	4	12	5
8	74	12	18	23	4	19	5
9	92	14	19	28	5	10	6
11	134	16	25	37	5	12	7
13	184	18	30	47	5	18	9
16	274	22	37	65	6	12	9
17	308	23	40	71	6	14	10
19	382	25	46	83	6	17	11
23	554	29	59	111	6	34	14
29	872	36	82	157	7	23	18
31	994	38	88	173	7	22	20
37	1408	45	107	226	7	55	23
41	1724	49	128	263	7	87	25
43	1894	51	132	282	7	85	28
47	2258	55	148	323	7	114	31
53	2864	62	161	386	8	52	32
59	3542	68	184	454	8	66	35
61	3784	70	190	477	8	68	35
67	4558	77	217	549	8	113	40
71	5114	81	234	599	8	121	43
73	5404	83	237	624	8	120	45
79	6322	89	265	703	8	176	45
83	6974	94	282	757	8	216	49
89	8012	100	305	840	8	239	54

2. Táblázat

A 2. táblázat BI oszlopában ismertetett szuboptimális megoldások további elemzését tartalmazza a 3. táblázat, amelyben eredményeinket a (2.1.)-(2.7.) feltételrendszer alapján mutatjuk be.

A CPLEX programcsomaggal HP 9000/720-as számítógéppel folytatott kísérleteinkről a 4. táblázatban számolunk be.

q	$\|B\|$	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8
8	18	29	18	7	19	–	–	–	–
13	30	70	45	30	20	18	–	–	–
16	37	98	76	52	24	11	12	–	–
17	40	104	95	47	34	13	14	–	–
19	46	128	104	68	42	22	17	–	–
23	59	166	145	112	60	36	34	–	–
29	82	199	233	184	123	64	45	23	–
31	88	255	220	196	163	95	42	22	–
37	107	310	368	294	201	125	54	55	–
41	128	312	422	358	280	167	97	87	–
43	132	372	447	404	302	171	112	85	–
47	148	429	505	474	357	251	127	114	–
53	161	570	683	629	454	271	144	60	52
59	184	662	808	815	542	365	185	98	66
61	190	729	856	821	620	369	233	87	68
67	217	769	1011	1007	764	505	257	131	113
71	234	822	1126	1109	874	559	326	176	121
73	237	914	1203	1184	885	582	349	166	120
79	265	982	1363	1320	1104	753	418	205	176
89	305	1148	1693	1740	1383	913	578	317	239

3. Táblázat:

A q a véges geometria rendje, $|B|$ a blokkoló halmaz számossága és $y_i, (i = 1, 2, ..., 8)$ az i-metszők száma.

A 4. táblázatban összefoglalt eredmények azt támasztják alá, amit a 3. fejezetben már említettünk, azaz nem alkalmasak az általános, egzakt algoritmusok az ilyen típusú feladatok hatékony megoldására. Mindezek alapján úgy gondoljuk, hogy numerikus eredményeink további javításához az ELP modellt kellene finomítani bevezetve a 2.3. Állításból adódó alsó- és felsőkorlátokat az $e^T x = m$ feltételre illetve a 2. táblázatból az α-ra nyerhető felsőkorlátokkal módosítanánk a modellt. A legkecsegtetőbb mégis a (2.1.)–(2.7.) rendszer megoldásainak a geometriai vizsgálata tűnik.

q	α	k	i	t
11	5	3730	233380	1512.17
13	5	9704	982028	4616.41
17	6	14544	4020982	68954.23
19	6	4877	1132761	22858.98
29	–	–	–	7210879

4. Táblázat:

A q a véges geometria rendje,
α a legkisebb metszésszám, amelyre sikerült $B(\alpha+1)$ típusú b.h.-t előállítani,
i az összes iteráció száma, míg t a felhasznált idő másodpercben kifejezve.

A $q = 17$ esetén nem állt rendelkezésre elég memória, ezért a futás leállt. A számítógép osztott üzemmódban működött. Az $\alpha = 6$ volt a legkisebb célfüggvényérték, amelyhez talált megengedett megoldást a CPLEX programcsomag. Ezzel szemben a $q = 29$ esetén a CPLEX programcsomaggal, alapértelmezésben (lásd [8]) nem sikerült 0-1 megoldást találni!

Köszönetnyilvánítás. A második szerző köszönetet mond A MAGYAR TUDOMÁNYÉRT alapítvány Kuratóriumának az 1993. március–december időszakban folyósított ösztöndíjért. A kutatást az OTKA T 014302 számú pályázata támogatta.

A delfti Műszaki Egyetem Sztochasztika, Statisztika és Operációkutatási Tanszékének köszönjük, hogy lehetővé tette számunkra a CPLEX programcsomag használatát a Convex C3820 és HP 9000/720 számítógépeken.

A szerzők megköszönik SZŐNYI TAMÁS hasznos tanácsait, amelyeket a dolgozat új változatának elkészítése során adott.

IRODALOM

[1] ABBOTT, H.L., LIU, A., „Property $B(s)$ and projective planes", *Ars Combinatoria* **20** (1985), 217–220.
[2] BOROS E., „$PG(2,q), p > 2$ has property $B(p+2)$", *Ars Combinatoria* **25** (1988), 111–114.
[3] BOROS E., Szóbeli közlés, 1992. október.
[4] BRUEN, A.A., „Baer subplanes and blocking sets", *Bull. Amer. Math. Soc.* **76** (1970), 342–344.
[5] BRUEN, A..A., THAS, J.A., „Blocking sets", *Geom. Dedicata* **6** (1977), 193–203.
[6] BRUEN, A.A., FISCHER, J.C., „Blocking sets and complete k-arcs", *Pacific Journal of Math.* **53** (1974), 73–84.
[7] BURGESS, D., „The distribition of quadratic residues and non-residues", *Mathematica* **4** (1957), 106–112.
[8] *CPLEX manual version 2.1.*, CPLEX Optimization, Inc. (1993).
[9] ERDŐS P., SILVERMANN, R., STEIN, A., „Intersection properties of families containig sets of nearly the same size", *Ars Combinatoria* **15** (1983), 247–259.
[10] FÜREDI Z., „Matchings and covers in hypergraphs", *Graphs and Combin.* **4** (1988), 115–206.
[11] ILLÉS T., *Maximális erős reprezentáns rendszerek és minimális blokkoló halmazok vizsgálata a véges projektív síkon*, Egyetemi doktori értekezés (Budapest, 1989).
[12] ILLÉS T., SZŐNYI T., WETTL F., „Blocking sets and maximal strong representative systems in finite projective planes", *Mitt. Math. Sem. Giessen* **201** (1991), 97–107.
[13] KÁRTESZI F., *Bevezetés a véges geometriákba* (Akadémiai Kiadó, Budapest, 1972).
[14] RÉNYI A., „Véges geometrák kombinatorikai alkalmazásai I", *Matematikai Lapok* **17** (1966), 33–76.

[15] SZÖNYI T., „Note on the existance of large minimal blocking sets in Galois planes", *Combinatorica* **12** (1992), 227–235.
[16] SZÖNYI T., „Blocking sets in finite planes and spaces", *Ratio Mathematica* **5** (1992), 93–106.
[17] VIZVÁRI B., *Egészértékű programozás* (Tankönyvkiadó, Budapest, 1989).

(Beérkezett: 1993. november 8.)

(Átdolgozva beérkezett: 1994. január 5.)

BÉRES LÁSZLÓ, PÉNZÜGYI ÉS SZÁMVITELI FŐISKOLA
SZÁMÍTÁSTECHNIKA TANSZÉK
BUZOGÁNY U. 10–12, 1149 BUDAPEST
ILLÉS TIBOR, ELTE TTK, OPERÁCIÓKUTATÁSI TANSZÉK
MÚZEUM KRT. 6–8., 1088 BUDAPEST
E-mail: illes@konig.elte.hu

COMPUTATIONAL INVESTIGATION OF THE COVERING NUMBER OF FINITE PROJECTIVE PLANES WITH SMALL ORDER

LÁSZLÓ BÉRES AND TIBOR ILLÉS

Our paper deals with P. Erdős' problem related to the covering number of blocking sets of finite projective planes. An integer linear programming (ILP) formulation of Erdős' problem is introduced for projective planes of given orders. The mathematical programming based approach is new in the area of finite projective planes. Considering the complexity of exact solution methods for integer programming problems and the available computational capacity a greedy algorithm for the constructed ILP is presented and implemented. We produced suboptimal solution for ILP (denote by $\alpha_G(q)$) for $7 \leq q \leq 89$, q prime and for $q = 8, 9, 16$. Our greedy algorithm is the only known method which for all integer α from a given interval (in our case $\alpha \in [\alpha_G(q), q]$) produces a blocking set with property $B(\alpha + 1)$. According to our computational results the approximated value of $\alpha_G(q)$ is $c \log q$, where $1 \leq c \leq 2$ holds.

If $q = 89$ then the ILP contains 8012 integer variables and 8011 range constraints. All the variables, except one, are binary. Finally, the computational results and the comparison with the known theoretical results are presented.

A kiadásért felelős a BJMT főtitkára
Szedte a KLTE Informatikai és Számítóközpont Kiadvány Szerkesztő Csoportja

Nyomta az MSZH Nyomda és Kiadó Kft., Budapest, 97.087
Felelős vezető: Nagy László

Budapest, 1997
Megjelent 18 (A/5) ív terjedelemben
250 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 sorszámozással 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 lehetséges, 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 a társszerzők esetén az első szerző neve szerint alfabetikus sorrendben úgy, hogy a cirill betűs szerzők nevét a Methematical 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-ás 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 researchon 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., *Mathods of Feasible Directions* (Elsevier 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: 425 Ft + ÁFA

TARTALOMJEGYZÉK

Klafszky Emil és *Hajdu Miklós*, Egy algoritmus a költségtervezési feladat megoldására tevékenység-élű tervütem hálón (CPM/cost feladat) 211
Hajdu Miklós, A költségtervezési feladat megoldása különböző függőségi kapcsolatokat tartalmazó tevékenység csomópontú háló esetén (MPM/cost feladat) 225
Vattai Zoltán András, Algoritmus az $F2 \mid \text{overlap} \mid C_{\max}$ megoldására 241
Csébfalvi A., Szakaszonként folytonosan differenciálható energia függvénnyel jellemezhető tartószerkezetek stabilitásvizsgálata 259
Bodócs László, Egy stabilitási feltétel a kvázi-izometrikus konjugált párok módszerére 267
Molnár Sándor, *Szigeti Ferenc* és *Vera Carmen E.*, Kalman-féle rangfeltételek az időtől függő lineáris rendszerekre 279
Molnár Sándor és *Szidarovszky Ferenc*, Dinamikus folyamatok konvergenciája és stabilitása 287
Vásárhelyiné Szabó Anna, A szerkezetanalízis egy matematikai modellje lokális egyensúlyi folyamatok esetén 293
Bagyinszki János, Véges halmazon értelmezett függvények *pr*-maximális és *pr*-teljes klónjai . 321
Kozák Imre, Megjegyzések Lámer G.: A szükséges és elégséges összeférhetőségi peremfeltételek meghatározása című cikkéhez 329
Benczúr András, *B. Novák Ágnes* és *Révész Z. Péter*, Klasszikus és súlyozott tudásbázisok transzformációi 347
Palágyi Kálmán, Lokális párhuzamos algoritmus bináris képek zajszűrésére 373
Béres László és *Illés Tibor*, Kis rendű projektív síkok metszésszámának számítógépes vizsgálata 397

INDEX

E. Klafszky and *M. Hajdu*, A new CPM time-cost trade-offs algorithm 211
M. Hajdu, An algorithm to solve the time-cost trade-offs problem in precedence diagramming . 225
Z. A. Vattai, Algorithm for solving $F2 \mid \text{overlap} \mid C_{\max}$ problem 241
A. Csébfalvi, Stability analysis of systems characterised by nonsmooth energy function 259
L. Bodócs, A stability condition for the method of quasi-isometric conjugate pairs 267
S. Molnár, *F. Szigeti* and *C. E. Vera*, Kalman's rank conditions for time dependent linear systems 279
S. Molnár and *F. Szidarovszky*, Convergence and Stability of Dynamic Processes 287
Anna Vásárhelyiné Szabó, Mathematical model of analysis of structures in the case of local equilibrium processes 293
J. Bagyinszki, *pr*-maximal and *pr*-complete clones of functions defined on a finite set 321
Imre Kozák, Remarks on the paper "Determination of the necessary and sufficient compatibility conditions on the boundary" written by G. Lámer 329
András Benczúr, *Ágnes B. Novák* and *Z. Péter Révész*, Propositional and weighted knowledge base transformations 347
Kálmán Palágyi, A Local Parallel Noise Reduction Algorithm for Binary Images 373
László Béres and *Tibor Illés*, Computational investigation of the covering number of finite projective planes with small order 397

ALKALMAZOTT MATEMATIKAI LAPOK

A MAGYAR TUDOMÁNYOS AKADÉMIA MATEMATIKAI ÉS FIZIKAI TUDOMÁNYOK OSZTÁLYÁNAK KÖZLEMÉNYEI

XVII. KÖTET

BJMT, BUDAPEST
1993

TARTALOMJEGYZÉK

Bagyinszki János, Véges halmazon értelmezett függvények *pr*-maximális és *pr*-teljes klónjai . 321
Bálint Erzsébet és *Deák István,* Párhuzamos számítógépek: optimalizálási programok 1
Benczúr András, B. Novák Ágnes és *Révész Z. Péter,* Klasszikus és súlyozott tudásbázisok transzformációi 347
Béres László és Illés Tibor, Kis rendű projektív síkok metszésszámának számítógépes vizsgálata 397
B. Novák Ágnes, Révész Z. Péter és *Benczúr András,* Klasszikus és súlyozott tudásbázisok transzformációi 347
Bodócs László, Egy stabilitási feltétel a kvázi-izometrikus konjugált párok módszerére 267
Csébfalvi A., Nemlineáris útkövető módszer tartószerkezetek stabilitásvizsgálatára, I. Reguláris pontok 57
Csébfalvi A., Nemlineáris útkövető módszer tartószerkezetek stabilitásvizsgálatára, II. Elágazási és határpontok 71
Csébfalvi A., Szakaszonként folytonosan differenciálható energia függvénnyel jellemezhető tartószerkezetek stabilitásvizsgálata 259
Csendes Tibor, Egy intervallum-aritmetikán alapuló algoritmus a szinthalmazok korlátainak megkeresésére 19
Deák István és *Bálint Erzsébet,* Párhuzamos számítógépek: optimalizálási programok 1
Faragó István, Haroten Hariton, Komáromi Nándor és *Pfeil Tamás,* A hővezetési egyenlet és numerikus megoldásának kvalitatív tulajdonságai, I. Az elsőfokú közelítések nemnegativitása 101
Faragó István, Haroten Hariton, Komáromi Nándor és *Pfeil Tamás,* A hővezetési egyenlet és numerikus megoldásának kvalitatív tulajdonságai, II. A másodfokú közelítés nemnegativitása, a maximum elv és az oszcillációmentesség 123
Farkas Henrik és *Simon L. Péter,* Polinomok gyökstruktúrájának vizsgálata a parametrikus reprezentáció módszerével 41
Hajdu Miklós, A költségtervezési feladat megoldása különböző függőségi kapcsolatokat tartalmazó tevékenység csomópontú háló esetén (MPM/cost feladat) 225
Hajdu Miklós és *Klafszky Emil,* Egy algoritmus a költségtervezési feladat megoldására tevékenység-élű tervütem hálón (CPM/cost feladat) 211
Haroten Hariton, Komáromi Nándor, Pfeil Tamás és *Faragó István,* A hővezetési egyenlet és numerikus megoldásának kvalitatív tulajdonságai, I. Az elsőfokú közelítések nemnegativitása 101
Haroten Hariton, Komáromi Nándor, Pfeil Tamás és *Faragó István,* A hővezetési egyenlet és numerikus megoldásának kvalitatív tulajdonságai, II. A másodfokú közelítés nemnegativitása, a maximum elv és az oszcillációmentesség 123
Illés Tibor és *Béres László,* Kis rendű projektív síkok metszésszámának számítógépes vizsgálata 397
Klafszky Emil és *Hajdu Miklós,* Egy algoritmus a költségtervezési feladat megoldására tevékenység-élű tervütem hálón (CPM/cost feladat) 211
Komáromi Nándor, Pfeil Tamás, Faragó István és *Haroten Hariton,* A hővezetési egyenlet és numerikus megoldásának kvalitatív tulajdonságai, I. Az elsőfokú közelítések nemnegativitása 101
Komáromi Nándor, Pfeil Tamás, Faragó István és *Haroten Hariton,* A hővezetési egyenlet és numerikus megoldásának kvalitatív tulajdonságai, II. A másodfokú közelítés nemnegativitása, a maximum elv és az oszcillációmentesség 123

Kozák Imre, Megjegyzések Lámer G.: A szükséges és elégséges összeférhetőségi peremfeltételek meghatározása című cikkéhez 329
Mészáros Csaba, Az affin skálázási algoritmus módosításairól 185
Mihálykó Csaba, A golyósmalmi őrlemény sűrűségfüggvényére felírt integro-differenciálegyenlet megoldhatósága és a megoldás speciális tulajdonságai 171
Molnár Sándor és *Szidarovszky Ferenc,* Dinamikus folyamatok konvergenciája és stabilitása 287
Molnár Sándor, Szigeti Ferenc és *Vera Carmen E.,* Kalman-féle rangfeltételek az időtől függő lineáris rendszerekre 279
Nagy Tamás, A szállítási feladat sztochasztikus variánsai 143
Németh Géza, Sorok a Mathieu függvények sajátértékeinek kiszámításához 195
Palágyi Kálmán, Lokális párhuzamos algoritmus bináris képek zajszűrésére 373
Pfeil Tamás, Faragó István, Haroten Hariton és *Komáromi Nándor,* A hővezetési egyenlet és numerikus megoldásának kvalitatív tulajdonságai, I. Az elsőfokú közelítések nemnegativitása 101
Pfeil Tamás, Faragó István, Haroten Hariton és *Komáromi Nándor,* A hővezetési egyenlet és numerikus megoldásának kvalitatív tulajdonságai, II. A másodfokú közelítés nemnegativitása, a maximum elv és az oszcillációmentesség 123
Révész Z. Péter, Benczúr András és *B. Novák Ágnes,* Klasszikus és súlyozott tudásbázisok transzformációi 347
Simon L. Péter és *Farkas Henrik,* Polinomok gyökstruktúrájának vizsgálata a parametrikus reprezentáció módszerével 41
Szidarovszky Ferenc és *Molnár Sándor,* Dinamikus folyamatok konvergenciája és stabilitása 287
Szigeti Ferenc, Vera Carmen E. és *Molnár Sándor,* Kalman-féle rangfeltételek az időtől függő lineáris rendszerekre 279
Szirmai Jenő, Néhány tércsoport optimális gömbkitöltése 87
Vásárhelyiné Szabó Anna, A szerkezetanalízis egy matematikai modellje lokális egyensúlyi folyamatok esetén 293
Vattai Zoltán András, Algoritmus az $F2 \mid \text{overlap} \mid C_{\max}$ megoldására 241
Vera Carmen E., Molnár Sándor és *Szigeti Ferenc,* Kalman-féle rangfeltételek az időtől függő lineáris rendszerekre 279

INDEX

Bagyinszki, J., *pr*-maximal and *pr*-complete clones of functions defined on a finite set 321
Bálint, E. and *Deák, I.*, Parallel computers: optimization software 1
Benczúr, András, B. Novák, Ágnes and *Révész, Z. Péter*, Propositional and weighted knowledge base transformations 347
Béres, László and *Illés, Tibor*, Computational investigation of the covering number of finite projective planes with small order 397
B. Novák, Ágnes, Révész, Z. Péter and *Benczúr, András*, Propositional and weighted knowledge base transformations 347
Bodócs, L., A stability condition for the method of quasi-isometric conjugate pairs 267
Csébfalvi, A., Nonlinear path-following method for stability of structures I. Regular points . 57
Csébfalvi, A., Nonlinear path-following method for stability of structures II. Bifurcation and limit points 71
Csébfalvi, A., Stability analysis of systems characterised by nonsmooth energy function 259
Csendes, T., An interval method for bounding level sets 19
Deák, I. and *Bálint, E.*, Parallel computers: optimization software 1
Faragó, I., Hariton, H. A., Komáromi, N. and *Pfeil, T.*, The differential equation of the heat transfer and qualitative properties its numerical solutions: I. The nonnegativity of the first order approximations 101
Faragó, I., Hariton, H. A., Komáromi, N. and *Pfeil, T.*, The differential equation of the heat transfer and qualitative properties its numerical solutions: II. The nonnegativity of the second order approximation, the maximum principle and the nonoscillation 123
Farkas, H. and *Simon, P. L.*, The investigation of the root structure of polynomials with the parametric representation method 41
Hajdu, M., An algorithm to solve the time-cost trade-offs problem in precedence diagramming . 225
Hajdu, M. and *Klafszky, E.*, A new CPM time-cost trade-offs algorithm 211
Hariton, H. A., Komáromi, N., Pfeil, T. and *Faragó, I.*, The differential equation of the heat transfer and qualitative properties its numerical solutions: I. The nonnegativity of the first order approximations 101
Hariton, H. A., Komáromi, N., Pfeil, T. and *Faragó, I.*, The differential equation of the heat transfer and qualitative properties its numerical solutions: II. The nonnegativity of the second order approximation, the maximum principle and the nonoscillation 123
Illés, Tibor and *Béres, László*, Computational investigation of the covering number of finite projective planes with small order 397
Klafszky, E. and *Hajdu, M.*, A new CPM time-cost trade-offs algorithm 211
Komáromi, N., Pfeil, T., Faragó, I. and *Hariton, H. A.*, The differential equation of the heat transfer and qualitative properties its numerical solutions: I. The nonnegativity of the first order approximations 101
Komáromi, N., Pfeil, T., Faragó, I. and *Hariton, H. A.*, The differential equation of the heat transfer and qualitative properties its numerical solutions: II. The nonnegativity of the second order approximation, the maximum principle and the nonoscillation 123
Kozák, Imre, Remarks on the paper "Determination of the necessary and sufficient compatibility conditions on the boundary" written by G. Lámer 329
Mészáros, Cs., On the modifications of the affine scaling algorithm for linear programming . 185
Mihálykó, Cs., Solubility of integrodifferential equation for the density function of ball mill granulate, and special properties of solution 171
Molnár, S. and *Szidarovszky, F.*, Convergence and Stability of Dynamic Processes 287

Molnár, S., Szigeti, F. and *Vera, C. E.,* Kalman's rank conditions for time dependent linear systems 279
Nagy, T., Stochastic variants of the entropy programming 143
Németh, G., Series approximations for the eigenvalues of Mathhien functions 195
Palágyi, Kálmán, A Local Parallel Noise Reduction Algorithm for Binary Images 373
Pfeil, T., Faragó, I., Hariton, H. A. and *Komáromi, N.,* The differential equation of the heat transfer and qualitative properties its numerical solutions: I. The nonnegativity of the first order approximations 101
Pfeil, T., Faragó, I., Hariton, H. A. and *Komáromi, N.,* The differential equation of the heat transfer and qualitative properties its numerical solutions: II. The nonnegativity of the second order approximation, the maximum principle and the nonoscillation 123
Révész, Z. Péter, Benczúr, András and *B. Novák, Ágnes,* Propositional and weighted knowledge base transformations 347
Simon, P. L. and *Farkas, H.,* The investigation of the root structure of polynomials with the parametric representation method 41
Szidarovszky, F. and *Molnár, S.,* Convergence and Stability of Dynamic Processes 287
Szigeti, F., Vera, C. E. and *Molnár, S.,* Kalman's rank conditions for time dependent linear systems 279
Szirmai, J., Optimale kugelpackungen unter einigen raumgruppen 87
Vásárhelyiné Szabó, Anna, Mathematical model of analysis of structures in the case of local equilibrium processes 293
Vattai, Z. A., Algorithm for solving $F2 \mid \text{overlap} \mid C_{\max}$ problem 241
Vera, C. E., Molnár, S. and *Szigeti, F.,* Kalman's rank conditions for time dependent linear systems 279