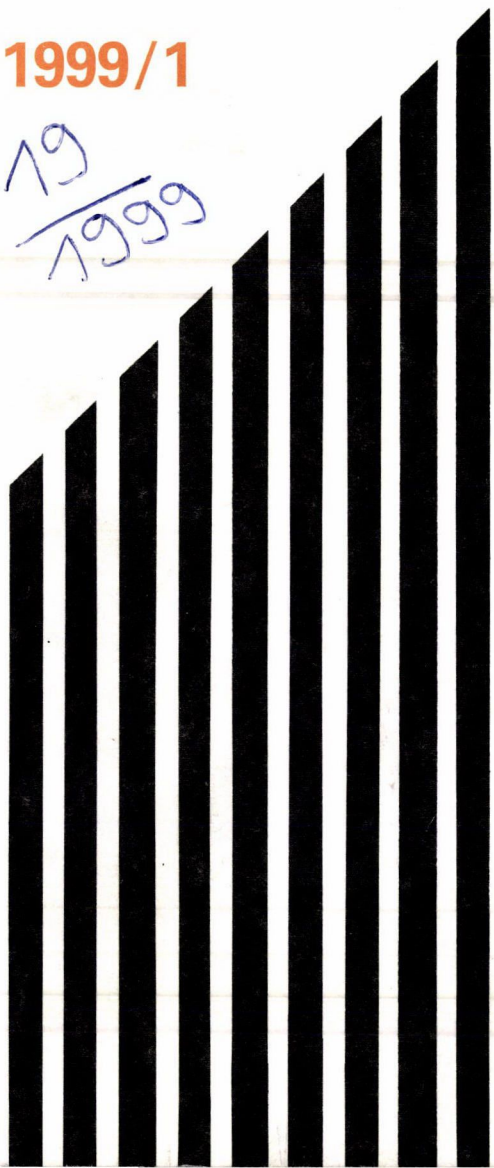
Alkalmazott matematikai lapok

1999/1



A MAGYAR TUDOMÁNYOS AKADÉMIA
MATEMATIKAI TUDOMÁNYOK
OSZTÁLYÁNAK KÖZLEMÉNYEI

19.
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

XIX. kötet 1. szám

Szerkesztőség és kiadóhivatal: 1027 Budapest, Fő u. 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 u. 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 Hungarica,
2. Acta Physica Hungarica,
3. Studia Scientiarum Mathematicarum Hungarica.

Alkalmazott Matematikai Lapok **19** (1999) 1–28

GRADIENS-MÓDSZER SZOBOLJEV-TÉRBEN: LINEÁRIS PEREMÉRTÉKFELADATOK KÖZELÍTŐ MEGOLDÁSA POLINOMOKKAL

KARÁTSON JÁNOS

Budapest

A cikk tárgya a gradiens-módszer Hilbert-térbeli általánosításának alkalmazása $2n$-edrendű lineáris elliptikus peremértékfeladatokra. Ebben a megközelítésben a gradiens-módszer a megfelelő Szoboljev-térben alkalmazható az általánosított differenciáloperátorra. A megvalósítás alapelve Czách László módszerén alapul: ha a tartomány gömbre transzformálható, polinomokból álló közelítő sorozatot készíthetünk. Megmutatjuk, hogy ez a gondolat a numerikus megvalósítás során is kedvező tulajdonságokhoz vezet. A számítógépes megvalósítás kérdései között részletesebben foglalkozunk a kétdimenziós esettel. A közelítő sorozat konstrukciójának köszönhetően az iteráció lépéseiben egyszerű szerkezetű lineáris algebrai egyenletrendszereket kell megoldanunk, s végeredményben könnyen megvalósítható, lineáris konvergenciát nyújtó módszerhez jutunk.

1. Bevezetés

A gradiens-módszer, amely különböző variációival együtt lineáris algebrai egyenletrendszerek megoldásának egyik leghatékonyabb iterációs módszere, elterjedtségét annak köszönheti, hogy diszkretizáció révén jól alkalmazható elliptikus peremértékfeladatok közelítő megoldására. Ezzel szemben az alkalmazások nemigen támaszkodnak a gradiens-módszer végtelen dimenziós általánosításaira, bár Kantorovics munkái óta e téren is számos eredmény született (lásd pl. [4], [8], [13]). A Hilbert-térbeli gradiens-módszert elsőként Czách L. alkalmazta lineáris peremértékfeladatra (*in* [8]); a gradiens-módszer variációi körében a konjugált gradiens-módszer peremértékfeladatokra is használható kidolgozása Daniel nevéhez fűződik ([5]).

E cikk célja előbb Czách L. módszerének kiterjesztése tetszőleges $2n$-edrendű lineáris Dirichlet-feladatra (beleértve a lépésköz technikai szempontból legegyszerűbb választását), majd a numerikus megvalósításhoz szükséges részletek kidolgozása. A kapott módszer, amely gömbön, ill. könnyen gömbre transzformálható tartományokon működik, egyszerűen realizálható és lineáris konvergenciát nyújt.

A 2. szakaszban a Szoboljev-térbeli gradiens-módszer itt szükséges eredményeit foglaljuk össze. Az alkalmazás szempontjából a 3. és 5. szakasz a középpont.

A 3. szakaszban módszerünk konstrukcióját és a fő eredményeket ismertetjük (a bizonyítást a 4. szakasz tartalmazza), az 5. szakasz pedig a számítógépes megvalósítás kérdéseivel foglalkozik másodrendű egyenlet esetén.

Köszönetnyilvánítás.

Ezúton szeretném megköszönni Dr. Czách Lászlónak, hogy megismertette velem a gradiens-módszerrel kapcsolatos eredményeit és felkeltette érdeklődésemet a témába való bekapcsolódáshoz.

2. A gradiens-módszer Szoboljev-térben

Az első két, itt idézett tétel a gradiens-módszer Hilbert-térbeli kiterjesztéseiről szól. A véges dimenziós gradiens-módszert Kantorovics általánosította korlátos lineáris önadjungált operátorra ([7]). Ennek technikailag egyszerűbb, de ugyanolyan konvergenciabecslést nyújtó változata az állandó lépésközű (egyszerű) iteráció:

2.1. TÉTEL (lásd pl. [12]). *Legyen H Hilbert-tér, $A : H \to H$ korlátos önadjungált lineáris operátor, melyre alkalmas $0 < m \leq M$ konstansokkal*

$$m\|x\|^2 \leq \langle Ax, x\rangle \leq M\|x\|^2 \qquad (x \in H).$$

Legyen $y \in H$, ekkor, mint ismeretes, az $Ax = y$ egyenletnek létezik egyetlen $x^* \in H$ megoldása.

Válasszunk tetszőleges $x_0 \in H$ kiindulási elemet és legyen

$$x_n := x_{n-1} - tz_n \qquad (n \in \mathbf{N}^+),$$

$$\text{ahol} \quad z_n := Ax_{n-1} - y \quad \text{és} \quad t := \frac{2}{M+m}.$$

Ekkor az (x_n) sorozat lineárisan konvergál x^*-hoz, éspedig

$$\|x_n - x^*\| \leq \frac{1}{m}\|Ax_0 - y\| \left(\frac{M-m}{M+m}\right)^n \qquad (n \in \mathbf{N}^+).$$

Szintén [12]-ben olvasható a gradiens-módszer nem korlátos operátorra történő kiterjesztésének összefoglalása. Ha ezt a 2.1. Tételre alkalmazzuk, az alábbi egyszerű iterációt kapjuk:

2.2. TÉTEL. *Legyen H Hilbert-tér, $D \subset H$ sűrű altér, $L : D \to H$ szimmetrikus lineáris operátor H-ban, melyre valamely $m_0 > 0$ mellett*

$$\langle Lx, x\rangle \geq m_0\|x\|^2 \qquad (x \in D),$$

valamint legyen $y \in R(L)$. Legyen $B : D \to H$ szimmetrikus lineáris operátor H-ban, melyre az alábbiak teljesülnek:
(1) $R(B) \supset R(L)$
(2) *B és L ekvivalens kvadratikus alakot határoz meg:* $\exists\ M \geq m > 0$, *hogy*

$$m\langle Bx, x\rangle \leq \langle Lx, x\rangle \leq M\langle Bx, x\rangle \qquad (x \in D).$$

Jelölje H_B a B operátor energetikai terét, azaz az $\langle x, y\rangle_B := \langle Bx, y\rangle$ skalárszorzattal ellátott D pre-Hilbert-tér teljessé tételét.
Tetszőleges $x_0 \in D$ mellett legyen

$$x_n := x_{n-1} - tz_n \qquad (n \in \mathbf{N}^+), \tag{2.1}$$

$$\text{ahol} \quad z_n := B^{-1}(Lx_{n-1} - y) \quad \text{és} \quad t := \frac{2}{M+m}.$$

Ekkor az (x_n) sorozat lineárisan konvergál az $Lx = y$ egyenlet x^* megoldásához a

$$\|x_n - x^*\|_B \leq \frac{1}{m\sqrt{p}}\|Lx_0 - y\| \left(\frac{M-m}{M+m}\right)^n \qquad (n \in \mathbf{N}^+) \tag{2.2}$$

($p = m_0 M^{-1}$) becslés szerint.

A 2.2. Tétel már lehetővé teszi, hogy elliptikus peremértékfeladatokra alkalmazzuk a gradiens-módszert a megfelelő Szoboljev-térben. Ezt az eredményt a következő általános tételben foglaljuk össze. (Ez az $N = 1$ speciális esetben tartalmazza közönséges differenciálegyenletek esetét is.) A numerikus megvalósításkor kiemelten fogjuk kezelni a másodrendű parciális differenciálegyenleteket, ennek speciális eseteit a tétel után említjük.

A tételben a multiindexekre szokásos jelöléseket használjuk: ha $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbf{N}^+$, akkor $\partial^\alpha := \partial_1^{\alpha_1} \cdots \partial_N^{\alpha_N}$ és $|\alpha| := \alpha_1 + \ldots + \alpha_N$. Ha $k \in \mathbf{N}$, akkor jelölje d_k azon α multiindexek számát, melyre $|\alpha| = k$.

Ismeretes (lásd pl. [10]), hogy bármely $\Omega \in \mathbf{R}^N$ sima peremű korlátos tartományhoz létezik olyan $\varrho > 0$ szám, hogy

$$\varrho \int_\Omega |u|^2 \leq \int_\Omega |\partial_i u|^2 \qquad \big(u \in H_0^1(\Omega),\ i = 1, \ldots, N\big). \tag{2.3}$$

2.3. Tétel. *Legyen $\Omega \subset \mathbf{R}^N$ korlátos tartomány, $0 < \nu < 1$, $\partial\Omega \in C^{2n,\nu}$. Legyen $n \in \mathbf{N}^+$, $a_{\alpha\beta} = a_{\beta\alpha} \in C^{k,\nu}(\overline{\Omega})$ $\big(|\alpha| = |\beta| = k,\ k = 0, \ldots, n\big)$ és $f \in C^{0,\nu}(\overline{\Omega})$ adott függvények. Legyen $\mu > 0$ és tegyük fel, hogy bármely $x \in \Omega$ esetén*

$$\sum_{|\alpha|=|\beta|=k} a_{\alpha\beta}(x)\xi_\alpha \overline{\xi}_\beta \geq 0 \qquad (k = 0, \ldots, n-1,\ \xi \in \mathbf{C}^{d_k})$$

és

$$\sum_{|\alpha|=|\beta|=n} a_{\alpha\beta}(x)\xi_\alpha\bar{\xi}_\beta \geq \mu \sum_{|\alpha|=n} |\xi_\alpha|^2 \quad (\xi \in \mathbf{C}^{d_n}).$$

(Az $a_{\alpha\beta}$ függvények folytonossága miatt bármely $k = 0, \dots, n$ esetén van olyan $m_k > 0$, hogy

$$\sum_{|\alpha|=|\beta|=k} a_{\alpha\beta}(x)\xi_\alpha\bar{\xi}_\beta \leq m_k \sum_{|\alpha|=k} |\xi_\alpha|^2, \quad \text{ha} \quad x \in \Omega,\ \xi \in \mathbf{C}^{d_k}.)$$

Legyen $D(L) := \left\{ u \in C^{2n,\nu}(\overline{\Omega}) : \partial^\alpha u|_{\partial\Omega} = 0 \ \left(|\alpha| \leq n-1\right) \right\}$ *és tekintsük a következő peremértékfeladatot:*

$$\text{(2.4)} \qquad \begin{cases} Lu := \displaystyle\sum_{|\alpha|=|\beta|\leq n} (-1)^{|\alpha|}\partial^\alpha(a_{\alpha\beta}\partial^\beta u) = f \\ \partial^\alpha u|_{\partial\Omega} = 0 \quad \left(|\alpha| \leq n-1\right). \end{cases}$$

Legyen $m := \frac{\mu}{n!}$, *valamint* $M := \sum_{k=0}^{n} m_k (N\varrho)^{k-n}$ *(ahol* ϱ *(2.3)-ból való).*

Készítsük el a következő iterációt: $u_0 \in D(L)$ *tetszőleges, majd* $k = 1, 2, \dots$ *esetén*

$$\text{(2.5)} \qquad u_k := u_{k-1} - \frac{2}{M+m} z_k \qquad (k \in \mathbf{N}^+),$$

ahol $z_k \in C^{2n,\nu}(\overline{\Omega})$ *a*

$$\text{(2.6)} \qquad \begin{cases} (-1)^n \Delta^n z_k = g_k := Lu_{k-1} - f \\ \partial^\alpha z_k|_{\partial\Omega} = 0 \quad \left(|\alpha| \leq n-1\right) \end{cases}$$

ún. iterációs egyenlet *megoldása.*

Ekkor (u_k) *lineárisan konvergál a (2.4) feladat* $u^* \in C^{2n,\nu}(\overline{\Omega})$ *megoldásához, éspedig*

$$\text{(2.7)} \qquad \|u_k - u^*\|_{H_0^n(\Omega)} \leq \frac{1}{m\sqrt{p}} \|Lu_0 - f\|_{L^2(\Omega)} \left(\frac{M-m}{M+m}\right)^n \qquad (n \in \mathbf{N}^+)$$

(ahol $p = m_0 M^{-1} = m(N\varrho)^n M^{-1}$*).*

Bizonyítás. Ismeretes, hogy $D(L)$ sűrű a $H_0^n(\Omega)$ térben, valamint L szimmetrikus és szigorúan pozitív. Emellett a simasági feltételek révén (2.4)-nek létezik (egyetlen) $u^* \in D(L)$ megoldása ([2]), így $f \in R(L)$. Legyen $B := (-\Delta)^n$ és $D(B) := D(L)$. Ha igazoljuk, hogy B-re teljesül a 2.2. Tétel (1)–(2) feltétele a

$H_0^n(\Omega)$ térben, akkor a tétel alkalmazható (2.4)-re, s ezzel megkapjuk a (2.7) becslést.

(1) A (2.4)-re felhasznált egzisztenciatétel ([2]) érvényes a $B = (-\Delta)^n$ operátor esetén is, így $R(B) = R(L) = C^{0,\nu}(\overline{\Omega})$.

(2) A peremfeltételek miatt

$$\int_\Omega (Lu)\overline{u} = \int_\Omega \sum_{|\alpha|=|\beta|\le n} a_{\alpha\beta}(\partial^\alpha u)(\partial^\beta \overline{u}) \qquad \big(u \in D(L)\big). \tag{2.8}$$

Hasonlóan, (2.3) iterált alkalmazásával tetszőleges $u \in D(B)$, valamint $k = 1, \dots, n$ esetén

$$\int_\Omega (Bu)\overline{u} = \int_\Omega (-1)^n \sum_{i_1,\dots,i_n=1}^{N} (\partial_{i_1}^2 \cdots \partial_{i_n}^2 u)\overline{u} = \int_\Omega \sum_{i_1,\dots,i_n=1}^{N} |\partial_{i_1} \cdots \partial_{i_n} u|^2 \ge$$

$$\ge (N\varrho)^{n-k} \int_\Omega \sum_{i_1,\dots,i_k=1}^{N} |\partial_{i_1} \cdots \partial_{i_k} u|^2 \ge (N\varrho)^n \int_\Omega |u|^2.$$

Emellett (szintén tetszőleges $u \in D(B)$, valamint $k = 1, \dots, n$ esetén)

$$\sum_{|\alpha|=k} |\partial^\alpha u|^2 \le \sum_{i_1,\dots,i_k=1}^{N} |\partial_{i_1} \cdots \partial_{i_k} u|^2 \le k! \sum_{|\alpha|=k} |\partial^\alpha u|^2.$$

Ezekből

$$m \int_\Omega (Bu)\overline{u} = m \int_\Omega \sum_{i_1,\dots,i_n=1}^{N} |\partial_{i_1} \cdots \partial_{i_n} u|^2 \le \mu \int_\Omega \sum_{|\alpha|=n} |\partial^\alpha u|^2 \le \int_\Omega (Lu)\overline{u} \le$$

$$\le \int_\Omega \sum_{k=0}^{n} m_k \sum_{|\alpha|=k} |\partial^\alpha u|^2 \le \int_\Omega \sum_{k=0}^{n} m_k (N\varrho)^{k-n} \sum_{i_1,\dots,i_n=1}^{N} |\partial_{i_1} \cdots \partial_{i_n} u|^2 =$$

$$= M \int_\Omega (Bu)\overline{u},$$

ahol $M := \sum_{k=0}^{n} m_k (N\varrho)^{k-n}$. □

2.1. Megjegyzés. A másodrendű esetben ($n = 1$) az operátor:

$$Lu := -. \sum_{i,j=1}^{N} \partial_i (a_{ij} \partial_j u) + qu.$$

Az előjelfeltételek ekkor a következők: léteznek olyan $\mu, m_0, m_1 > 0$ állandók, hogy

$$\mu|\xi|^2 \le \sum_{i,j=1}^{N} a_{ij}(x)\xi_i\overline{\xi}_j \le m_1|\xi|^2 \qquad (x \in \Omega,\ \xi \in \mathbf{C}^N),$$

valamint $0 \le d(x) \le m_0$ $(x \in \Omega)$. Az alábbi speciális esetekben a μ és m_1 állandók egyszerűen becsülhetők:

(a) Ha az $\big(a_{ij}(x)\big)$ $(x \in \overline{\Omega})$ mátrix egyenletesen diagonálisan domináns, azaz

$$\alpha := \min_i \Big\{ \min_{\overline{\Omega}} a_{ii} \Big\} - \max_i \sum_{j \ne i} \|a_{ij}\|_{C(\overline{\Omega})} > 0,$$

akkor $\mu \ge \alpha$, valamint $m_1 \le \max_i \sum_j \|a_{ij}\|_{C(\overline{\Omega})}$.

(b) Ha az operátor

$$Lu := -\sum_{i=1}^{N} \partial_i(a_i\partial_i u) + qu$$

alakú, akkor az a_i függvényekre az előjelfeltétel jelentése: $\alpha := \min_i\{\min_{\overline{\Omega}} a_i\} > 0$. Ekkor $\mu = \alpha$, valamint $m_1 = \max_i \|a_i\|_{C(\overline{\Omega})}$. (Lásd Czách L. idézett eredményét, [8]).

3. A polinomokkal történő közelítés konstrukciója és a módszer konvergenciája

A következőkben ismertetett módszer akkor alkalmazható, ha az Ω tartomány sima transzformációval gömbbe vihető át.

Ekkor a módszer lényege, hogy az (egységgömbre transzformált) egyenlet együtthatóit polinomokkal approximáljuk, és az eredeti egyenlet helyett a perturbált egyenlet megoldását közelítjük. Ez utóbbi esetén ugyanis elérhető, hogy a közelítő sorozat polinomokból álljon, és ezáltal az iterációs egyenletek megoldása lineáris algebrai egyenletrendszerre redukálódjon. Az approximáció javításával a gömbön a pontos egyenlet megoldását közelítjük, s ezt a közelítő sorozatot végül visszatranszformáljuk az eredeti tartományra.

Numerikus szempontból a módszer alapja az, hogy a gömbön az iterációs egyenleteket lineáris algebrai egyenletrendszerrel megoldhatjuk. Ez a következő tényen alapul (l. [8]): legyen $S \subset \mathbf{R}^n$ az egységgömb, P pedig (N-változós) polinom. Ekkor a

$$\begin{cases} \Delta^n u = P \\ \partial^\alpha u|_{\partial S} = 0 \quad (|\alpha| \le n-1) \end{cases} \tag{3.1}$$

egyenlet u megoldása is ($2n$-nel magasabb fokú) polinom, amely a következőképp kapható meg. Legyen a P polinom foka r, s keressük u-t (a peremfeltételeket eleve kielégítő)

$$u(x_1,\dots,x_N) = \left(\sum_{i=1}^{N} x_i^2 - 1\right)^n Q(x_1,\dots,x_N)$$

alakban, ahol Q is r-edfokú polinom. Ekkor a $\Delta^n u$ r-edfokú polinom együtthatói a Q együtthatóinak lineáris kombinációi, így Q együtthatóit a $\Delta^n u$ és P egyenlővé tételéből kapott lineáris egyenletrendszer megoldása adja. (A fellépő mátrix determinánsa nem 0, ugyanis a megfelelő homogén egyenletnek csak triviális megoldása van a homogén (3.1) egyenlet konstans 0 megoldásának egyértelműsége miatt. Így a lineáris algebrai egyenletrendszernek létezik megoldása.)

Az alábbiakban összefoglaljuk a módszer konstrukcióját és a konvergenciáról szóló eredményeket. Az állítások bizonyításait a 4. szakasz tartalmazza. A fent említett lineáris algebrai egyenletrendszer alakját a számítógépes megvalósítás kérdéseiről szóló 5. szakaszban vizsgáljuk meg.

(a) A konstrukció gömbön

Legyen $p \in \mathbf{N}$ és teljesüljenek az S egységgömbön a (2.4) egyenlet együtthatóira az alábbi simasági feltételek:

$$a_{\alpha\beta} \in C^{p+1,\nu}(\overline{S}) \quad \big(|\alpha| = |\beta| = i,\ i = 0,\dots,n\big), \quad f \in C^{p,\nu}(\overline{S}).$$

Ekkor a Jackson-féle approximációs tételeknek köszönhetően ([11]) konstruálhatók olyan $\left(a_{\alpha\beta}^{[k]}\right)_{k\in\mathbf{N}}$ és $\left(f^{[k]}\right)_{k\in\mathbf{N}}$ k-adfokú polinomokból álló sorozatok, melyekre az alábbi becslések teljesülnek:

(1) $|\alpha| = \beta| = i$ $(i = 0,\dots,n)$, $j = 0,\dots,p+i$ esetén:

$$\left\|a_{\alpha\beta} - a_{\alpha\beta}^{[k]}\right\|_{C^j(\overline{S})} \le \frac{A}{k^{p+i-j+\nu}} \qquad (k \in \mathbf{N}); \tag{3.2}$$

(2) $j = 0,\dots,p$ esetén

$$\left\|f - f^{[k]}\right\|_{C^j(\overline{S})} \le \frac{A}{k^{p-j+\nu}} \tag{3.3}$$

(ahol $A > 0$ k-tól független állandó).

Definiáljuk az L operátor közelítéseit a következőképpen: $u \in D\big(L^{[k]}\big) = D(L)$ esetén legyen

$$L^{[k]}u := \sum_{|\alpha|=|\beta|\le n} (-1)^{|\alpha|}\partial^\alpha\big(a_{\alpha\beta}^{[k]}\partial^\beta u\big).$$

A (2.4) feladatot ezáltal az alábbiakkal közelítjük:

$$(3.4)_k \qquad \begin{cases} L^{[k]}u = f^{[k]} \\ \partial^\alpha u|_{\partial S} = 0 \quad (|\alpha| \le n-1) \end{cases}$$

Alkalmazzuk a $(3.4)_k$ feladatokra a 2.3. Tétel-beli gradiens-módszert az $u_0^{[k]} := 0$ kezdőfüggvénnyel! Ekkor e szakasz bevezetőjében írtak következményeképp a $(3.4)_k$ egyenlethez készített $(u_j^{[k]})$ sorozat mindvégig polinomokból áll, amelyeket lépésenként egy lineáris algebrai egyenletrendszerrel határozhatunk meg.

Legyen végül $(m_k)_{k\in\mathbf{N}} \to \infty$ szigorúan növekedő indexsorozat és

$$u_{(k)} := u_k^{[m_k]},$$

azaz az $u_j^{[m_k]}$ $(k, j \in \mathbf{N})$ kétindexű sorozatból alkalmas „átlós" típusú egyindexű sorozatot választunk.

Azt várjuk, hogy az $u_{(k)}$ sorozat az eredeti (2.4) $Lu = f$ feladat megoldásához konvergál.

(b) A konstrukció egyéb tartományon

Legyen $p \in \mathbf{N}$, és teljesüljenek az $\Omega \subset \mathbf{R}^N$ tartományon a (2.4) egyenlet együtthatóira az (a) ponthoz hasonlóan az alábbi simasági feltételek:

$$a_{\alpha\beta} \in C^{p+i,\nu}(\overline{\Omega}) \quad \big(|\alpha| = |\beta| = i,\ i = 0, \dots, n\big), \quad f \in C^{p,\nu}(\overline{\Omega}).$$

Tegyük fel, hogy található olyan $T \in C^{2n+1+p,\nu}$ diffeomorfizmus, melyre $\det T' \neq 0$ $\overline{S}$-on. (Akkor T^{-1} is $C^{2n+1+p,\nu}$-beli.)

Legyen $u \in L^2(\Omega)$ esetén

$$\hat{T}u := (u \circ T)|\det T'|^{1/2}. \tag{3.5}$$

Ekkor $\hat{T} : L^2(\Omega) \to L^2(S)$ izometria, emellett (T simasága miatt) $\hat{T}$ bijekció $C^{p+2,\nu}(\overline{\Omega})$ és $C^{p+2,\nu}(\overline{S})$ között is. Kézenfekvő, de terjedelmes számolással igazolható, hogy $\hat{L} := \hat{T}L\hat{T}^{-1}$ szintén másodrendű lineáris differenciáloperátor, melynek együtthatói öröklik L együtthatóinak simaságát. Itt $\hat{L}$ definíciója szerint $\hat{L}(\hat{T}u) = \hat{T}(Lu)$ $(u \in D(L))$, ami azt jelenti, hogy a

$$(3.6)_a \qquad \begin{cases} Lu = f \quad \Omega\text{-ban} \\ u|_{\partial\Omega} = 0 \end{cases} \qquad \text{és} \qquad (3.6)_b \qquad \begin{cases} \hat{L}w = f \quad S\text{-ben} \\ w|_{\partial S} = 0 \end{cases}$$

egyenletek ekvivalensek. Azaz:

$$u \text{ megoldása } (3.6)_a\text{-nak} \iff \hat{T}u \text{ megoldása } (3.6)_b\text{-nek.}$$

A gömbre transzformált $(3.6)_b$ egyenlethez az (a) pontban megadott eljárással elkészíthetjük a $(w_{(k)})$ közelítő sorozatot. Ebből az

$$(u_{(k)}) := (\hat{T}^{-1} w_{(k)}) = \left((w_{(k)} \circ T^{-1}) \left| \det (T^{-1})' \right|^{1/2} \right)$$

sorozat az eredeti egyenletet közelítő sorozat lesz.

Ez az eset csak akkor vizsgálható érdemben, ha a szóban forgó T bijekció könnyen megadható. Erre az 5. szakaszban látunk példát (b. pont).

(c) A módszer konvergenciája

3.1. TÉTEL. *Legyen* $\frac{M-m}{M+m} < q_2 < 1$, $r \in \mathbf{N}$ *adott*, $1 < h < \left(\frac{1}{q_2}\right)^{\frac{1}{N+2r}}$ *és* $m_k := \lfloor h^k \rfloor$ *(h^k egészrésze), ha* $k \in \mathbf{N}$. *Ekkor*

(i) $\exists\, c > 0$, $0 < q < 1$:

$$\left\| u_{(k)} - u^* \right\|_{H^{2n}(\Omega)} \leq c q^k \qquad (k \in \mathbf{N}); \tag{3.7}$$

(ii) *Ha* $2n \geq \left\lfloor \frac{N}{2} \right\rfloor + 1$, *akkor* $\exists\, c_1 > 0 : n' := 2n - \left\lfloor \frac{N}{2} \right\rfloor - 1$ *mellett*

$$\left\| u_{(k)} - u^* \right\|_{C^{n'}(\overline{\Omega})} \leq c_1 q^k \qquad (k \in \mathbf{N}); \tag{3.8}$$

(iii) *Ha* $p \geq N$, *akkor* $r \leq \frac{p-N}{2}$ *esetén a* (2.4) *egyenlet megoldására* $u^* \in C^{2n+r}(\overline{\Omega})$ *teljesül, valamint* $\exists c_r > 0$, $0 < q_r < 1$:

$$\left\| u_{(k)} - u^* \right\|_{C^{2n+r}(\overline{\Omega})} \leq c_r q_r^k \qquad (k \in \mathbf{N}); \tag{3.9}$$

(iv) *Ha* $p \geq \frac{N}{2}$, *akkor* $r < p + \nu - \frac{N}{2}$ *esetén* $u^* \in C^{2n+r}(\Omega)$, *és* $\exists 0 < \hat{q}_r < 1$, *továbbá* $\forall\, K \subset \Omega$ *kompakt részhalmazhoz* $\exists\, c_r(K) > 0$:

$$\left\| u_{(k)} - u^* \right\|_{C^{2n+r}(K)} \leq c_r(K) (\hat{q}_r)^k \qquad (k \in \mathbf{N}). \tag{3.10}$$

3.1. Megjegyzés. Említést érdemel a tétel következő speciális esete: másodrendű egyenlet közelítésekor az eredeti (2.4) feladat simasági feltételeivel (azaz $p = 0$ mellett) a (ii) becslésből $N = 2$ és 3 dimenzió esetén egyenletes konvergenciát nyerünk.

(d) Analitikus együtthatók esete

Ebben a pontban megemlítjük, hogy ha (síkbeli feladat esetén) az együtthatófüggvények analitikusak, akkor ezek jobb approximációs tulajdonságai miatt a konvergenciabecslések is javulnak.

Legyen tehát $N = 2$, valamint az $a_{\alpha\beta}$ és f függvények analitikusak $\overline{\Omega}$-on (azaz egy $\overline{\Omega}$-at tartalmazó nyílt halmazon).

Ha van olyan $T : \overline{S} \to \overline{\Omega}$ valós-analitikus bijekció, melyre $\det T' \neq 0$ $\overline{S}$-on, akkor a (2.4) egyenlet — a fentebb látott transzformációval — olyan egyenletbe vihető át, ahol az együtthatók és jobboldal analitikusak az $\overline{S}$ egységkörlapon. (A megfelelő transzformáció létezésével az 5.c) pontban foglalkozunk.) Ekkor az $\overline{S}$ körlapon most olyan $\left(a_{\alpha\beta}^{[k]}\right)_{k\in\mathbf{N}}$ és $\left(f^{[k]}\right)_{k\in\mathbf{N}}$ k-adfokú polinomokból álló sorozatok készíthetőek, melyekre alkalmas $A > 0$ és $0 < q < 1$ állandók mellett, bármely $|\alpha| = |\beta| \leq n$ esetén

$$\left\|a_{\alpha\beta} - a_{\alpha\beta}^{[k]}\right\|_{C^j(\overline{S})} \leq aq^k \quad \text{és} \quad \left\|f - f^{[k]}\right\|_{C^j(\overline{S})} \leq Aq^k \quad (k \in \mathbf{N}).$$

A továbbiakban a módszer ugyanúgy folytatódik, mint az (a)-(b) pontokban, csupán most elég az $m_k := k$ indexsorozattal értelmezni az $u_{(k)}$ függvényeket.

3.2. TÉTEL. *$u^* \in C^\infty(\overline{\Omega})$, emellett bármely $r \in \mathbf{N}$ esetén $\exists\, c_r > 0, 0 < q_r < 1$:*

$$\left\|u_{(k)} - u^*\right\|_{C^r(\overline{\Omega})} \leq c_r q_r^k \qquad (k \in \mathbf{N}).$$

A módszer megvalósítása akkor a legegyszerűbb, ha az $a_{\alpha\beta}$ és f függvények Taylor-sorba fejthetőek egy $\overline{\Omega}$-at tartalmazó nyílt halmazon. Ekkor a sor szeletei adják az $a_{\alpha\beta}^{[k]}$ és $f^{[k]}$ polinomokat.

4. A konvergencia bizonyítása

A bizonyítást külön végezzük az egységgömbön, majd erre visszavezetve egyéb tartományokon. Ezek előtt, az első pontban a bizonyításhoz szükséges approximációelméleti eredményeket foglaljuk össze.

(a) Approximációelméleti segédtételek

4.1. TÉTEL ([8]). *Legyen $S \subset \mathbf{R}^N$ a nyílt egységgömb, $K \subset S$ kompakt részhalmaz, $\alpha, \beta \in \mathbf{N}$, $p : S \to \mathbf{R}$ pedig n-edfokú polinom. Ekkor léteznek olyan $A, B > 0$ n-től független konstansok, hogy*

$$\|p\|_{C(\overline{S})} \leq An^N \|p\|_{L^2(\overline{S})} \quad \textit{ill.} \quad \|p\|_{C(K)} \leq Bn^{\frac{N}{2}} \|p\|_{L^2(\overline{S})}.$$

(A tétel [8]-ban $N = 2$ esetén szerepel, az általános eset bizonyítása ezzel teljesen analóg.)

A következő tétel [9] 2. tételének módosítása.

4.2. TÉTEL. *Legyen $S \subset \mathbf{R}^N$ a nyílt egységgömb, $f \in L^2(S)$. Legyen $(p_n) : S \to \mathbf{R}$ tetszőleges polinomsorozat, ahol $(\operatorname{gr} p_n)$ (azaz a p_n polinomok fokaiból álló számsorozat) monoton növő, továbbá (ε_n) monoton fogyó 0-sorozat, melyre*

$$\|p_n - f\|_{L^2(S)} \le \varepsilon_n \qquad (n \in \mathbf{N}).$$

Legyen $r \in \mathbf{N}$. Ekkor a következők teljesülnek:
(i) *Ha $\sum_j (\operatorname{gr} p_{j+1})^{N+2r} \varepsilon_j < \infty$, akkor $f \in C^r(\overline{S})$ és $\exists c > 0$:*

$$\|p_n - f\|_{C^r}(\overline{S}) \le c \sum_{j=n}^{\infty} (\operatorname{gr} p_{j+1})^{N+2r} \varepsilon_j \qquad (n \in \mathbf{N}).$$

(ii) *Ha $\sum_j (\operatorname{gr} p_{j+1})^{\frac{N}{2}+r} \varepsilon_j < \infty$, akkor $f \in C^r(S)$ és $\forall\, K \subset S$ kompakt részhalmaz esetén $\exists\, c_K > 0$:*

$$\|p_n - f\|_{C^r}(K) \le c_K \sum_{j=n}^{\infty} (\operatorname{gr} p_{j+1})^{\frac{N}{2}+r} \varepsilon_j \qquad (n \in \mathbf{N}).$$

Bizonyítás. Analóg az eredeti [9] 2. tétel bizonyításával, ha abban a polinomok fokát (ami ott n) az általános $\operatorname{gr} p_n$-re cseréljük, valamint a felhasznált segédtételben a fenti 4.1. tételt használjuk az L^1-normabecslés helyett.

4.1. Következmény. Legyen $S \subset \mathbf{R}^N$ a nyílt egységgömb, $K \subset S$ kompakt részhalmaz, $a, c > 0$, $0 < q < 1$, $k \in \mathbf{N}^+$, $f \in L^2(S)$. Legyen $(p_n) : S \to \mathbf{R}$ polinomsorozat, melyre $\operatorname{gr} p_n \le an^k$, emellett

$$\|p_n - f\|_{L^2(S)} \le cq^n \qquad (n \in \mathbf{N}).$$

Ekkor $f \in C^\infty(\overline{S})$ és $\forall\, r \in \mathbf{N}$ esetén $\exists\, c_r > 0$, $0 < q_r < 1$, hogy

$$\|p_n - f\|_{C^r(\overline{S})} \le c_r q_r^n \qquad (n \in \mathbf{N}).$$

($f \in C^\infty(\overline{S})$ itt azt jelenti, hogy f majdnem mindenütt megegyezik egy $C^\infty(\overline{S})$-beli függvénnyel. A továbbiakban is így értjük egy $L^2(S)$-beli függvény simaságát.)

Bizonyítás. Alkalmazhatjuk a 4.2. Tétel (ii) részét $\varepsilon_n := cq^n$ mellett, hiszen alkalmas $c_m > 0$ és $q < q_m < 1$ konstansokkal

$$\sum_{j=n}^{\infty} (\operatorname{gr} p_{j+1})^m \varepsilon_j \le ac \sum_{j=n}^{\infty} (j+1)^{km} q^j \le \sum_{j=n}^{\infty} c_m q_m^j = \frac{c_m}{1-q_m} q_m^n \qquad (m \in \mathbf{N}). \quad \square$$

Közvetlenül megfogalmazható a 4.2. Tétel következménye, ha a polinomsorozat k-adikig bezárólag vett deriváltjainak L^2-beli konvergenciáját tesszük fel, s a tételt ezekre alkalmazzuk.

4.2. Következmény. Legyen $S \subset \mathbf{R}^N$ a nyílt egységgömb, $k \in \mathbf{N}^+$, $f \in H^k(S)$. Legyen $(p_n) : S \to \mathbf{R}$ tetszőleges polinomsorozat, ahol $(\operatorname{gr} p_n)$ monoton növő, továbbá (ε_n) monoton fogyó 0-sorozat, melyre

$$\|p_n - f\|_{H^k(S)} \le \varepsilon_n \qquad (n \in \mathbf{N}).$$

(i) Ha $r \in \mathbf{N}$ és $\sum_j (\operatorname{gr} p_{j+1})^{N+2r} \varepsilon_j < \infty$, akkor $f \in C^{k+r}(\overline{S})$ és $\exists\, c > 0$:

$$\|p_n - f\|_{C^{k+r}(\overline{S})} \le c \sum_{j=n}^{\infty} (\operatorname{gr} p)^{N+2r} \varepsilon_j \qquad (n \in \mathbf{N}).$$

(ii) Ha $r \in \mathbf{N}$ és $\sum_j (\operatorname{gr} p_{j+1})^{\frac{N}{2}+r} \varepsilon_j < \infty$, akkor $f \in C^{k+r}(S)$ és $\forall\, K \subset S$ kompakt részhalmaz esetén $\exists\, c_K > 0$:

$$\|p_n - f\|_{C^{k+r}(K)} \le c_K \sum_{j=n}^{\infty} (\operatorname{gr} p_{j+1})^{\frac{N}{2}+r} \varepsilon_j \qquad (n \in \mathbf{N}).$$

(b) A konvergencia bizonyítása gömbön

Tekintsük a (2.4) egyenletet az $\Omega = S(0,1) \subset \mathbf{R}^N$ egységgömbön a 2(a) szakaszbeli feltételekkel, és készítsük el az ott megadott konstrukció szerinti „átlós" közelítő sorozatot. (Az (m_k) indexsorozatot tehát egyelőre még nem rögzítjük.)

A 2.3. Tételben láttuk, hogy az eredeti $Lu = f$ egyenletre alkalmazott gradiensmódszer $\frac{M-m}{M+m}$ kvóciense az L operátornak a $B = (-\Delta)^n$ operátorra nézve vett m és M határaitól függ. Először e határoknak (az approximált operátorokban) k-tól való függéséről szóló lemmára van szükségünk.

4.1. LEMMA. *Bármely $0 < m' < m$ és $M' > M$ esetén található olyan $N_0 \in \mathbf{N}$ index, hogy ha $k > N_0$, akkor az $L^{[k]}$ operátor $m^{[k]}$ és $M^{[k]}$ határaira $m^{[k]} > m'$ és $M^{[k]} < M'$ teljesül.*

Bizonyítás. Abból következik, hogy a (3.2–3) becslések révén (2.8) perturbációja tetszőlegesen kicsivé tehető $\langle Bu, u\rangle$-hoz képest, így $\langle L^{[k]}u, u\rangle$ határai tetszőlegesen közel kerülhetnek $\langle Lu, u\rangle$ eredeti határaihoz. A részletes számolást az olvasóra bízzuk.

4.3. Következmény. Található olyan $N_0 \in \mathbf{N}$ index, hogy ha $k > N_0$, akkor az $L^{[k]}u = f^{[k]}$ egyenleteknek létezik (egyetlen) $u \in D(L)$ megoldása.

Bizonyítás. Mivel $L^{[k]}$ együtthatói és $f^{[k]}$ polinomok, így (bőségesen) teljesítik a 2.3. Tételben idézett simasági feltételeket. Ezért ha $k > N_0$ esetén $L^{[k]}$ alsó határa pozitív, akkor létezik egyetlen megoldás.

4.4. Következmény. Található olyan $N_0 \in \mathbf{N}$ index és $0 < q_1 < 1$, hogy ha $k > N_0$, akkor az $L^{[k]}u = f^{[k]}$ egyenletekre alkalmazott gradiens-módszer lineárisan konvergál q_1 kvócienssel.

Bizonyítás. Legyen $0 < m' < m$ és $M' > M$. Ekkor a 4.1. Lemma alapján az ott kapott $N_0 \in \mathbf{N}$ index, valamint $q_1 := \frac{M'-m'}{M'+m'}$ megfelelő.

4.2. LEMMA. *Ha $C^{p+\nu}$-beli együtthatók esetén a* 3(a) *pontbeli konstrukciót alkalmazzuk, akkor $\varepsilon_k := \frac{A}{k^{p+\nu}}$ mellett, ha pedig analitikus együtthatók esetén a* 3(d) *pontbeli konstrukciót alkalmazzuk, akkor $\varepsilon_k := Aq^k$ mellett az alábbi becslés teljesül. Alkalmas $A_1 > 0$ mellett*

$$\left\|u^{[k]} - u^*\right\|_{H^{2n}(S)} \le A_1\varepsilon_k \qquad (k \in \mathbf{N}), \tag{4.1}$$

ahol u^* a (2.4), *$u^{[k]}$ pedig a $(3.4)_k$ egyenlet megoldása.*

Bizonyítás. Mivel $Lu^* = f$ és $L^{[k]}u^{[k]} = f^{[k]}$, így

$$L^{[k]}\big(u^{[k]} - u^*\big) = \big(L - L^{[k]}\big)u^* + f^{[k]} - f.$$

Így

$$\left\|L^{[k]}\big(u^{[k]} - u^*\big)\right\|_{L^2(S)} \le \left\|\big(L - L^{[k]}\big)u^*\right\|_{L^2(S)} + \left\|f^{[k]} - f\right\|_{L^2(S)}. \tag{4.2}$$

Itt

$$\big(L - L^{[k]}\big)u^* = \sum_{j\le n}\sum_{|\alpha|=|\beta|=j}(-1)^{|\alpha|}\partial^\alpha\Big(\big(a_{\alpha\beta} - a_{\alpha\beta}^{[k]}\big)(\partial^\beta u^*)\Big).$$

Az összeg minden tagjában a deriválások elvégzése után $a_{\alpha\beta} - a_{\alpha\beta}^{[k]}$ legfeljebb j-edik deriváltjai szorzódnak u^* legfeljebb $2j$-edik deriváltjaival. A 3(a) vagy (d)-beli konstrukciókat éppen úgy definiáltuk, hogy e szorzatok első tényezőinek maximumnormája ε_k-val becsülhető. A második tényezőkben vehetjük u^* szereplő deriváltjának L^2-normáját. Összegezve (háromszög-egyenlőtlenség után):

$$\left\|\big(L - L^{[k]}\big)u^*\right\|_{L^2(S)} \le \text{const.} \cdot \varepsilon_k\|u^*\|_{H^{2n}(S)}.$$

Emellett (szintén a konstrukció szerint)

$$\left\|f^{[k]} - f\right\|_{L^2(S)} \le \text{const.} \cdot \left\|f^{[k]} - f\right\|_{C(\overline{S})} \le \text{const.} \cdot \varepsilon_k,$$

Így (4.2)-ből

$$\left\|L^{[k]}\big(u^{[k]} - u^*\big)\right\|_{L^2(S)} \le \text{const.} \cdot \varepsilon_k \qquad (k \in \mathbf{N}).$$

A $2n$-edrendű Bernstein-Ladüzsenszkaja-egyenlőtlenség szerint

$$\left\|L^{[k]}w\right\|_{L^2(S)} \geq \text{const.} \cdot \|w\|_{H^{2n}(S)} \qquad \left(w \in H^{2n}(S) \cap H_0^n(S)\right),$$

amiből már következik a kívánt (4.1) becslés.

4.3. SEGÉDTÉTEL. *Tekintsük a* 3(a) *vagy* (d) *pontbeli konstrukciót, és legyen* $\varepsilon_k > 0$ *mint az előbbi* 4.2. *lemmában. Legyen* $(m_k)_{k\in\mathbf{N}} \to \infty$ *növekedő indexsorozat és* $u_{(k)} := u_k^{[m_k]}$.

Ekkor az alábbi négy állítás teljesül:

(i) $\exists\ c_2 > 0,\ 0 < q_2 < 1$, *hogy az* $\left(u_{(k)}\right)$ *sorozat az*

$$\left\|u_{(k)} - u^*\right\|_{H^{2n}(S)} \leq \varepsilon_{m_k} + c_2 q_2^k \qquad (k \in \mathbf{N}) \tag{4.3}$$

becslés szerint konvergál u^**-hoz.*

(ii) *Ha* $2n \geq \left\lfloor \frac{N}{2} \right\rfloor + 1$, *akkor* $\exists\ c_3 > 0 : n' := 2n - \left\lfloor \frac{N}{2} \right\rfloor - 1$ *mellett*

$$\left\|u_{(k)} - u^*\right\|_{C^{n'}(\overline{S})} \leq c_3(\varepsilon_{m_k} + c_2 q_2^k) \qquad (k \in \mathbf{N}) \tag{4.4}$$

(iii) *Ha valamely* $r \in \mathbf{N}$ *esetén* $\sum_j \left[(j+1)(m_{j+1}+2n)\right]^{N+2r}(\varepsilon_{m_j} + c_2 q_2^j) < \infty$, *akkor* $u^* \in C^{2n+r}(\overline{S})$ *és* $\exists\ c_4 > 0 : \forall\ k \in \mathbf{N}$ *esetén*

$$\left\|u_{(k)} - u^*\right\|_{C^{2n+r}(\overline{S})} \leq c_4 \sum_{j=k}^{\infty} \left[(j+1)(m_{j+1}+2n)\right]^{N+2r}(\varepsilon_{m_j} + c_2 q_2^j) \tag{4.5}$$

(iv) *Ha* $r \in \mathbf{N}$ *és* $\sum_j \left[(j+1)(m_{j+1}+2n)\right]^{\frac{N}{2}+r}(\varepsilon_{m_j} + c_2 q_2^j) < \infty$, *akkor* $u^* \in C^{2n+r}(S)$ *és* $\forall\ K \subset S$ *kompakt részhalmaz esetén* $\exists\ c_K > 0$:

$$\left\|u_{(k)} - u^*\right\|_{C^{2n+r}(K)} \leq c_K \sum_{j=k}^{\infty} \left[(j+1)(m_{j+1}+2n)\right]^{\frac{N}{2}+r}(\varepsilon_{m_j} + c_2 q_2^j). \tag{4.6}$$

Bizonyítás. (a) Könnyen látható, hogy a $(3.4)_k$ egyenlet közelítésekor $u_j^{[k]}$ polinom marad, melynek foka

$$\operatorname{gr} u_j^{[k]} = j(k+2n) \qquad (k, j \in \mathbf{N}). \tag{4.7}$$

Ui. $\operatorname{gr} u_0^{[k]} = \operatorname{gr} 0 = 0$, és ha $u_{j-1}^{[k]}$ $(j-1)(k+2n)$-adfokú polinom $(j \in \mathbf{N}^+)$, akkor $\left(L^{[k]}u_{j-1}^{[k]} - f^{[k]}\right)$ $\left((j-1)(k+2n)+k\right)$-adfokú polinom lesz (hisz $a_{\alpha\beta}^{[k]}$ és $f^{[k]}$ k-adfokú polinomok). A

$$\begin{cases} \Delta^n z_j^{[k]} = (-1)^n\left(L^{[k]}u_{j-1}^{[k]} - f^{[k]}\right) \\ \partial^\alpha z_j^{[k]}\big|_{\partial S} = 0 \quad (|\alpha| \leq n-1) \end{cases}$$

egyenletből $z_j^{[k]}$ polinom és $\operatorname{gr} z_j^{[k]} = (j-1)(k+2n)+k+2n = j(k+2n)$, végül $u_j^{[k]} := u_{j-1}^{[k]} - tz_j^{[k]}$ révén $\operatorname{gr} u_j^{[k]} := \max\{\operatorname{gr} u_{j-1}^{[k]}, \operatorname{gr} z_j^{[k]}\} = j(k+2n)$.

(i) A $(3.4)_k$ feladatra vonatkozó (2.7) becslésre — felhasználva (4.7)-et, és hogy $w \in H_0^n(S)$ esetén $\|w\|_{L^2(S)} \le \text{const.} \cdot \|w\|_{H_0^n(S)}$ — alkalmazhatjuk a 4.1. következményt. Így $u^{[k]} \in C^{2n}(\overline{S})$ és $\exists\, d_2 > 0$, $q_1 < q_2 < 1$:

$$\left\|u^{[k]} - u_j^{[k]}\right\|_{C^{2n}(\overline{S})} \le d_2 q_2^j \qquad (j \in \mathbf{N}).$$

Ebből alkalmas $c_2 > 0$ konstanssal

$$\left\|u^{[k]} - u_j^{[k]}\right\|_{H^{2n}(S)} \le c_2 q_2^j \qquad (j \in \mathbf{N}). \tag{4.8}$$

Innen rögtön következik (4.3), hiszen (4.1) és (4.8) alapján

$$\left\|u^* - u_{(k)}\right\|_{H^{2n}(S)} = \left\|u^* - u_k^{[m_k]}\right\|_{H^{2n}(S)} \le$$

$$\le \left\|u^* - u^{[m_k]}\right\|_{H^{2n}(S)} + \left\|u^{[m_k]} - u_k^{[m_k]}\right\|_{H^{2n}(S)} \le \varepsilon_{m_k} + c_2 q_2^k \quad (k \in \mathbf{N}).$$

(ii) (4.4) annak folyománya, hogy ha $n := 2n - \left\lfloor \frac{N}{2} \right\rfloor - 1$, akkor a Szoboljev-féle beágyazási tétel szerint $H^{2n}(S) \subset C^{n'}(\overline{S})$, és alkalmas $c_3 > 0$ mellett $\|w\|_{C^{n'}(\overline{S})} \le c_3 \|w\|_{H^{2n}(S)}$, ha $w \in H^{2n}(S)$ (l. [1]).

(iii)–(iv) Mivel (4.7) szerint $\operatorname{gr} u_{(k)} = \operatorname{gr} u_k^{[m_k]} = k(m_k + 2n)$ $(k \in \mathbf{N})$, így (4.3)-ra alkalmazva a 4.2. következmény (i)–(ii) állítását (ezt a $H_0^n(S)$-beli megfelelő normák ekvivalenciája miatt megtehetjük) megkapjuk bizonyítandó tételünk (iii)–(iv) részét. □

(A továbbiakban az egyszerűség kedvéért feltesszük, hogy a 4.1. lemmában és két következményben $N_0 = 1$. Ez nem megszorítás, hiszen véges sok tag módosulása minden felhasznált becslésben csak a konstans szorzó értékét változtatja meg.)

A 3.1. Tétel bizonyítása (az egységgömbön)

A 4.3. segédtételt alkalmazhatjuk $\varepsilon_k := \frac{A}{k^{p+\nu}}$ mellett.

(i) Most $\varepsilon_{m_k} = \frac{A}{m_k^{p+\nu}} \le \frac{B}{(h^{p+\nu})^k}$ ($B \ge A$ állandó). Mivel $h > 1$, így

$$\tilde{q} := h^{-(p+\nu)} < 1.$$

Legyen $c := \max\{\beta, c_2\}$, $q := \max\{\tilde{q}, q_2\}$, ahol q_2 a (4.8)-ban szereplő kvóciens. Ekkor

$$\varepsilon_{m_k} + c_2 q_2^k \le \beta(\tilde{q})^k + c_2 q_2^k \le c q^k \qquad (k \in \mathbf{N}).$$

Így a kívánt (3.7) becslést (4.3)-ból kapjuk.

(ii) A Szoboljev-féle beágyazási tétel következménye.
(iii) Most

$$\sum_{j=k}^{\infty} \left((j+1)(m_{j+1}+2n)\right)^{N+2r}(\varepsilon_{m_j}+c_2q_2^j) \le$$

$$\le K\sum_{j=k}^{\infty} \left((j+1)h^j\right)^{N+2r}\left(\frac{\beta}{(h^\gamma)^j}+c_2q_2^j\right) \le$$

$$\le K_1\sum_{j=k}^{\infty}(j+1)^{N+2r}(h^{N+2r-\gamma})^j + K_2\sum_{j=k}^{\infty}(j+1)^{N+2r}(h^{N+2r}q_2)^j.$$

(K, K_1, K_2 alkalmas pozitív konstansok.) Itt az $N+2r-\gamma<0$ feltétel épp azt biztosítja, hogy a $h^{N+2r-\gamma}$ kvóciens 1-nél kisebb legyen, így az első tagban szereplő sor konvergens, sőt megadható olyan $K_3>0$, $0<q_3<1$, hogy összege felülről becsülhető $K_3q_3^k$-nal. Ugyanígy becsülhető a másik sor $K_4q_4^k$-nal (ott a h-ra tett feltevésből $h^{N+2r}q_2<1$), így fennáll a 4.3. segédtétel (iii) pontjának feltétele, s ezért (4.5) szerint, $c' := c_4\max\{K_3,K_4\}$, $q' := \max\{q_3,q_4\}$ mellett fennáll a kívánt (3.9) becslés.

(iv) Ugyanúgy igazolható, mint (iii); itt a $h^{\frac{N}{2}+r-\gamma}<1$ feltételt kell felhasználni. □

A 3.2. Tétel bizonyítása (az egységgömbön)

Ugyanúgy történik, mint az előző, a 4.3. segédtételt most $\varepsilon_k := Aq^k$ mellett alkalmazhatjuk. A (iii) pontban

$$\sum_{j=k}^{\infty}\left[(j+1)(m_{j+1}+2n)\right]^{N+2r}(\varepsilon_{m_j}+c_2q_2^j) \le M_1\sum_{j=k}^{\infty}(j+1)^{2(N+2r)}(Aq^j+c_2q_2^j),$$

itt a majoráns sor az előzőekhez hasonlóan becsülhető $c_rq_r^k$-nal. □

(c) A konvergencia bizonyítása egyéb tartományon

Tekintsük az eredeti (2.4) egyenletet és alkalmazzuk a (3.5) transzformációt!

4.3. LEMMA. *Legyen* $T\in C^{2n+1+p,\nu}(\overline{S},\overline{\Omega})$, $k\in\{1,\dots,2n+p\}$. *Ekkor az alábbi normák ekvivalensek:*

(i) $\qquad \|u\|_{H^k(\Omega)}$ *és* $\|\hat{T}u\|_{H^k(S)}$;

(ii) $\qquad \|u\|_{C^k(\overline{\Omega})}$ *és* $\|\hat{T}u\|_{C^k(\overline{S})}$.

(Ha T analitikus, akkor (i) *és* (ii) *tetszőleges $k \in \mathbf{N}$ esetén igaz.)*

Bizonyítás. Legyen először $k = 1$. Az $R := \sqrt{|\det T'|}$ jelöléssel $\hat{T}u := (u \circ T)R$, így $i = 1, \dots, N$ és $u \in H^1(\Omega)$ esetén

$$(4.9)\quad \partial_i(\hat{T}u) = \sum_k (\partial_i T_k)(\partial_k u \circ T)R + (u \circ T)\partial_i R = \sum_k (\partial_i T_k)\hat{T}(\partial_k u) + \frac{\partial_i R}{R}\hat{T}u.$$

Itt $R \in C^{2n+p}(\overline{S})$ és $R \neq 0$ $\overline{S}$-on. Ebből

$$\left\|\partial_i(\hat{T}u)\right\|_{L^2(S)} \leq \max_k \|\partial_i T_k\|_{C(\overline{S})} \sum_k \left\|\hat{T}(\partial_k u)\right\|_{L^2(S)} + \left\|\frac{\partial_i R}{R}\right\|_{C(\overline{S})} \|\hat{T}u\|_{L^2(S)} =$$

$$= \max_k \|\partial_i T_k\|_{C(\overline{S})} \sum_k \|\partial_k u\|_{L^2(\Omega)} + \left\|\frac{\partial_i R}{R}\right\|_{C(\overline{S})} \|u\|_{L^2(\Omega)} \leq \text{const.} \cdot \|u\|_{H^1(\Omega)},$$

amiből már következik, hogy

$$\|\hat{T}u\|_{H^1(S)} \leq \text{const.} \cdot \|u\|_{H^1(\Omega)}.$$

(A másik irány szerepcserével hasonló). Ugyanígy adódik (4.9)-ből

$$\left\|\partial_i(\hat{T}u)\right\|_{C(\overline{S})} \leq \text{const.} \cdot \|u\|_{C^1(\overline{\Omega})},$$

amiből az előbbi módon (ii) következik.

A $k \geq 1$ eset ugyanígy, több számolással igazolható. (A fellépő együtthatófüggvények T simasága miatt C^k-beliek, s a nevezőbe mindig a pozitív alsó korláttal rendelkező R függvény kerül.)

4.4. Következmény. Az $\hat{L}$ operátor 3.b. pontbeli értelmezése miatt tetszőleges $u \in L^2(\Omega)$, $v = \hat{T}u \in L^2(S)$ esetén $\langle \hat{L}v, v\rangle = \langle Lu, u\rangle$. Mivel $\langle Lu, u\rangle$ és $\|u\|_{H_0^n(\Omega)}$ ekvivalensek, így a 4.3. lemma alapján $\langle \hat{L}v, v\rangle$ és $\|v\|_{H_0^n(S)}$ is azok. Emellett $\hat{L}$ együtthatói öröklik L együtthatóinak simaságát, így a transzformált $(3.6)_b$ egyenletre igaz a 3.1, ill. 3.2. tétel.

4.5. Következmény. A transzformált $(3.6)_b$ egyenletre tehát igazak a (3.7–10) becslések az egységgömbön, így pedig a 4.3. lemma szerint a becslések (alkalmas konstansokkal) $(3.6)_a$-ra is teljesülnek. Ezzel a 3.1. és 3.2. tételt igazoltuk.

5. A módszer számítógépes megvalósítása

Amint a 3. szakaszban láttuk, a módszer alapja az, hogy az egységgömbön az iterációs egyenleteket lineáris algebrai egyenletrendszerrel megoldhatjuk. Ebben a szakaszban részletesebben megvizsgáljuk a fellépő egyenletrendszert másodrendű peremfeladat esetén. A gömbre való transzformáció két speciális esetének megadása után összefoglaljuk a módszer előnyeit és hátrányait. Végül egy példán szemléltetjük a módszer alkalmazását.

(a) Az iterációs egyenletrendszer

A gömbre transzformált másodrendű egyenlet közelítésekor olyan

$$\begin{cases} \Delta u = p \\ u|_{\partial S} = 0 \end{cases} \tag{5.1}$$

alakú iterációs egyenleteket kell megoldanunk, amelyekben a p jobboldal polinom. Erre a következő módszer javasolható (l. a 3. szakasz bevezetésében): legyen a p polinom foka n, s keressük u-t (a peremfeltételeket eleve kielégítő)

$$u(x_1,\dots,x_N) = \left(\sum_{i=1}^{N} x_1^2 - 1\right) q(x_1,\dots,x_N)$$

alakban, ahol q is n-edfokú polinom. Ekkor a Δu n-edfokú polinom együtthatói a q együtthatóinak lineáris kombinációi, így q együtthatóit a Δu és p egyenlővé tételéből kapott lineáris egyenletrendszer megoldása adja.

(1) Milyen alakú ez az egyenletrendszer?

Legyen a p n-edfokú polinom

$$p(x_1,\dots,x_N) = \sum_{m=0}^{n} \sum_{k_1+\dots+k_N=m} c_{k_1,\dots,k_N} x_1^{k_1} \cdots x_N^{k_N}.$$

Az említettek szerint (5.1) megoldását a következő alakban keressük:

$$u(x_1,x_2,\dots,x_N) = \left(\sum_{i=1}^{N} x_i^2 - 1\right) q(x_1,x_2,\dots,x_N),$$

ahol

$$q(x_1,\dots,x_N) = \sum_{m=0}^{n} \sum_{k_1+\dots+k_N=m} a_{k_1,\dots,k_N} x_1^{k_1} \cdots x_N^{k_N}$$

meghatározandó n-edfokú polinom. Ekkor a Δu n-edfokú polinomban — rögzített $k := (k_1,\dots,k_n)$ multiindex esetén, ahol $|k| := k_1 + \dots + k_n \le n$ — az $x_1^{k_1} \cdots x_N^{k_N}$

tag együtthatója a q polinom együtthatóinak lineáris kombinációja lesz. A q polinom ismeretlen együtthatóira megoldandó egyenletrendszer tehát azt írja le, hogy ez a kombináció megegyezik $c_{k_1,\dots,k_N}$-nel minden k multiindex esetén, melyre $|k| \le n$.

Rögzített $x \in \mathbf{R}^N$ és $|k| \le n$ esetén legyen $x^k := x_1^{k_1} \cdots x_N^{k_N}$; $|k| \le n-2$ és adott $1 \le i \le N$ index esetén $x_{i+}^k := x_i^2 x^k$; tetszőleges k multiindex és $i \ne j$ indexek esetén pedig

$$x_{i+,j-}^k := \begin{cases} x_i^2 x_j^{-2} x^k & (k_j \ge 2) \\ 0 & (k_j < 2). \end{cases}$$

Jelölje továbbá a_k, a_k^{i+}, $a_k^{i+,j-}$ az x^k, x_{i+}^k, $x_{i+,j-}^k$ hatvány együtthatóját a q polinomban; legyen végül $c_k := c_{k_1,\dots,k_N}$.

A $\Delta u = p$ egyenlőség alapján könnyen látható, hogy u-ban x_{i+}^k együtthatója

$$a_k + \sum_{\substack{j=1 \\ j \ne i}}^{N} a_k^{i+,j-} - a_k^{i+},$$

ebből pedig felírható a keresett egyenletrendszer:

$$a_k \sum_{i=1}^{N} (k_i+2)(k_i+1) + \tag{5.2}$$

$$+ \sum_{i=1}^{N} \left(\sum_{\substack{j=1 \\ j \ne i}}^{N} a_k^{i+,j-} - a_k^{i+} \right) (k_i+2)(k_i+1) = c_k \quad (|k| \le n)$$

Az egyenletrendszer együtthatói *ritka (sparse)* mátrixot alkotnak, a c_k jobboldal esetén ui. csak olyan a_k, ismeretlenek szerepelnek, ahol a k' multiindex legfeljebb két helyen, egyféle lehetséges módon különbözik a k multiindextől. Ez azt mutatja, hogy n növekedésével a nem 0 együtthatók száma korlátos marad, éspedig (5.2) szerint legfeljebb N^2+1 lehet. Másfelől, mivel az említett különbség a multiindexek megfelelő koordinátái között legfeljebb 2 lehet, az együtthatókból képzett mátrix *sávmátrix* lesz az (5.2) rendszert alkotó egyenletek bármely olyan sorrendje esetén, amikor a c_k jobboldalak multiindexeinek rendje növekszik.

(2) Érdemes megvizsgálni egyszerűség kedvéért az $N = 2$ esetet.

Hogyan írható fel ilyenkor $\operatorname{gr} p = n$ esetén az egyenletrendszer $[A_{(n)}]$ mátrixa? A fellépő egyenletek száma ($|k| = k_1 + k_2 \le n$ miatt) $\alpha_n := \frac{(n+1)(n+2)}{2}$, ezek a c_{k_1,k_2} együtthatók multiindexei szerint rendezhetők, ezzel (5.2) jobboldala

$$c^T = (c_{00}, c_{10}, c_{01}, \dots, c_{n0}, c_{n-1,1}, \dots, c_{0n}) \tag{5.3}$$

lesz. Írjuk fel az i-edik egyenletet ($i = 1, 2, \ldots, \alpha_n$)! Itt az i index egyértelműen felírható $i = \frac{m(m+1)}{2} + j$ ($m = 0, 1, \ldots, n$; $j = 1, 2, \ldots, m+1$) alakban, ekkor az i. egyenlet jobboldalán $c_{m-j+1,j-1}$ szerepel.

Bevezetve az alábbi jelöléseket:

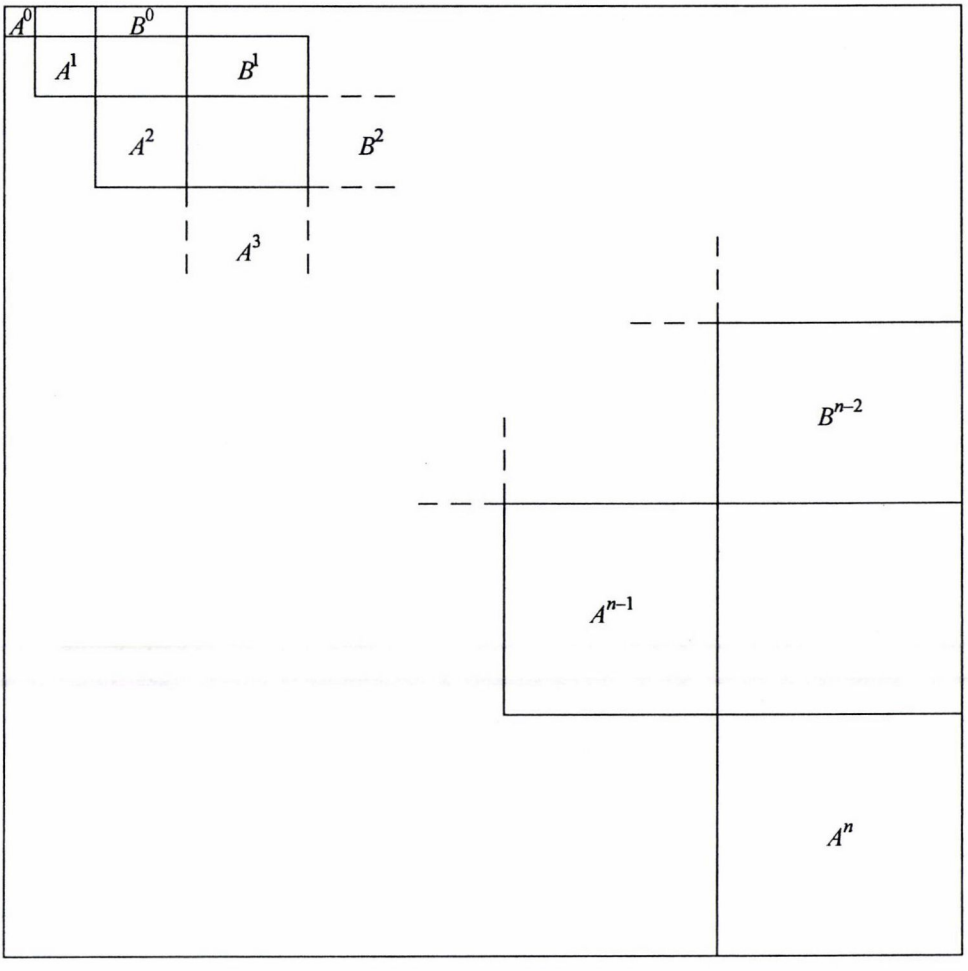
$$
\text{(5.4)} \quad \begin{cases} A^{(m)}_{i,i-2} := (m+2-j)(m+3-j) & (j = 3, 4, \ldots, m+2), \\ A^{(m)}_{i,i} := j(j+1) + (m+2-j)(m+3-j) & (j = 1, 2, \ldots, m+1), \\ A^{(m)}_{i,i+2} := j(j+1) & (j = 1, 2, \ldots, m-1), \\ B^{(m)}_{i,i} := -(m+2-j)(m+3-j) & (j = 1, 2, \ldots, m+1), \\ B^{(m)}_{i,i+2} := j(j+1) & (j = 1, 2, \ldots, m+1), \\ \text{ill. } x_i := a_{m-j+1,j-1} & (j = 1, 2, \ldots, m+1), \\ \quad \hat{c}_i := c_{m-j+1,j-1} & (j = 1, 2, \ldots, m+1), \end{cases}
$$

az i. egyenlet a következő lesz:

(5.5)

$$
\begin{cases} \text{(a)} \quad \text{ha} \quad j = 1, 2: \\ \qquad A^{(m)}_{i,i} x_i + A^{(m)}_{i,i+2} x_{i+2} + B^{(m)}_{i,i} x_{i+2m+3} + B^{(m)}_{i,i+2} x_{i+2m+5} = \hat{c}_i \\ \text{(b)} \quad \text{ha} \quad 3 \le j \le m-1: \\ \qquad A^{(m)}_{i,i-2} x_{i-2} + A^{(m)}_{i,i} x_i + A^{(m)}_{i,i+2} x_{i+2} + B^{(m)}_{i,i} x_{i+2m+3} + B^{(m)}_{i,i+2} x_{i+2m+5} = \hat{c}_i \\ \text{(c)} \quad \text{ha} \quad j = m, m+1: \\ \qquad A^{(m)}_{i,i-2} x_{i-2} + A^{(m)}_{i,i} x_i + B^{(m)}_{i,i} x_{i+2m+3} + B^{(m)}_{i,i+2} x_{i+2m+5} = \hat{c}_i. \end{cases}
$$

Tehát az $[A_{(n)}] \in \mathbf{R}^{\alpha_n \times \alpha_n}$ mátrix az alábbi alakú:



ahol rögzített $0 \leq m \leq n$ esetén az A^m, ill. B^m mátrix (j_1, j_2)-edik eleme (5.4) jelöléseivel $A^{(m)}_{i_1,i_2}$, ill. $B^{(m)}_{i_1,i_2}$ $(j_r := i_r - \frac{m(m+1)}{2},\ r = 1,2)$. Így (5.5) szerint az $A^m \in \mathbf{R}^{(m+1)\times(m+1)}$ blokk $m > 1$ esetén tridiagonális, a $B^m \in \mathbf{R}^{(m+1)\times(m+3)}$ blokk pedig 2 átlóból áll.

Pl. $n = 4$ esetén

$$[A_{(4)}] = \begin{array}{|cccccccccccccccc|}
\hline
4 & & -2 & & -2 & & & & & & & & & & & \\
 & 8 & & & & & -6 & & -2 & & & & & & & \\
 & & 8 & & & & & -2 & & -6 & & & & & & \\
\cline{3-11}
 & & & 14 & & 2 & & & & & -12 & & -2 & & & \\
 & & & & 12 & & & & & & & -6 & & -6 & & \\
 & & & 2 & & 14 & & & & & & & -2 & & -12 & \\
\cline{5-13}
 & & & & & & 22 & & 2 & & & & & & & \\
 & & & & & & & 18 & & 6 & & & & & & \\
 & & & & & & 6 & & 18 & & & & & & & \\
 & & & & & & & 2 & & 22 & & & & & & \\
\cline{9-16}
 & & & & & & & & & & 32 & & 2 & & & \\
 & & & & & & & & & & & 26 & & 6 & & \\
 & & & & & & & & & & 12 & & 24 & & 12 & \\
 & & & & & & & & & & & 6 & & 26 & & \\
 & & & & & & & & & & & & 2 & & 32 & \\
\hline
\end{array}$$

Az $[A_{(n)}]$ mátrixoknak a sávos szerkezet mellett további előnyös tulajdonsága, hogy egymásba skatulyázottak, azaz $0 \le k \le n$ esetén az $[A_{(k)}] \in \mathbf{R}^{\alpha_k \times \alpha_k}$ mátrix az $[A_{(n)}] \in \mathbf{R}^{\alpha_n \times \alpha_n}$ mátrixnak α_k-adik főminorja. Ezért, ha n lépésig akarjuk közelíteni egyenletünket, akkor csak $[A_{(n)}]$-et kell meghatározni, és minden lépésben a megfelelő főminort használhatjuk az (5.2) egyenlet felírásához.

Végül érdemes megjegyezni, hogy a főátló alatti elemeket fölülről lefelé sorban eltüntethetjük, s így $\alpha_n - 2n - 1$ lépésben felsőháromszög-mátrixot kapunk, melynek soraiban szintén legfeljebb 5 nemzérus elem szerepel, s ezek továbbra is az $[A_{(n)}]$-ben talált sávban helyezkednek el.

(b) Példák gömbre való transzformációra

A 3. szakasz szerint ahhoz, hogy a (2.4) egyenletre az ismertetett gradiensmódszert alkalmazhassuk, az egyenletet az egységgömbre kell transzformálnunk. Ehhez arra van szükség, hogy explicit alakban ismerjünk egy $T : \overline{S} \to \overline{\Omega}$ bijekciót, melyre $\det T' \neq 0$ $\overline{S}$-on. Az alábbi két esetben megmutatjuk, hogy ha ismerjük a tartomány peremét alkalmas módon leíró $C^{k,\nu}$-beli, ill. analitikus függvényt, akkor a megkívánt simaságú bijekcióra egyszerű formula adható.

1. Legyen $\Omega \subset \mathbf{R}^N$ olyan csillagtartomány, melynek $\partial\Omega$ határa $C^{k,\nu}$-beli.

Feltehető, hogy az Ω tartomány 0-ra nézve csillagszerű és tartalmazza $\overline{S}$-et (hiszen ha nem, egy eltolással és nagyítással segíthetünk ezen). Legyen $\partial\Omega$ $C^{k,\nu}$-beli. Ekkor — áttérve $(r, \varphi_1, \dots, \varphi_{N-1})$ polárkoordinátákra és a $\varphi := (\varphi_1, \dots, \varphi_{N-1})$ jelölést használva — a határ pontjai előállíthatók $r = d(\varphi)$ alakban, ahol $d \in C^{k,\nu}(\mathbf{R}^{N-1})$ pozitív értékű, $[0, \pi)^{N-2} \times [0, 2\pi)$-periodikus függvény; az $\Omega \supset \overline{S}$ feltétel miatt $d > 1$.

A d függvény segítségével megadunk egy $T : \overline{S} \to \overline{\Omega}$ $C^{k,\nu}$ simaságú bijekciót, melyre $\det T' \neq 0$.

A $T : \overline{S} \to \overline{\Omega}$ bijekció a 0-t hagyja helyben, 0-t kivéve pedig polárkoordinátákkal adjuk meg és ezt θ-val jelöljük. Éspedig, $r \in (0,1]$ és $\varphi \in [0,\pi)^{N-2} \times [0,2\pi)$ esetén legyen

$$\theta(r,\varphi) := \big(F(r,\varphi),\varphi\big)$$

(azaz sugárirányú nagyítást végzünk), ahol

$$F(r,\varphi) := \begin{cases} r, & \text{ha } 0 < r \leq 1/2 \\ r + \big(d(\varphi) - 1\big)(2r-1)^{k+1}, & \text{ha } 1/2 < r \leq 1 \text{ és } \varphi \text{ tetszőleges.} \end{cases}$$

Mivel tetszőleges φ esetén $r \mapsto F(r,\varphi)$ szigorúan növő és $F(1,\varphi) = d(\varphi)$, ezért látható, hogy θ valóban egy $T : \overline{S} \to \overline{\Omega}$ bijekciót határoz meg.

Az $r \mapsto F(r,\theta)$ függvény definíciójából az is következik, hogy $\det \theta'(r,\varphi) = \partial_r F(r,\varphi) \neq 0$ ($r \neq 0$, φ tetszőleges). Könnyen látható továbbá, hogy F, és így θ is örökli d simaságát, azaz $C^{k,\nu}$-beli. Mivel az utóbbi két tulajdonságot a polártranszformáció megőrzi, a 0 egy környezetében pedig T az identitás, így teljesül a T-re megkívánt két további tulajdonság is: $T \in C^{k,\nu}$ és $\det T' \neq 0$ $\overline{S}$-on.

2. Legyen $\Omega \subset \mathbf{R}^2$ és $\Gamma := \partial\Omega$ analitikus görbe. Keresendő olyan $T : \overline{S} \to \overline{\Omega}$ analitikus bijekció, melyre $\det T' \neq 0$.

A kívánt T transzformáció mindig létezik. Ugyanis, ha az S és Ω síkbeli halmazokat $\mathbf{C}$ részhalmazaiként tekintjük, akkor [3] 49a tétele alapján van olyan $f : S \to \Omega$ konform leképezés, amely ∂S-et Γ-ba viszi és analitikusan kiterjeszthető egy $\overline{S}$-at tartalmazó nyílt halmazra. Emiatt a $T : \overline{S} \to \overline{\Omega}$,

$$T(x,y) := \big(\operatorname{Re} f(x+iy), \operatorname{Im} f(x+iy)\big)$$

leképezés megfelel a kívánalmaknak: mivel f komplex analitikus bijekció, ezért T valós analitikus bijekció, és mivel [3] 51. tétele alapján $f' \neq 0$ $\overline{S}$-on, ezért $\det T' = |f'|^2 \neq 0$ $\overline{S}$-on.

T képlettel való megadására pl. az alábbi esetben nyílik lehetőség.

Legyen $\gamma : [0,2\pi] \to \mathbf{C}$ olyan görbe, amely a

$$\gamma(t) = \sum_{n=-\infty}^{\infty} a_n e^{\mathrm{i}nt}$$

Fourier-sor összegeként áll elő. Ha valamely $c > 0$ és $q \in (0,1)$ mellett $a_{-n} = 0$ és $|a_n| \leq cq^n$ ($n \in \mathbf{N}^+$), akkor $\Gamma = R(\gamma)$ analitikus. Emellett az

$$f(z) := \sum_{n=0}^{\infty} a_n z^n \tag{5.6}$$

komplex függvény, amely ekkor analitikus az origó közepű, $1/q$ sugarú nyílt körlapon, ∂S-et Γ-ba viszi. Az f által (5.6) szerint meghatározott T leképezés tehát megfelel a célnak, ha f bijekció és $f' \neq 0$ $\overline{S}$-on. (Ekkor ugyanis $\overline{S}$ T-képe csak $\overline{\Omega}$ lehet, s a fentiek szerint így analitikus bijekciót kapunk $\overline{S}$ és $\overline{\Omega}$ között, melyre $\det T' \neq 0$ $\overline{S}$-on). Ehhez pl. egyszerű elégséges feltételt ad meg a fenti konstansokra tett $|a_1| > \sigma$ kikötés, ahol $\sigma := c\sum_{n=2}^{\infty} nq^n = c(2-q)q^2(1-q)^{-2}$. Ekkor ugyanis bármely $z, w \in \overline{S}$ esetén

$$|f(z) - f(w)| = \left| \sum_{n=1}^{\infty} a_n (z-w) \sum_{k=0}^{n-1} z^k w^{n-1-k} \right| =$$

$$= \left| a_1 + \sum_{n=2}^{\infty} a_n \sum_{k=0}^{n-1} z^k w^{n-1-k} \right| |z-w| \geq \left(|a_1| - c\sum_{n=2}^{\infty} nq^n \right) |z-w| = \eta |z-w|$$

(ahol $\eta = |a_1| - \sigma > 0$), így f bijekció és $\forall\, z \in \overline{S}$ esetén $|f'(z)| \geq \eta > 0$.

(c) A módszer előnyei és hátrányai

(a) Előnyök

1. A módszer egyszerűen *algoritmizálható*. A megoldandó egyenletet approximációs polinomokkal pótolt egyenletekkel közelítjük. Az approximált egyenletekre alkalmazzuk a gradiens-módszert, s ekkor az iteráció minden lépése a következő két részből áll. Ha ismerjük az u_n közelítést, akkor

(i) meghatározzuk a $g_n := Lu_n - f$ polinomot, ami polinomok összeadását, szorzását és deriválását jelenti;

(ii) g_n jobboldallal megoldjuk a $\Delta z_n = g_n$, $z_n|_{\partial S} = 0$ feladatot, ami egy lineáris algebrai egyenletrendszer megoldásából áll; a kapott z_n megoldásból az $u_{n+1} := u_n - \frac{2}{m'+M'} z_n$ formula adja meg a következő közelítést. (Itt m' és M' az approximált operátorok közös határait jelölik).

2. A fentiek következtében a módszer numerikusan egyszerűen realizálható. A polinomok szorzása és deriválása nem okoz nehézséget, s az egyenletrendszerek megoldása is — minthogy a fellépő mátrixok *ritka (sparse)* és *sávos struktúrájú* mátrixok — könnyen elvégezhető.

3. A módszer segítségével a közelítéseket *folytonos alakban,* azaz a megoldási tartományon értelmezett formulával állítjuk elő, éspedig *polinom* formájában. Ennek a diszkretizációs eljárásokkal (véges differencia-, végeselem-módszer) összehasonlítva az alábbi előnyei vannak:

— Jobban kihasználja a tartomány alakját. A diszkretizációs módszerek ugyanis különösen téglalap (téglatest) alakú vagy sokszög (poliéder) által határolt tartományon alkalmazhatók jól. Esetünkben azonban — vagyis *gömbön* — a tartomány rácsfelbontásánál nehézségek lépnek fel, a tartomány polár típusú transzformációjakor pedig az egyenletben jelenik meg szingularitás.

— A tartomány minden pontjában a közelítő függvényérték egyszerűen meghatározható.
— Ha a kapott közelítő megoldással további számításokat kell végezni, ezek polinomokkal egyszerűen elvégezhetők (differenciálás és integrálás helyett például lineáris kombináció).

4. Amint láttuk, elég sima együtthatók esetén a közelítések megfelelő deriváltjaikkal együtt egyenletesen konvergálnak az u^* megoldáshoz és annak megfelelő deriváltjaihoz. Ha ez fennáll a második deriváltakig bezárólag (ehhez elég a $p \geq N$ feltétel), akkor az u^* megoldás alakjának megőrződését kapjuk. Ha ugyanis pl. $u^* > 0$ (vagy $\partial_i u^* > 0$, ill. $D^2 u^*$ pozitív definit) $\overline{\Omega}$-on, akkor egy index után ez igaz lesz az u_n közelítésekre is. Ezért ha a megoldás pozitív (vagy az i. változóban növekedő, ill. konvex), akkor ez a (megfelelő pontosságú) közelítésekre is öröklődik.

(A kvalitatív tulajdonságok ilyen típusú megmaradását más feladattípus, ill. módszer során is vizsgálták e lap hasábjain [6].)

b) Hátrányok

1. A módszer csak egyszerűen gömbre transzformálható tartományon alkalmazható.

2. Az egyenletrendszerekben fellépő mátrixok sávszélessége nagyobb, mint pl. a véges differenciák módszerében.

3. A nagyobb sávszélességek miatt a szükséges műveletigény is nagyobb.

(d) Egy példa

A modellfeladat

$$\text{(5.7)} \qquad \begin{cases} \partial_x\left(e^{\frac{x+y}{2}}\partial_x u\right) - \partial_y\left(e^{\frac{x+y}{2}}\partial_y u\right) = 4 - x - y \text{ az } S \subset \mathbf{R}^2 \text{ egységkörön} \\ u|_{\partial S} = 0. \end{cases}$$

Legyen $a(x,y) := e^{\frac{x+y}{2}}$, valamint $f(x,y) := 4 - x - y$ $(x, y \in \overline{S})$. Itt

$$m := \min a = e^{-1/\sqrt{2}} \quad \text{és} \quad M := \max a = e^{1/\sqrt{2}}.$$

Az m és M számok az

$$Lu := -\sum_{i=1}^{2} \partial_i(a\partial_i u) = f$$

operátor diagonális együtthatómátrixának határai (l. 2.1. megjegyzés) és egyben a 2.3. tételben szereplő m és M állandók is.

Az a együtthatófüggvény analitikus, így a közelítő polinomokat Taylor-sorának szeleteiből vehetjük:

$$a^{[n]}(x,y) := \sum_{k=0}^{n} \frac{(x+y)^k}{2^k k!}.$$

A közelítés pontosságaként pl. az

$$\|a^{[n]} - a\|_{C(S)} \le e^{1/\sqrt{2}} \cdot \left(\frac{1}{4}\right)^n \qquad (n \in \mathbf{N})$$

becsléssel számolhatunk. Így ha $n \ge 2$, akkor az

$$L^{[n]}u := -\sum_{i=1}^{n} \partial_i\left(a^{[n]}\partial_i u\right)$$

operátor alsó határa már pozitív:

$$0 < m' := e^{-1/\sqrt{2}} - e^{1/\sqrt{2}} \cdot \frac{1}{16} \le e^{+1/\sqrt{2}} + e^{1/\sqrt{2}} \cdot \frac{1}{16} =: M'.$$

Ebből következik, hogy ha az iterációt a

$$t := \frac{2}{m' + M'} = \frac{1}{ch\left(1/\sqrt{2}\right)} = 0.7933$$

lépéshosszal végezzük, akkor a konvergencia elvi kvóciense

$$\frac{M' - m'}{M' + m'} = 0.7094$$

lesz.

A numerikus kísérletben $a(x,y)$ helyett az $a^{[3]}(x,y)$ közelítést helyettesítettük. Az $u_0 := 0$ kiindulási függvénnyel végeztük el a gradiens-módszer által definiált iterációt: ha megvan u_n, akkor

$$\begin{cases} z_n \quad \text{a} \quad -\Delta z_n = L^{[3]}u_n - f,\ z_n|_{\partial S} = 0 \quad \text{feladat megoldása,} \\ u_{n+1} := u_n - 0.7933 \cdot z_n. \end{cases}$$

Az Lu_n függvények kiszámításához a z_n és u_n polinomokat együtthatómátrix alakjában tároltuk, így a parciális deriváltak és szorzatok egyszerűen meghatározhatók. Az iterációs egyenletek megoldásához a polinomok együtthatóit az (5.3)-ban megadott oszlopvektorba rendeztük. Az algoritmust (a lineáris algebrai egyenletrendszerek megoldását is beleértve) a MATLAB programcsomaggal futtattuk le.

Az (5.7) feladat pontos megoldását ismerjük:

$$u^*(x,y) = e^{-\frac{x+y}{2}}(1 - x^2 - y^2).$$

Ez lehetővé teszi a közelítő sorozat tényleges hibájának meghatározását.

A

$$h_n := \|u^* - u_n\|_{H_0^1(S)}$$

hibákat alkalmas numerikus kvadratúrával határoztuk meg (közelítőleg). Emellett a 3.2. tétel szerint a sorozat egyenletesen is konvergál, ezért kiszámítottuk az

$$e_n := \max_{\overline{S}} |u^* - u_n|$$

hibákat is. (Ezeket szintén közelítőleg, polárkoordinátás rácsháló n határoztuk meg.) Az iterációt az $e_n < 0.01$ pontossági kritériumig végeztük. A h_n-re és e_n-re kapott értékeket az alábbi táblázat tartalmazza.

n	1	2	3	4	5	6	7
h_n	3.0291	1.2042	0.5791	0.2919	0.1487	0.0740	0.0332
e_n	1.1250	0.4204	0.1948	0.0926	0.0455	0.0213	0.0090

IRODALOM

[1] Adams, R. A., *Sobolev spaces* (Academic Press, New York–London, 1975).

[2] Agmon, S., Douglis, A., Nirenberg, L., Estimates near the boundary for solutions of elliptic partial differential equations satisfying general boundary conditions, *Communications on Pure and Applied Mathematics* **12** (1959) 623–727.

[3] Behnke, H., Sommer, F., *Theorie der analytischen Funktionen einer komplexen Veranderlichen* (Springer–Verlag, 1972).

[4] Cea, J., *Lectures on optimization — theory and algorithms* (Springer–Verlag, 1978).

[5] Daniel, J. W., The conjugate gradient method for linear and nonlinear operator equations, *SIAM Journal on Numerical Analysis* **4** (1967) 10–28.

[6] Faragó I., Haroten H., Komáromi N., Pfeil T., A hővezetési egyenlet és numerikus megoldásának kvalitatív tulajdonságai I–II., *Alk. Mat. Lapok* **17** (1993) 101–121 és 123–141.

[7] Kantorovics, L. V., On an effective method of solution of extremal problems for quadratic functionals, *Dokl. Akad. Nauk. SSSR* **48** (1945) 483–487.

[8] Kantorovich, L. V., Akilov, G. P., *Functional Analysis* (Pergamon Press, 1982).

[9] Karátson, J., On uniform convergence of polynomials converging in L^p norm, *Annales Univ. Sci. Budapest* **37** (1994) 291–298.

[10] Simon L., Baderko, E. A., *Másodrendű lineáris parciális differenciálegyenletek* (Tankönyvkiadó, Budapest, 1983).

[11] Tyiman, A. F., *Tyeorija priblizsenyija funkcij gyejsztvityelnovo peremennovo* (Goszudarsztvennoje Izdatyelsztvo, Moszkva, 1960).

[12] Wouk, A., *A Course of Applied Functional Analysis* (J. Wiley & Sons, 1979).

[13] Zeidler, E., *Vorlesungen über nichtlineare Funktionalanalysis* (Teubner-Texte zur Mathematik, Leipzig, 1977).

(Beérkezett: 1995. március 17.)

KARÁTSON JÁNOS
ELTE, ALKALMAZOTT ANALÍZIS TANSZÉK
1088 BUDAPEST, RÁKÓCZI ÚT 5.
E-mail: karatson@cs.elte.hu

GRADIENT METHOD IN SOBOLEV SPACES: APPROXIMATE SOLUTION OF LINEAR BOUNDARY VALUE PROBLEMS USING POLYNOMIALS

JÁNOS KARÁTSON

The subject of the paper is the application of the Hilbert space version of the gradient method (GM) to linear elliptic boundary value problems of $2n$-th order. In this setting the GM is applied to the generalized differential operator in the corresponding Sobolev space, thus reducing the original problem to auxiliary Poisson equations. The principle of realization is based on László Czách's method in the case of domains that can be transformed to a ball. Namely, the approximating sequence is constructed to consist of polynomials. This yields that the solution of the auxiliary equations can be achieved by solving systems of linear algebraic equations (SLAE-s) with sparse matrices. The method yields linear convergence in Sobolev norm, moreover, for smooth enough coefficients we have uniform convergence. Numerical performance is investigated with the two-dimensional case in focus. The method is easy to realize, requiring the solution of SLAE-s of simple structure, and yields linear convergence.

Alkalmazott Matematikai Lapok **19** (1999) 29–38

NAGYSEBESSÉGŰ INFORMATIKAI HÁLÓZAT ADATFORGALMÁNAK MATEMATIKAI STATISZTIKAI JELLEMZÉSE*

GÁL ZOLTÁN, IGLÓI ENDRE ÉS DR. TERDIK GYÖRGY

Debrecen

A kommunikációs hálózatok modellezésének és a matematikai statisztikának is aktuális problémája a nagysebességű hálózatok tanulmányozása. Ebben a cikkben a Debreceni Universitas Egyesülés FDDI gyűrűje routereinek forgalmi adatait az idősoranalízis módszereivel vizsgáljuk. A szezonális komponensek kivonása után az idősorok mindegyike a hosszú memóriájú folyamatok jellegzetes tulajdonságát mutatja. A memória paraméterekre három különböző módszerrel is becslést adunk.

1. Az FDDI gyűrű

Debrecenben az Universitas Egyesülés intézményeinek helyi adatátviteli hálózatait egy homogén, nagysebességű optikai hálózat kapcsolja össze. Ez a világon jelenleg legelterjedtebb, FDDI (Fibre Distributed Data Interface) 100 Mbps átviteli sebességű MAN technológia. Erre a gyűrű topológiájú városi hálózatra pillanatnyilag hat darab CISCO AGS+/4 típusú router berendezés segítségével kapcsolódnak az intézmények (KLTE: Kossuth Lajos Tudományegyetem, ATOMKI: MTA Atommagkutató Intézete, DATE: Debreceni Agrártudományi Egyetem, DOTE: Debreceni Orvostudományi Egyetem, DRK: Debreceni Református Kollégium, MFK: KLTE Műszaki Főiskolai Kar) helyi, csillag topológiájú optikai Ethernet hálózatai. Az intézmények többségének van bizonyos saját erőforrás gépparkja, ami sajnos az igényeket nem képes kielégíteni, többek között ezért is gyakori, hogy a felhasználók a városi hálózat segítségével a szomszédos intézmények számítógépes szolgáltatásait is jelentős mértékben igénybe veszik. Az interaktív terminálhasználat, a fájlátvitel, az elektronikus levelezés, könyvtári keresőrendszerek, valamint az Internet látványos grafikus szolgáltatásai mind olyan alkalmazások, amelyek a debreceni városi hálózatot egyre inkább adatok átvitelével terhelik le. Lehetőségünk van a gerinchálózati eszközök interfészein áthaladó keretek számának időegységenkénti mérésére és letárolására, ill. az FDDI gyűrűre kapcsolódó router berendezések procesz-

* Ez a munka az OTKAT019501 támogatásával készült.

szorainak időegységenkénti terhelését mintavételezni egy erre a célra fejlesztett, Cabletron Spectrum nevű menedzsment szoftverrel. Jelen esetben a mintavételezési időköz 10 perc.

A használt CISCO routerek egyprocesszorosak, így az interfészek között routolt minden egyes csomag processzálását a router processzora végzi. Minden routernek van egy FDDI interfésze és négy, hat vagy nyolc Ethernet interfésze. Van néhány olyan interfész, amelyeken több IP network is definiálva van, ezek száma viszont kicsi. A routerek nemcsak IP csomagok routolását, hanem IPX és DECnet csomagokat is processzálnak. Ezen túlmenően a nem routolható protokollokat, mint például a LAT (Local Area Transfer), hidalják.

A router adott interfészére kapcsolt, ugyanazon network két gépe közötti forgalom nem jelenik meg a routerben, így annak processzorát sem terheli. Egy csomópont akkor küld a routernek csomagokat, ha routing és címtábláját kell frissítenie, ill. amikor más networkhez címzi a csomagját. A routerek a csomópontok forgalmától függetlenül a hálózati kapcsolatokra vonatkozó adminisztrációs, ú.n. routing információkat is küldenek egymásnak. Ez a típusú adminisztrációs forgalom lényegesen kisebb az adatforgalomnál, viszont ez is a routerek processzorait terheli.

A csomópontok közötti forgalom kétfajta. Az egyik a batch jellegű adatforgalom, ami a levelező szerverek közötti forgalomból, valamint az adatbankok tükrözését végző szerverek közötti adatforgalomból származik. A másik az interaktív jellegű adatforgalom. Ezt a felhasználói kliens gépek közötti, ill. a kliens-szerver adatcsere okozza, miközben a felhasználók a hálózati alkalmazásokat használják. Ilyenek például a terminálemulació, a WWW használat és az esetleges fájlátvitelek, valamint a WWW szerverek cache funkciójának biztosítása.

Megállapítható, hogy a router processzorának terhelése függ egyrészt az intézményi belső hálózatos forgalomtól, másrészt az adott intézmény városi forgalmától.

2. Az idősorok analízise

2.1. Szezonalitás

Az analízisünk alapjául szolgáló öt idősor az ATOMKI, a DATE, a DOTE, a DRK, és az MFK routerek processzorainak a megfigyelési időpontsorozatokban mért pillanatnyi terhelési adatai. A KLTE adatai technikai okokból hiányoznak (mivel a menedzsment szerver a többi gerinchálózati eszköz — repeater és bridge — portjai forgalmi adatainak gyüjtése miatt túlterhelt volt). Az adatok százalékban vannak megadva, mégpedig egész értékre kerekítve. Az 1. és a 2. ábrán az ATOMKI ill. a DATE idősor látható.

Látszik, hogy nincs trend, ami várható is volt. Így a tulajdonképpeni első lépés az egyes idősorokon belüli függőségi viszonyok tanulmányozása, vagyis a tapasztalati autokorreláció sorozatok kiszámítása. A 3. ábrán az ATOMKI-é található, az első 1000 lépésig. Észrevehető az egy napos, $24 \times 6 = 144$ periódusú (a megfigyelési időköz 10 perc, azaz 1/6 óra), és az egy hetes, $7 \times 24 \times 6 = 1008$ periódusú

szezonális komponens. Ez látható a 4. ábrán. Ezeket a determinisztikus periódikus függvényeknek tekintett komponenseket (ill. a maradékosztály átlagolásos módszerrel becsült értéküket) kivontuk az idősorokból, és maradékokként kaptuk a most már stacionárius sorozatoknak tekinthető öt idősort. A továbbiakban ezekkel foglalkozunk.

2.2. A hosszú memóriájúság és a memória paraméter becslése

2.2.1. A hipotézis felállítása. A szezonalitás eltávolítása után visszamaradó sorozat tekinthető zajnak is. Ebben az esetben azt várnánk, hogy a klasszikus Box–Jenkins modell szerinti autoregressziós mozgó átlag (ARMA) sorozatról van szó, aminek a spektrálsűrűségfüggvénye (a továbbiakban spektrum) korlátos. Ezért a következő lépésben kiszámítottuk a zaj idősorok periodogramjait. A periodogram a spektrum becslése. Az 5. ábrán az ATOMKI zaj periodogramja látható. A többi négy is hasonlóan jól mutatja az

$$f(\lambda) \sim c|\lambda|^{-2\delta}, \qquad \lambda \to 0 \tag{1.1}$$

(c pozitív konstans) tulajdonságot, ami a hosszú memóriájú ú.n. FARMA (Fractional ARMA) idősorok spektrumaira jellemző.

Egy stacionárius idősort hosszú memóriájúnak szokás nevezni, ha az autokovariancia sorozata, $R(k)$, $k = 0, 1, 2, \ldots$, olyan lassan csökken, ha $k \to \infty$, hogy $\sum |R(k)|$ divergens. Ez ekvivalens azzal, hogy az $f(\lambda)$, $\lambda \in [-\pi, \pi]$ spektrum nem korlátos. Az irodalomban és a gyakorlatban szinte kizárólag csak olyan hosszú memóriájú folyamat fordul elő, amely spektrumának pólusa $\lambda = 0$-ban van és az $|\lambda|^{-2\delta}$ rendű. A $\delta \in \left[0, \frac{1}{2}\right)$ paramétert a folyamat memória paraméterének nevezzük. Egy ilyen idősor jól közelíthető FARMA folyamattal, amelynek előállítása

$$x_t = (I - B)^{-\delta} y_t, \quad t \in \mathbf{Z} \tag{1.2}$$

ahol y_t egy véges rendű ARMA folyamat, I az identitás és B az eltolás (Back-shift) operátor. (1.2)-ből az x_t folyamat spektruma

$$f_x(\lambda) = \left|1 - e^{-i\lambda}\right|^{-2\delta} f_y(\lambda), \quad \lambda \in [-\pi, \pi], \quad \lambda \neq 0, \quad \text{ha} \quad \delta > 0. \tag{1.3}$$

Az y_t folyamat spektruma, $f_y(\lambda)$ folytonos $[-\pi, \pi]$-n és pozitív, ezért igaz (1.1). A $\delta = 0$ az ARMA esetnek felel meg, ekkor a spektrum korlátos, az autokovariancia sorozat pedig a végtelenben exponenciálisan csökken. Ezért a $\delta = 0$ esetben a folyamat rövid memóriájú, a $0 < \delta < \frac{1}{2}$ esetben hosszú memóriájú. Tehát a stacionárius ARMA modell rövid memóriájú és a FARMA modell speciális, elfajult esetének tekinthető.

Itt kell megjegyeznünk, hogy a hosszú memóriájú folyamatokra sok szempontból egész más törvények érvényesek, mint a rövid memóriájúakra, tehát az idősoranalízis klasszikus módszereivel óvatosnak kell lenni (ld. [1]). Erre a jelenségre egy egyszerű példa a következő alpontban található (2.1) összefüggés.

Tehát a hipotézisünk az, hogy az idősorok, azaz mostmár a zaj sorozatok hosszú memóriájúak, pontosabban hosszú memóriájú FARMA folyamatok. A továbbiakban néhány, ennek a hipotézisnek az eldöntésére használható módszert fogunk alkalmazni. Nem célunk a modellek (a sztochasztikus differenciaegyenletek ill. az ARMA részek) pontos megadása, csak a hosszú memóriájúság és a memóriaparaméter vizsgálata. Az alábbi három módszernek éppen az a közös előnye, hogy a modell pontos ismerete nélkül is tesztelhető velük a hosszú vagy rövid memóriájúság, ill. becsülhető a memóriaparaméter.

2.2.2. Szórásnégyzet-idő grafikon. A módszer ([1], 92. o., [6], 75. o.) lényege, hogy a δ memóriaparaméterű hosszú memóriájú idősorokra a rövid memóriájúaktól eltérően az $\overline{x}_m$ szórásnégyzete m^{-1}-nél lassabban konvergál 0-hoz:

$$D^2\overline{x}_m \sim cm^{2\delta-1}, \qquad m \to \infty, \tag{2.1}$$

c egy pozitív konstans. Logaritmust véve,

$$\log D^2\overline{x}_m \approx \log c + (2\delta - 1)\log m, \qquad m \to \infty.$$

Szedjük szét az idősort k darab m hosszúságú diszjunkt részre, és becsüljük $D^2\overline{x}_m$-ot a k darab rész átlagainak $s_k^2(\overline{x}_m)$ tapasztalati szórásnégyzetével! Csináljuk meg ezt több (m,k) párra, úgy, hogy m is és k is nagy legyen, és ábrázoljuk $\log s_k^2(\overline{x}_m)$-ot $\log m$ függvényeként! Példaként a DRK megfelelő függvénye látható a 6. ábrán.

A folytonos vonal $\delta = 0$-nak felel meg. Rövid memóriájú idősor esetén a csillagos és a folytonos vonal párhuzamos lenne. A módszer használható a δ becslésére is (bár a $\hat{\delta}$ becslés aszimptotikus viselkedése elméletileg még nem tisztázott). Ugyanis $\hat{\delta}$ a csillagokból álló „egyenes" legkisebb négyzetek módszerével becsült meredeksége.

	ATOMKI	DATE	DOTE	DRK	MFK
$\hat{\delta}$	0.28	0.13	0.26	0.42	0.35

δ szórásnégyzet-idő módszerrel való becslései

2.2.3. R/S módszer. A módszer ([1], 33. o.) a tapasztalati szórással (**S**tandard deviation) való osztással skálainvariánssá tett „adjusted **R**ange" statisztikát alkalmazza, innen származik a neve. A hosszú memóriájú folyamatok felfedezője, Hurst arra használta, hogy kiszámítsa, mekkora kapacitású víztároló kellene a Nílus vízhozamának egyenletessé tételéhez. A becslési eljárást ezen a példán keresztül vázoljuk.

Vizsgáljuk a következő, diszkrét idejű stacionárius rendszert! Tegyük fel, hogy a víztárolóból való egyenletes sebességű kifolyáson kívül nincs vízveszteség (pl. párolgás)! Ha a víztároló szintje a t-edik évben ugyanakkora, mint a $t+n$-edik évben,

és x_i az i-edik évben befolyt vízmennyiség, továbbá

$$y_j = \sum_{i=1}^{j} x_i$$

a j-edik évig összesen befolyt vízmennyiség, akkor

$$R(t,n) = \max_{0\le i\le n}\left[y_{t+i} - y_t - \frac{i}{n}(y_{t+n} - y_t)\right] - \min_{0\le i\le n}\left[y_{t+i} - y_t - \frac{i}{n}(y_{t+n} - y_t)\right]$$

az a minimális kapacitás, ami ahhoz kell, hogy ne csorduljon túl és ne ürüljön ki a víztároló a t és $t+n$ időpontok között. Ez az „adjusted range". Jelöljük $S(t,n)$-nel az $x_{t+1},\dots,x_{t+n}$ tapasztalati szórását! Hurst ábrázolta a $\log\big(R(t,n)/S(t,n)\big)$-eket a $\log n$ függvényében, különböző n-ekre és t-kre, és észrevette, hogy nagy n-ekre

$$\log E\left(\frac{R(t,n)}{S(t,n)}\right) \sim a + H\log n, \quad \text{ahol} \quad \frac{1}{2} < H < 1, \tag{3.1}$$

és E most a t-kre való átlag képzést jelenti, minden egyes rögzített n-re. Ez a tulajdonság egyértelműen jellemzi a hosszú memóriájú folyamatokat, ugyanis rövid memóriájúakra

$$E\left(\frac{R(1,n)}{S(1,n)}\right) \sim c\,n^{\frac{1}{2}}, \qquad n\to\infty,$$

c egy pozitív konstans ([5]). Itt meg kell jegyeznünk, hogy a memória paraméterre kétféle jelölés szokásos, δ és a Mandelbrot által bevezetett, Hurst-re emlékeztető $H = \delta + \frac{1}{2}$.

A fenti módszert az MFK idősorra a 7. ábra szemlélteti. Minden egyes n-re 10 darab csillag jelenti a tíz féle t értékhez tartozó $\log(R/S)$ értéket.

Becslést is kaphatunk δ-ra, ha a csillagokból álló ponthalmazra legkisebb négyzetes becsléssel egyenest illesztünk, ugyanis (3.1) szerint az egyenes meredeksége $\hat{H} = \hat{\delta} + \frac{1}{2}$. A 7. ábrán látható egyenesek a két szélsőséges esetnek, a $\delta = 0$-nak és a $\delta = \frac{1}{2}$-nek felelnek meg.

	ATOMKI	DATE	DOTE	DRK	MFK
$\hat{\delta}$	0.36	0.18	0.23	0.4	0.32

δ R/S módszerrel való becslései

2.2.4. Periodogramot használó legkisebb négyzetes becslés. Az először [2]-ban vizsgált módszer a spektrum (1.1) és (1.3) tulajdonságain alapszik. Ha $I_n(\lambda_j^{(n)})$-vel

jelöljük a periodogram $\lambda_j^{(n)}$ helyen felvett értékét, akkor 0-hoz közeli $\lambda_j^{(n)}$ frekvenciákra

$$(4.1)\quad \log I_n(\lambda_j^{(n)}) = \log f(\lambda_j^{(n)}) + \log \frac{I_n(\lambda_j^{(n)})}{f(\lambda_j^{(n)})} \approx -2\delta \log \lambda_j^{(n)} + \log c + \log \frac{I_n(\lambda_j^{(n)})}{f(\lambda_j^{(n)})}$$

(c egy pozitív konstans). [3]-ban és [4]-ben sikerült bebizonyítani, hogy az y_t ARMA folyamatot végtelen mozgó átlagként előállító innovációs folyamat eloszlására vonatkozó, itt nem részletezendő, egyébként egyszerű és elég gyenge feltétel mellett van olyan $\lambda_j^{(n)}$ alappont rendszer sorozat, hogy az ezen $\lambda_j^{(n)}$-kre felírt (4.1) lineáris modellből δ-ra kapott legkisebb négyzetes $\hat{\delta}_n$ becslés sorozat konzisztens és aszimptotikusan normális eloszlású. A $\hat{\delta}_n$ aszimptotikus szórásnégyzet sorozata is megadható, az alappont rendszer függvényeként.

A következő táblázat az ezen módszerrel kapott δ-becsléseket tartalmazza.

	ATOMKI	DATE	DOTE	DRK	MFK
$\hat{\delta}$	0.31	0.25	0.28	0.39	0.48

δ periodogramot használó legkisebb négyzetes becslései

Továbbá, mivel $\hat{\delta_n}$ aszimptotikus eloszlása normális, vizsgálhatjuk a

$$H_0 : \delta = 0, \text{ azaz az idősor rövid memóriájú}$$
$$H_1 : \delta > 0, \text{ azaz az idősor hosszú memóriájú}$$

hipotézisrendszert. A hipotézisvizsgálat eredménye: a megfigyelt elsőfajú hiba mind az öt esetben 10^{-5} alatt van, azaz gyakorlatilag nulla, tehát minden szokásos szintnél, azaz a lehető legbizonyosabban elutasíthatjuk a H_0 hipotézist.

A 8. ábrán a DATE idősor esetében a periodogramnak és a becsült δ-hoz tartozó spektrumnak (a periodogram által részben takart, hiperbola alakú függvény) a módszer által használt $\lambda_j^{(n)}$ alappontokhoz tartozó része látható, mindkettő azonos konstans szorzótól eltekintve.

3. Összefoglalás

A Debreceni Universitas FDDI csomópontjaiban elhelyezkedő routerek processzorainak időegységenkénti terheltségét vizsgáltuk. Sikerült megmutatni, hogy a terheltséget a hosszú memória jellemzi. Az alkalmazott három becslési módszerrel kapott memória paraméterek mind az öt idősorra közel vannak egymáshoz, átla-

gaikat pedig a következő táblázat tartalmazza.

	ATOMKI	DATE	DOTE	DRK	MFK
$\hat{\delta}$	0.32	0.19	0.26	0.4	0.38

δ becslések átlagai

Megjegyezzük, hogy hasonló felismerés — video jelsorozat ill. Ethernet hálózat forgalmi adatai hosszú memóriájúsága — található [1]-ben ill. [6]-ban.

IRODALOM

[1] Beran, J., *Statistics for Long-Memory Processes* (Chapman and Hall, 1994).

[2] Geweke, J., Porter-Hudak, S., The estimation and application of long-memory time series models, *Jour. of Time Series Analysis* **4/4** (1983).

[3] Iglói, E., On periodogram based least squares estimation of the long-memory parameter of FARMA processes, *Publ. Math. Debrecen* **44/3–4** (1994).

[4] Iglói, E., Consistency of the periodogram based linear regression estimator of the long-memory parameter of FARMA processes, *Tech. Rep.* **94/100**, *Dept. of Math., L. Kossuth Univ.* (1994).

[5] Mandelbrot, B. B., Van Ness, J. W., Fractional Brownian motions, fractional noises and applications, *SIAM Rev.* **10** (1968).

[6] Willinger, W., Taqqu, M. S., Leland, W. E., D. W. Wilson, Self-Similarity in High-Speed Packet Traffic: Analysis and Modeling of Ethernet Traffic Measurements, *Statistical Science* **10/1** (1995).

[7] SPECTRUM: System Administrator's Guide (Cabletron Systems, 1995).

[8] SPECTRUM: Report Generator User's Guide (Cabletron Systems, 1995).

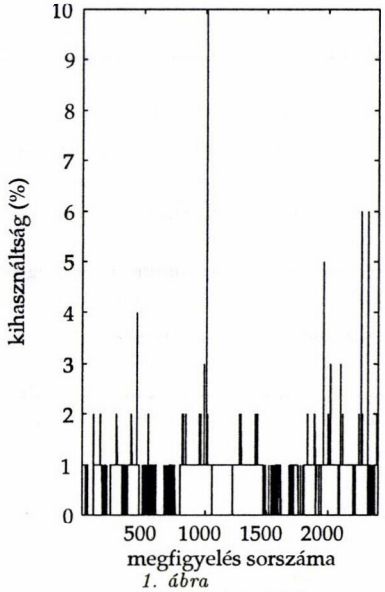
(Beérkezett: 1997. június 21.)

KOSSUTH LAJOS TUDOMÁNYEGYETEM
INFORMATIKAI ÉS SZÁMÍTÓ KÖZPONT
4010 DEBRECEN, EGYETEM TÉR 1.
PF.: 58.
E-mail: zgal@tigris.klte.hu
igloi@tigris.klte.hu
terdik@tigris.klte.hu

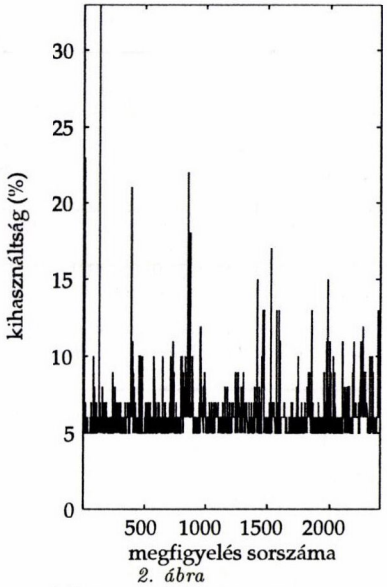
MATHEMATICAL STATISTICAL CHARACTERIZATION OF HIGH-SPEED COMPUTER NETWORK TRAFFIC DATA

ZOLTÁN GÁL, ENDRE IGLÓI AND GYÖRGY TERDIK

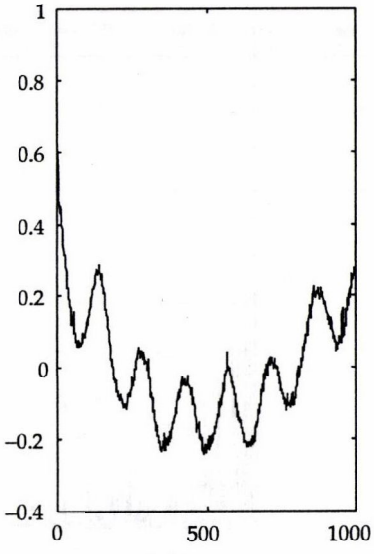
The studying of high-speed networks is an actual problem in both fields of modeling of communication networks and in mathematical statistics. In this paper the methods of time series analysis are applied to the traffic data of the routers by the FDDI ring of Universitas of Debrecen. After deseasonalization the five data series examined are found to be long-memory processes. The long-memory parameters are estimated by three different methods.



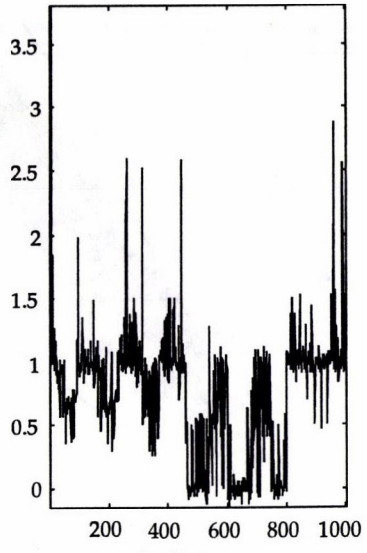
1. ábra
Kihasználtság, ATOMKI



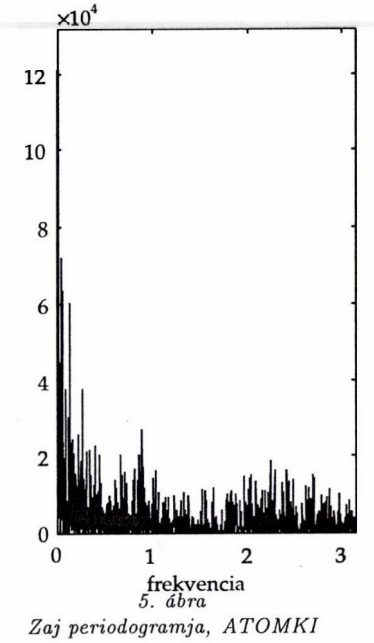
2. ábra
Kihasználtság, DATE



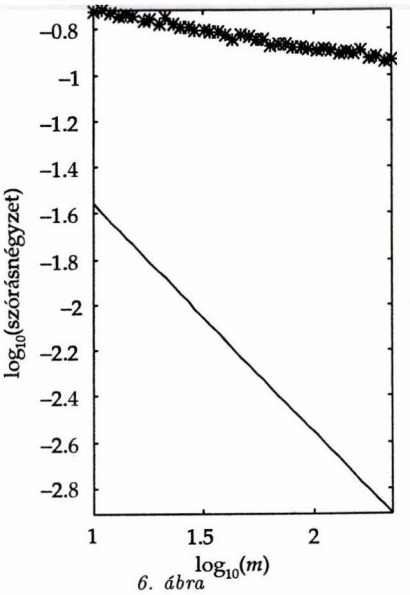
3. ábra
Autokorrelációk, ATOMKI



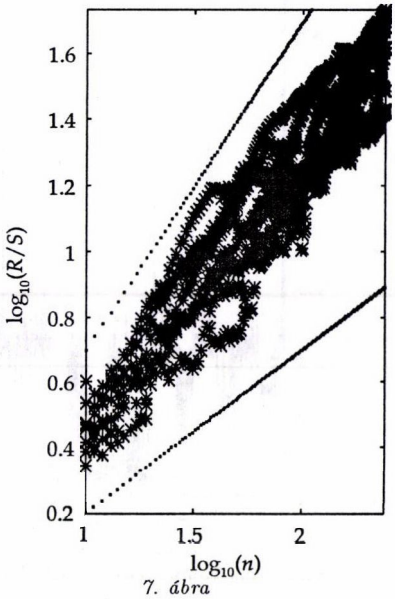
4. ábra
Szezonális komponens, ATOMKI



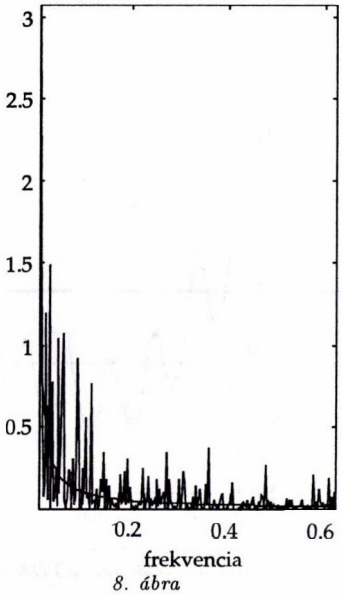
5. ábra
Zaj periodogramja, ATOMKI



6. ábra
Szórásnégyzet-idő graf., DRK



7. ábra
„R/S" grafikon, MFK



8. ábra
Periodogram és spektrum része, DATE

Alkalmazott Matematikai Lapok **19** (1999) 39–68

EGY ASZINKRON SZTOCHASZTIKUS APPROXIMÁCIÓS TÉTEL ÉS NÉHÁNY ALKALMAZÁSA

SZEPESVÁRI CSABA*

Szeged

A dolgozatban Markov döntési problémák adaptív optimális kontrolljával foglalkozunk. Bebizonyítunk egy aszinkron sztohasztikus approximációs tételt. A tétel legfőbb alkalmazása a Markov döntesi problémákban az optimális költségfüggvények becslése. Ezért a dolgozatban a Markov döntési problémák elméletét is áttekintjük, majd bemutatjuk a tétel három alkalmazását. Végül megkonstruálunk néhány aszimptotikusan optimális adaptív politikát.

1. Bevezetés

Tegyük fel, hogy adott egy véges $\mathcal{X}$ állapottér és egy szintén véges $\mathcal{A}$ akciótér és megfigyelhetjük a $\left\{\left(\xi_t^{(\pi)}, \alpha_t^{(\pi)}\right)\right\}$ kontrollált Markov-folyamatot, ahol $\xi_t^{(\pi)} \in X$ a π politikával vezérelt rendszer t-edik pillanatbeli állapota és $\alpha_t^{(\pi)} \in A$ a π politika által előírt akció szintén a t-edik pillanatban, ahol az állapotváltozások a $p(x,a,y) \in [0,1]$ állapotátmenet függvény által adottak ($x,y \in \mathcal{X}, a \in \mathcal{A}$), mégpedig:

$$P\left(\xi_{t+1}^{(\pi)} | \xi_t^{(\pi)}, \alpha_t^{(\pi)}, \ldots, \xi_0^{(\pi)}, \alpha_0^{(\pi)}\right) = p\left(\xi_t^{(\pi)}, \alpha_t^{(\pi)}, \xi_{t+1}^{(\pi)}\right).$$

Tegyük fel, hogy minden lépésben jelentkezik egy $c_t^{(\pi)} \in \mathbf{R}$ költség melyet megfigyelhetünk és amelyre áll, hogy

$$E\left[c_t^{(\pi)} | \xi_{t+1}^{(\pi)}, \alpha_t^{(\pi)}, \xi_t^{(\pi)}, \ldots, \xi_0^{(\pi)}, \alpha_0^{(\pi)}\right] = c\left(\xi_t^{(\pi)}, \alpha_t^{(\pi)}, \xi_{t+1}^{(\pi)}\right),$$

ahol $c(x,a,y) \in \mathbf{R}$ rögzített ($x,y \in \mathcal{X}, a \in \mathcal{A}$), de nem ismerjük sem a p átmenetvalószínűségeket, azaz az egyes akciók hatását, sem a $c_t^{(\pi)}$ közvetlen költségek a pillanatnyi állapotátmenettől való függését, azaz c-t. Ebben a cikkben azzal a kérdéssel foglalkozunk, hogy hogyan adható meg olyan eljárás, amely tetszőleges *véges*

* A cikk elkészítésének ideje alatt a szerző a JATE-MTA Mesterséges Intelligencia Kutatócsoportjának munkatársa volt.

$\mathcal{X}, \mathcal{A}$ terekre, átmenetvalószínűségekre és költségekre egy aszimptotikusan optimális π kontrollt eredményez abban az értelemben, hogy a π által vezérelt rendszerre igaz, hogy

$$\lim_{n\to\infty} P\left(\alpha_n^{(\pi)} \in A^*\left(\xi_n^{(\pi)}\right)\right) = 1. \tag{1}$$

Itt $\xi_0 = \xi_0^{(\pi)}$ egy tetszőleges véletlen kezdőállapot, és $A^*(x)$ az $x \in \mathcal{X}$ állapotbeli optimális akciók halmaza abban az értelemben, hogy ha $\alpha_t^{(\pi)} \in A^*\left(\xi_t^{(\pi)}\right)$ minden t-re, akkor a bejövő költségek $c_t^{(\pi)}$ sorozatának lecsengetett összegének várható értéke tetszőleges kezdőállapotra minimális lesz. Azaz egy ilyen akciókat előíró π^* politikára

$$E\left[\sum_{t=0}^{\infty} \gamma^t c_t^{(\pi^*)} | \xi_0 = x\right] = \inf_{\pi} E\left[\sum_{t=0}^{\infty} \gamma^t c_t^{(\pi)} | \xi_0 = x\right].$$

A mondott $A^*(x)$ halmazrendszer létezése jól ismert (a teljesség kedvéért az ide vonatkozó eredményeket röviden ismertetjük a 3. fejezetben). Nyilván az optimális akciók halmaza, $A^*(x)$, függ az átmenetvalószínűségektől és a közvetlen költségektől. $A^*(x)$-et a következő nem-lineáris, Bellman típusú fixpontegyenlet megoldásával lehet megadni:

$$Q^*(x,a) = \sum_{y\in\mathcal{X}} p(x,a,y)\Big\{c(x,a,y) + \gamma \min_{b\in\mathcal{A}} Q^*(y,b)\Big\}.$$

A mindkét oldalon szereplő $Q^* : \mathcal{X} \times \mathcal{A} \to \mathbf{R}$ függvény az ún. optimális akció költség függvény és az x állapotbeli $A^*(x)$ optimális akciók épp a $Q^*(x,a)$-t minimalizáló akciók.

Ezek alapján logikus, ha az (1), aszimptotikus optimalitást kielégítő politikák konstruálását p és c és ezáltal Q^*, vagy esetleg közvetlenül Q^* becslésére alapozzuk. A továbbiakban a következő feltevésekkel élünk:

1.1. Feltevés. Adott egy $\left(\xi_t^{(\pi)}, \alpha_t^{(\pi)}\right)$ kontrollált folyamat, mely egy véges állapot- és akció-terű $(\mathcal{X}, \mathcal{A}, p, c)$ Markov döntési folyamathoz és valamely π politikához tartozik. Legyen $\mathcal{F}_t$ a $\left(\xi_t^{(\pi)}, \alpha_t^{(\pi)}, c_t^{(\pi)}\right)$ folyamat teljes múltja által generált σ-algebra. Adott továbbá véletlen költségek egy $c_t^{(\pi)}$ sorozata úgy, hogy $E\left[c_t^{(\pi)} | \mathcal{F}_t\right] = c\left(\xi_{t-1}^{(\pi)}, \alpha_{t-1}^{(\pi)}, \xi_t^{(\pi)}\right)$ és $\mathrm{Var}\left[c_t^{(\pi)} | \mathcal{F}_t\right] < C$ valamely $\infty > C > 0$ számra. Továbbá $c_t^{(\pi)}$ és $\xi_{t+1}^{(\pi)}$ függetlenek, ha adott a múlt.

Az aszimptotikusan optimális politikák konstruálásakor fellépő problémákat a következő eljárással illusztráljuk: Tételezzük fel, hogy úgy döntöttünk, hogy p-t és c-t becsüljük, ebből kiszámoljuk Q^* egy becslését, majd Q^* becslése segítségével megadjuk a választandó akciókat. Speciálisan, tegyük fel most, hogy p-t és c-t egyszerű átlagolással becsüljük: Jelöljük $n_t(x,a)$-val azt, hogy $\left(\xi_i^{(\pi)} = x,\right.$

$\alpha_i^{(\pi)} = a\big)$ hányszor fordult elő az első t-lépésben: $n_t(x,a) = \sum_{i=0}^{t} \chi\big(\xi_i^{(\pi)} = x, \alpha_i^{(\pi)} = a\big)$. Ekkor legyen

$$p_t(x,a,y) = \begin{cases} \frac{1}{n_t(x,a)} \sum_{i=0}^{t} \chi\big(\xi_i^{(\pi)} = x, \alpha_i^{(\pi)} = a, \xi_{i+1}^{(\pi)} = y\big), & \text{ha } n_t(x,a) > 0; \\ 0, & \text{különben,} \end{cases}$$

a p becslése a t-edik lépésben és hasonlóan legyen

$$c_t(x,a,y) = \begin{cases} \frac{1}{n_t(x,a)} \sum_{i=0}^{t} c_i \chi\big(\xi_i^{(\pi)} = x, \alpha_i^{(\pi)} = a, \xi_{i+1}^{(\pi)} = y\big), & \text{ha } n_t(x,a) > 0; \\ 0, & \text{különben,} \end{cases}$$

c becslése a t-edik lépésben. Világos, hogy $p_t \to p$ és $c_t \to c$ m.m., ha minden pozitív valószínűségű (x,a,y) átmenetet (azaz az olyan átmeneteket, amelyekre $p(x,a,y) > 0$) végtelen sokszor jár be a kontrollált folyamat. Ez ekvivalens azzal, hogy minden állapotot végtelen sokszor jár be $\xi_t^{(\pi)}$ és, hogy minden állapotban minden akciót π végtelen sokszor próbál ki. Egy ilyen π-t *erősen elegendően felfedezőnek* nevezzünk. Ilyenkor könnyű látni, hogy az $M_t = (\mathcal{X}, \mathcal{A}, p_t, c_t)$ MDP-khoz tartozó Q_t^* optimális akció költség függvények sorozata is tart Q^*-hoz (lásd a 4.1.2. fejezetet is) és elég nagy t-re a megfelelő $A_t^*(x)$ is optimális akciókat ad majd meg. Tehát ha π nagy t-kre az „optimálisnak tűnő", $A_t^*\big(\xi_t^{(\pi)}\big)$-beli akciókat írja elő, akkor aszimptotikusan optimális lehet (1) értelmében. Azonban, ha π csak ilyen akciókat ír elő, akkor már nem feltétlenül lesz erősen elegendően felfedező — tehát π-nek időnként szuboptimálisnak tűnő akciókat is elő kell írnia: Ezek az akciók nem feltétlenül rontják π teljesítményét amennyiben egy pontosabb modell becslését teszik lehetővé. Az optimális akciók választásának kívánalma mindenesetre ellentmondani látszik a döntési probléma megismerésére való törekvésnek. Meg fogjuk mutatni, hogy ez az ellentmondás látszólagos és feloldható, ha a politika által előírt szuboptimális akciók aránya megfelelő ütemben csökken (pl. az x állapot t-edik látogatásakor a szuboptimálisnak tűnő akciók választásának valószínűsége $1/t$-vel arányos).[1]

A fenti módszerben pontosan elkülöníthető két rész: a modell (p, c vagy épp Q^*) becslését célzó, valamint a becslést lehetővé tevő és aszimptotikusan optimális irányítást megadó rész. A következő, 2. fejezetben bemutatunk egy általános módszert, mellyel sokféle becslő algoritmus konvergenciáját tudjuk bizonyítani. Ezután röviden ismertetjük a Markov folyamatok optimális vezérlésének elméletét a 3. fejezetben, majd a 4. fejezetben rátérünk az adaptív politikák konstruálására. Ennek

[1] Az átlagos költség kritérium irodalmában ez a „randomizálás" módszer néven ismert. Egy másik fontos módszer a becslések torzításának módszere, amikoris a becsült mennyiségeket úgy torzítják, hogy azok az akciók, amelyeket kevesebbszer használt a politika jobbnak tűnjenek, mint amit a torzítatlan becslésük ad. Ezzel a módszerrel nemrégiben az átlagos költség kritérium esetére olyan aszimptotikusan optimális politikákat sikerült megadni, amelyek konvergencia sebessége is optimális [6]. A lecsengetett várható összköltség kritériumra Robbins javasolta ezt a torzításos módszert.

keretén belül először bemutatjuk, hogy a 2. fejezetben ismertetett eredmény miképp alkalmazható különféle becslő algoritmusok konvergenciájának bizonyítására (4.1. fejezet), majd a 5. fejezetben megmutatjuk, hogy ha a becslő rész „ideális" feltételek mellett konvergál, akkor megadható olyan politika, amely (i) a becsült mennyiségektől függ, (ii) egyhez tartó valószínűséggel az optimálisnak tűnő akciókat választja, de (iii) még lehetővé teszi a becslő rész konvergenciáját is.

2. Aszinkron sztochasztikus approximáció

Legyen $\mathcal{B}$ egy normált vektortér, $T : \mathcal{B} \to \mathcal{B}$ egy tetszőleges, fixponttal rendelkező operátor és legyen $\mathcal{T} = (T_0, T_1, \ldots)$ véletlentől függő, $\mathcal{B} \times \mathcal{B} \to \mathcal{B}$ leképezések egy sorozata. A [10, 15] dolgozatokban a szerzők a $\mathcal{B} = B(\mathcal{X})$, $\mathcal{X}$ feletti korlátos függvények terében (itt a norma természetesen a szuprémum norma) azt vizsgálják, hogy a $v_{t+1} = T_t(v_t, v_t)$ függvénysorozat milyen, a T_t-kre kirótt feltételek mellett konvergál a T egy fixpontjához majdnem mindenütt (m.m.) a $\mathcal{B}$ normájában, feltéve, hogy a $\mathcal{T} = (T_0, T_1, \ldots)$ operátor sorozat a T-t az alábbi értelemben approximálja:

2.1. Definíció. Legyen $F \subseteq \mathcal{B}$ és $\mathcal{F}_0 : F \to 2^{\mathcal{B}}$. Azt mondjuk, hogy a $\mathcal{T}$ operátor sorozat approximálja T-t az F halmazon és az $\mathcal{F}_0$ által meghatározott kezdeti értékek mellett, ha minden $v \in F$-re és $v_0 \in \mathcal{F}_0(v)$-re az $v_{t+1} = T_t(v_t, v)$ sorozat m.m. konvergál Tv-hez a $\mathcal{B}$ normájában. Ha $F = \{v\}$ egyelemű, akkor azt mondjuk, hogy $\mathcal{T}$ approximálja T-t v-nél az $F_0 = \mathcal{F}_0(v)$ kezdeti értékek mellett.

A következő tétel, melynek részletes bizonyítása a [15] dolgozatban található meg (a bizonyítás vázlatát az Appendixben közöljük) bizonyos kvázi-kontraktív $B(\mathcal{X})$ feletti operátorok approximálhatóságáról szól. A tétel ismertetéséhez szükségünk van egy további fogalomra:

2.2. Definíció. Az $F \subseteq \mathcal{B}$ halmazt invariánsnak nevezzük a $T : \mathcal{B} \times \mathcal{B} \to \mathcal{B}$ operátorra nézve, ha minden $u, v \in F$-re, $T(u, v) \in F$. Továbbá azt mondjuk, hogy F invariáns a $\mathcal{T} = (T_0, T_1, \ldots)$ operátor sorozatra, ha invariáns minden $T_t, t \geq 0$ operátorra.

2.1. TÉTEL. *Legyen $\mathcal{X}$ egy tetszőleges halmaz, $\mathcal{B} = \big(B(\mathcal{X}), \|\cdot\|\big)$, ahol $\|v\| = \sup_{x \in \mathcal{X}} |v(x)|$ és legyen $T : B(\mathcal{X}) \to B(\mathcal{X})$ egy fixponttal rendelkező operátor. Jelöljük v^*-val T egy fixpontját és tegyük fel, hogy a $\mathcal{T} = (T_0, T_1, \ldots)$ véletlen operátorok sorozata approximálja T-t a v^*-nál, az $F_0 \subseteq B(\mathcal{X})$ kezdeti értékek mellett. Tegyük fel továbbá, hogy $v^* \in F_0$, F_0 invariáns $\mathcal{T}$-re és hogy léteznek olyan $g_t : \mathcal{X} \to [0, 1]$ mérhető függvények és egy olyan $0 < \gamma < 1$ konstans, hogy az alábbi feltételek elegendően nagy t-re m.m. teljesülnek:*

1. $\big|T_t(u_1, v^*)(x) - T_t(u_2, v^*)(x)\big| \leq g_t(x)\big|u_1(x) - u_2(x)\big|$, *ahol* $t \in \mathbf{N}$, $x \in \mathcal{X}$ *és* $u_1, u_2 \in F_0$.
2. $\big|T_t(u, v)(x) - T_t(u, v^*)(x)\big| \leq \gamma\big(1 - g_t(x)\big)\big(\|v - v^*\| + \lambda_t\big)$, *ahol* $t \in \mathbf{N}$, $x \in \mathcal{X}$ *és* $u, v \in F_0$ *és* $\lambda_t \to 0$ *m.m.*

3. $\lim_{n\to\infty} \left\| \prod_{t=k}^{n} g_t(\cdot) \right\| = 0$, $k \geq 0$.

Ekkor tetszőleges $v_0 \in F_0$*-ra, a* $v_{t+1} = T_t(v_t, v_t)$ *függvénysorozat* v^**-hoz tart m.m. a* $B(\mathcal{X})$ *normájában.*

Megmutatható, hogy ha a tétel feltételei teljesülnek akkor T valóban kvázikontrakció v^*-nál, mégpedig a tételben szereplő γ kontrakciós faktorral (azaz $\|Tv - Tv^*\| \leq \gamma\|v - v^*\|$, $v \in F_0$). A tételt azért interpretálhatjuk aszinkron szukcesszív approxációs eredményként, mert a tétel feltételei lehetővé teszik, hogy a $v_{t+1} = T_t(v_t, v_t)$ iterációban $v_{t+1}(x) = v_t(x)$ legyen bizonyos (t-től és a véletlentől is függő) x-ekre, azaz a v_t komponensei különböző időpontokban változhatnak. Mindazonáltal, mivel a tétel szerint $\|v_t - v^*\| \to 0$ m.m., így (a tétel feltételei mellett) v_t egyes komponenseinek változási sebességei mégsem térhetnek el túlságosan egymástól.

A rövidség kedvéért bevezetjük azt a konvenciót, hogy az aritmetikai műveleteket, egyenlőtlenségeket, stb. kiterjesztjük az azonos értelmezési tartomány feletti függvényekre is, mégpedig a megfelelő műveletek, egyenlőtlenségek, stb. komponensenként való értelmezésével. Így pl. a tétel 1. feltétele $|T_t(u_1, v^*) - T_t(u_2, v^*)| \leq g_t|u_1 - u_2|$ rövidített alakban írható, a 2. feltétele pedig a $|T_t(u, v) - T_t(u, v^*)| \leq \gamma(1 - g_t)(\|v - v^*\| + \lambda_t)$ alakot ölti. A továbbiakban a $\|\cdot\|$ mindig a szuprémum normát jelöli majd.

3. Markov döntési folyamatok

3.1. Definíció. Egy Markov döntési probléma (MDP) egy $(\mathcal{X}, \mathcal{A}, p, c)$ négyes, ahol

1. $p : \mathcal{X} \times \mathcal{A} \times \mathcal{X} \to \mathbf{R}$ és minden $a \in \mathcal{A}$-ra $p(\cdot, a, \cdot)$ egy átmenetvalószínűség mátrix.
2. $c : \mathcal{X} \times \mathcal{A} \times \mathcal{X} \to \mathbf{R}$.

$\mathcal{X}$-et a Markov döntési probláma állapotterének, $\mathcal{A}$-t az akcióhalmaznak, p-t az átmenetvalószínűség függvénynek, c-t pedig a költség függvénynek nevezzük. A továbbiakban feltesszük, hogy $\mathcal{X}$ és $\mathcal{A}$ végesek.[2]

A Markov döntési problémát magát a következőképp interpretálhatjuk: tekintsünk egy olyan véletlen folyamatot melyet diszkrét időközönként ($t = 0, 1, 2 \ldots$) észlelünk és melynek lehetséges értékei $\mathcal{X}$ elemei. Miután észleltük a folyamat valamely x állapotát egy a akciót kell választunk $\mathcal{A}$-ból. Ezután két dolog történik:

1. a rendszer a $p(x, a, \cdot)$ átmenetvalószínűségek által előírt módon átmegy a következő állapotába, legyen ez y.

[2] A nem tanulásra vonatkozó eredmények kiterjeszthetők nem véges állapotterekre is (ekkor $\mathcal{X}$ Borel kell legyen). Ha az akció halmaz végtelen, akkor még további folytonossági feltevésekkel feltételekkel is kell élni ahhoz, hogy az eredmények átvihetők legyenek. E kérdések tárgyalására nézve lásd pl. [3].

2. elkönyveljük a $c(x,a,y)$ költséget.

Vegyük észre, hogy a rendszer következő állapota csak a pillanatnyi állapottól és a választott akciótól függ. Hasonlóképpen, a t-edik lépésbeli költség független a rendszer korábbi állapotaitól és a korábban választott akcióktól. Az akciók választásának módját valamely π politika szabja meg. Egy politika minden lépésben a (teljes) múlt alapján előír egy akciót, vagy általánosabban egy valószínűségeloszlást az akciók halmazán. A politikák között fontos szerepet játszanak a determinisztikus stacionér politikák. Egy ilyen politika minden lépésben csak a pillanatnyi állapottól függő akciót ír elő és ezért azonosítható egy $\mathcal{X} \to \mathcal{A}$ leképezéssel. Ha egy μ stacionér politikát alkalmazunk akkor az állapotok ξ_t sorozata egy a $\{p_{xy} = p(x,\mu(x),y)\}$ átmenetvalószínűségekkel adott Markov láncot alkot — emiatt hívják a szóban forgó döntési problémákat Markov döntési problémáknak. Eddig a pontig nem határoztuk meg a politikák értékelésének módját. A legkönyebben kezelhető értékelő függvények a politika alkalmazása során keletkező lecsengetett összköltséget veszik alapul, azaz $V_\pi(x) = \sum_{t=0}^{\infty} \gamma^t c_t^{(\pi,x)}$-et, ahol $c_t^{(\pi,x)}$ a π politika alkalmazásakor a t-edik lépésben keletkező költséget jelöli, x a folyamat kezdőállapot és $0 < \gamma < 1$ az ún. lecsengetési faktor. A költségek lecsengetésének egyik indoka az lehet, ha a távoli jövőbeni költségek kevésbé fontosak, mint a jelenbeliek. Közgazdaságtani okoskodással úgy is indokolható a lecsengetés, hogy ha az éves kamatláb mértéke r, akkor ma $\big(1/(1+r)\big)^t c$ forintot kell a bankba tenni a t év múlva jelentkező c költség fedezésére — feltéve, hogy a kamatlábak időben nem változnak és függetlennek a bankba tett pénz mennyiségétől. Azonban $V_\pi(x)$ maga is véletlen mennyiség és így általában nem alkalmas különböző politikák összehasonlítására. Ha $V_\pi(x)$ helyett a várható értékét tekintjük, akkor már összehasonlításra alkalmas értékelést kapunk — a megfelelő költséget a várható lecsengetett összköltségnek nevezzük. Hasonlóképp $V_\pi(x)$ lényeges szuprémuma is megfelelő, a megfelelő költséget a politika pesszimista összköltségének nevezzük. Persze, más értékelések is elképzelhetőek és használatosak. A következő részben a várható lecsengetett összköltség kritériumhoz tartozó elméletet mutatjuk be, de a későbbiekben még visszatérünk egy alkalmazás kapcsán a pesszimista összköltség kritériumhoz is.

A várható lecsengetett összköltség kritérium

Ebben a részben a tárgyalás Ross könyvét követi [18]. Egy $\pi = (\pi_0, \pi_1, \ldots)$ végtelen sorozatot politikának nevezünk, ha minden $t \geq 0$-ra $\pi_t : (\mathcal{A} \times \mathcal{X})^{t+1} \to [0,1]$ és $\pi_t(\cdot; x_t, a_{t-1}, \ldots, a_0, x_0)$ egy valószínűségeloszlás $\mathcal{A}$-n minden

$$(x_t, a_{t-1}, \ldots, a_0, x_0) \in \mathcal{X} \times (\mathcal{A} \times \mathcal{X})^t$$

múltra (itt és a továbbiakban, kihasználva a Descartes-szorzat asszociativitását, megengedjük a zárójelek szabad átcsoportosítását a Descartes szorzatokon belül, pl. $\mathcal{X} \times (\mathcal{A} \times \mathcal{X})^t$-t azonosítjuk $\mathcal{X} \times \mathcal{A} \times \mathcal{X} \ldots \mathcal{A} \times \mathcal{X}$-vel).

3.2. Definíció. A $\{(\xi_n, \alpha_n)\} \subset (\mathcal{X} \times \mathcal{A})^{\mathbf{N}}$ stochasztikus folyamatot a π politikához és a p_0 kezdeti eloszláshoz tartozó *kontroll folyamatnak* nevezzük, ha minden

$n \geq 0$-ra és $(x_n, a_n, \ldots, a_0, x_0)$ múltra teljesülnek az alábbiak:

$$P(\xi_0 = x_0) = p_0(x_0)$$

$$P(\xi_n = x_n \,|\, \alpha_{n-1} = a_{n-1}, \xi_{n-1} = x_{n-1}, \ldots, \alpha_0 = a_0, \xi_0 = x_0) = p(x_{n-1}, a_{n-1}, x_n)$$

$$P(\alpha_n = a_n \,|\, \xi_n = x_n, \alpha_{n-1} = a_{n-1}, \ldots, \alpha_0 = a_0, \xi_0 = x_0) = \pi_n(a_n; x_n, \ldots, a_0, x_0).$$

Ha π feltüntetése lényeges, akkor a π-hez és p_0-hoz tartozó kontroll folyamatot $\left\{\left(\xi_n^{(p_0,\pi)}, \alpha_n^{(p_0,\pi)}\right)\right\}$-vel jelöljük. Ha p_0 egyetlen állapotra, pl. x-re koncentrált, akkor a megfelelő folyamatot $\left\{\left(\xi_n^{(x)}, \alpha_n^{(x)}\right)\right\}$-vel, illetve ha π feltüntetése fontos, akkor $\left\{\left(\xi_n^{(x,\pi)}, \alpha_n^{(x,\pi)}\right)\right\}$-vel jelöljük.

Egy π-hez tartozó kontroll folyamat például a következőképp konstruálható meg: Tetszőleges p_0 $\mathcal{X}$-feletti valószínűség eloszlás és π politika meghatároz egy $(\mathcal{X} \times \mathcal{A})^{\mathbf{N}}$ feletti $P_{p_0,\pi}$ valószínűség eloszlást a következőképpen: Legyen

$$\begin{aligned} P_{p_0,\pi}(x_0, a_0, \ldots, x_n) = {} & p_0(x_0)\pi_0(a_0; x_0) \\ & p(x_0, a_0, x_1)\pi_1(a_1; x_1, a_0, x_0) \\ & \vdots \\ & p(x_{n-1}, a_{n-1}, x_n), \quad n \geq 0 \end{aligned}$$

és

$$P_{p_0,\pi}(x_0, a_0, \ldots, x_n, a_n) = P_{p_0,\pi}(x_0, a_0, \ldots, x_n)\pi_n(a_n; x_n, \ldots, a_0, x_0)$$

és terjesszük ki a természetes módon $P_{p_0,\pi}$-t az $(\mathcal{X} \times \mathcal{A})^{\mathbf{N}}$-feletti cilinderhalmazokra, majd a cilinderhalmazok által generált $\mathcal{F}$ σ-algebrára. Ekkor az $\left(\Omega = (\mathcal{X} \times \mathcal{A})^{\mathbf{N}}, \mathcal{F}, P = P_{\xi_0,\pi}\right)$ valószínűségi tér feletti n-edik koordináta függvények, azaz $\xi_n(\omega) = x_n$, $\alpha_n(\omega) = a_n$, ahol $\omega = (x_0, a_0, x_1, a_1, \ldots)$, épp megfelelnek a 3.2. definíció követelményeinek, azaz egy p_0-hoz és π-hez tartozó kontroll folyamatot adnak meg.

3.3. Definíció. A π politika alkalmazásának összköltsége az x kezdőállapotban

$$v_\pi(x) = E\left[\sum_{t=0}^{\infty} \gamma^t c\left(\xi_t^{(x,\pi)}, \alpha_t^{(x,\pi)}, \xi_{t+1}^{(x,\pi)}\right)\right].$$

Könnyű látni, hogy ha ξ_0 olyan, hogy $P(\xi_0 = x) > 0$, akkor

$$v_\pi(x) = E\left[\sum_{t=0}^{\infty} \gamma^t c\left(\xi_t^{(\pi)}, \alpha_t^{(\pi)}, \xi_{t+1}^{(\pi)}\right) \Big| \xi_0 = x\right].$$

Legyen

$$v^*(x) = \inf_{\pi} v_\pi(x), \quad x \in \mathcal{X}.$$

v^* neve: optimális költségek függvénye, röviden, optimális költség függvény. Egy π^* politikát optimálisnak nevezünk, ha

$$v_{\pi^*}(x) = v^*(x), \quad x \in \mathcal{X}.$$

Az alábbi tétel szerint v^* egy nem-lineáris függvényegyenlet (a Bellman egyenlet) megoldása.

3.1. TÉTEL.

$$v^*(x) = \min_{a \in \mathcal{A}} \sum_{y \in \mathcal{X}} p(x,a,y)\{c(x,a,y) + \gamma v^*(y)\}. \tag{2}$$

Bizonyítás. A teljes valószínűség tétele miatt

$$v_\pi(x) = \sum_{a \in \mathcal{A}} \pi_0(a;x) \sum_{y \in \mathcal{X}} p(x,a,y)\{c(x,a,y) + \gamma v_{\pi^{x,a}}(y)\}, \tag{3}$$

ahol $\pi^{x,a}$ azt a politikát jelöli, amelynek végrehatjását π előírja, feltéve, hogy π-t x-ből indulva kezdjük végrehajtani és hogy az elsőként választott akció az a. Formálisan, $\pi^{x,a} = (\pi_0^{x,a}, \pi_1^{x,a}, \ldots)$, ahol

$$\pi_t^{x,a}(a_t; x_t, a_{t-1}, \ldots, a_0, x_0) = \pi_{t+1}(a_t; x_t, a_{t-1}, \ldots, a_0, x_0, a, x).$$

Mivel $v_{\pi^{x,a}} \geq v^*$, így

$$v_\pi(x) \geq \sum_{a \in \mathcal{A}} \pi_0(a;x) \sum_{y \in \mathcal{X}} p(x,a,y)\{c(x,a,y) + \gamma v^*(y)\} \geq$$

$$\geq \min_{a \in \mathcal{A}} \sum_{y \in \mathcal{X}} p(x,a,y)\{c(x,a,y) + \gamma v^*(y)\}.$$

Mivel π és x tetszőlegesek voltak, így

$$v^*(x) \geq \min_{a \in \mathcal{A}} \sum_{y \in \mathcal{X}} p(x,a,y)\{c(x,a,y) + \gamma v^*(y)\}, \quad x \in \mathcal{X}. \tag{4}$$

A másik irányú egyenlőtlenség a következőképp adódik: Legyen $\mu(x) \in \mathcal{A}$ az az x-hez tartozó akció, amelyre

$$\sum_{y \in \mathcal{X}} p(x,\mu(x),y)\{c(x,\mu(x),y) + \gamma v^*(y)\} =$$

$$= \min_{a \in \mathcal{A}} \sum_{y \in \mathcal{X}} p(x,a,y)\{c(x,a,y) + \gamma v^*(y)\}, \quad x \in \mathcal{X},$$

(ilyen akció van, mivel feltevésünk szerint $\mathcal{A}$ véges) és legyen ${}_x\pi = ({}_x\pi_0, {}_x\pi_1, \ldots)$ egy olyan politika, amelyre $v_{{}_x\pi}(x) \le v_{{}_x\pi}(x) + \varepsilon$, $x \in \mathcal{X}$, ahol $\varepsilon > 0$ tetszőlegesen rögzített. Legyen $\pi = (\pi_0, \pi_1, \ldots)$ a következő: $\pi_0(\mu(x_0); x_0) = 1$ és $\pi_0(a, x_0) = 0$ különben és

$$\pi_t(a_t; x_t, a_{t-1}, \ldots, a_1, x_1, a_0, x_0) = {}_{x_1}\pi_{t-1}(a_t; x_t, a_{t-1}, \ldots, a_1, x_1)$$

(π az első lépésben az x állapotból a $\mu(x)$ akciót végrehajtását írja elő, majd a második lépéstől a ${}_y\pi$ politikát, feltéve, hogy a második lépésben a folyamat állapota y). Ekkor

$$v_\pi(x) = \sum_{y\in\mathcal{X}} p(x, \mu(x), y)\left\{c(x, \mu(x), y) + \gamma v_{{}_y\pi}(y)\right\} \le$$

$$\le \sum_{y\in\mathcal{X}} p(x, \mu(x), y)\left\{c(x, \mu(x), y) + \gamma v^*(y)\right\} + \gamma\varepsilon,$$

amiből a $v^* \le v_\pi$ egyenlőtlenség és μ választása miatt következik, hogy

$$v^*(x) \le \sum_{y\in\mathcal{X}} p(x, \mu(x), y)\left\{c(x, \mu(x), y) + \gamma v^*(y)\right\} + \gamma\varepsilon = \tag{5}$$

$$= \min_{a\in\mathcal{A}} \sum_{y\in\mathcal{X}} p(x, a, y)\left\{c(x, a, y) + \gamma v^*(y)\right\} + \gamma\varepsilon.$$

Mivel ε tetszőleges volt, így (4) és (5) együtt adják a tétel állítását. □

Mielőtt továbbmennénk szükségünk van két fogalomra: a nem-nyújtások és kontrakciók fogalmára. Legegyszerűbb ezt a két fogalmat egy általánosabb fogalomból származtatni, a Lipschitzségből. Egy $T : \mathcal{B}_1 \to \mathcal{B}_2$ normált vektorterek közötti leképezést α-Lipschitznek nevezünk, ha $\mathcal{B}_1$ bármely két f, g elemére $\|Tf - Tg\| \le \alpha\|f - g\|$. Ha $\alpha \le 1$, akkor T-t nem-nyújtásnak nevezzük, ha $\alpha < 1$, akkor pedig kontrakciónak. Könnyű látni, hogy T_1 α- és egy T_2 β-Lipschitz operátorok kompozíciója $\alpha\beta$-Lipschitz, az eltolás és az átlag képzés nem-nyújtások, és hogy az 1-nél abszolútértékben kisebb számmal való szorzás kontrakció. A Banach fixponttétel szerint, ha T egy $\mathcal{B} \to \mathcal{B}$ Banach-terek közötti kontrakció, akkor T-nek egyetlen fixpontja van és $\lim_{n\to\infty} T^n f$ ehhez a fixponthoz tart $\mathcal{B}$ normájában tetszőleges $f \in \mathcal{B}$-re.

A (3) egyenlőséget a Markov döntési folyamatok alapegyenletének nevezik. Ebből az egyenlőségből következik, hogy ha $\pi = (\pi_0, \pi_1, \ldots)$ egy *determinisztikus stacionér politika* (röviden: stacionér politika), azaz ha

$$\pi_t(a_t; x_t, \ldots, a_0, x_0) = \pi_0(a_t; x_t)$$

és $\pi_0(\cdot;x)$ minden x-re egyetlen pontra koncentrált, akkor

$$v_\pi = T_\mu v_\pi, \tag{6}$$

ahol $\mu : \mathcal{X} \to \mathcal{A}$ az a leképezés, amelyre $\pi_0(\mu(x),x) = 1$ $(x \in \mathcal{X})$ és ahol T_μ : $B(\mathcal{X}) \to B(\mathcal{X})$ az az operátor, amelyre

$$(T_\mu v)(x) = \sum_{y\in\mathcal{X}} p(x,\mu(x),y)\big\{c(x,\mu(x),y) + \gamma v(y)\big\}.$$

(6) bizonyításához elég annyit megjegyezni, hogy ha π stacionér politika, akkor $v_{\pi^x} = v_\pi$. Könnyű látni azt is, hogy T_μ kontrakció (a szuprémum normára nézve) és így a Banach fixponttétel miatt v_π (6) egyetlen megoldása és $\|T_\mu^n v - v_\pi\| \to 0$, ahol $v \in B(\mathcal{X})$ tetszőleges. A π stacionér politikát a továbbiakban azonosítjuk a hozzá tartozó $\mu : \mathcal{X} \to \mathcal{A}$ leképezéssel. Így adott $\mu : \mathcal{X} \to \mathcal{A}$ leképezésről beszélhetünk mint stacionér politikáról, és értelmezhetjük v_μ-t is, mint a megfelelő π stacionér politika költségfüggvényét. Az eddigieg alapján a következő, igen fontos tétel már könnyen bizonyítható:

3.2. TÉTEL. *Legyen μ egy olyan $\mathcal{X} \to \mathcal{A}$ leképezés, amelyre*

$$\sum_{y\in\mathcal{X}} p(x,\mu(x),y)\big\{c(x,\mu(x),y) + \gamma v^*(y)\big\} =$$

$$= \min_{a\in\mathcal{A}} \sum_{y\in\mathcal{X}} p(x,a,y)\big\{c(x,a,y) + \gamma v^*(y)\big\}, \quad x \in \mathcal{X}.$$

Ekkor $v_\mu = v^*$, azaz μ optimális.

Bizonyítás. A μ leképezés definíciójából adódik, hogy

$$(T_\mu v^*)(x) = \min_{a\in\mathcal{A}} \sum_{y\in\mathcal{X}} p(x,a,y)\big\{c(x,a,y) + \gamma v^*(y)\big\} = v^*(x),$$

ahol a második egyenlőség a 3.1. tételből következik. Így

$$T_\mu v^* = v^*,$$

amiből

$$T_\mu^2 v^* = T_\mu(T_\mu v^*) = T_\mu v^* = v^*$$

és indukcióval

$$T_\mu^n v^* = v^*.$$

Azonban T_μ kontrakció volta miatt $T_\mu^n v^* \to v_\mu$ és így kapjuk, hogy $v_\mu = v^*$. □

A fenti tétel szerint tehát léteznek optimális politikák, sőt léteznek stacionér optimális politikák is — és ezeket v^*-ból (2) alapján meghatározhatjuk. Így, ha v^*-ot meg tudjuk határozni, akkor meg tudunk adni optimális politikákat is.

Legyen mostmár $T : B(\mathcal{X}) \to B(\mathcal{X})$ a következőképp definiálva:

$$(Tv)(x) = \min_{a \in \mathcal{A}} \sum_{y \in \mathcal{X}} p(x,a,y)\{c(x,a,y) + \gamma v^*(y)\}, \quad x \in \mathcal{X}.$$

A 3.1. tétel miatt $Tv^* = v^*$. A Lipschitz operátoroknál elmondottakból könnyen adódik, hogy T kontrakció és így $T^n v$ tetszőleges $v \in B(\mathcal{X})$-re v^*-hoz tart a szuprémum normában, és v^* a (2) egyetlen megoldása. Megjegyezzük, hogy $T^n v$-t iterációval szokás kiszámolni: ha $v_n = T^n v$, akkor $v_{n+1} = Tv_n$. Ennek az iterációnak számtalan nevet adtak: szokás szukcesszív approximációnak, érték iterációnak, vagy dinamikus programozásnak is hívni. Mi szukcesszív approximációnak fogjuk nevezni. Megmutatható, hogy véges állapottér esetén elég nagy n-re v_n már olyan közel lesz v^*-hoz, hogy az a stacionér politika, amely minden $x \in \mathcal{X}$-re a

$$(Qv_n)(x,a) = \sum_{y \in \mathcal{X}} p(x,a,y)\{c(x,a,y) + \gamma v_n(y)\}, \quad x \in \mathcal{X}. \tag{7}$$

-t minimalizáló akciót írja elő már optimális lesz [15]. A $(Qv)(x,a)$ minimalizáló akciókat *v-re és x-re mohó akcióknak* nevezzük, az olyan stacionér politikát pedig, amely minden x állapotra a v-re és x-re mohó akciókat írja elő *v-re mohó politikának* nevezzük. A 3.2. tétel tehát azt mondja ki, hogy a v^*-ra mohó politikák optimálisak. Ennek a fordítottja is igaz: ha μ stacionér optimális politika, akkor mohó a v^*-ra. Az $A^*(x) = \{a^* \in \mathcal{A} \mid (Qv)(x,a^*) = \min_{a \in \mathcal{A}}(Qv)(x,a)\}$ halmazt az x-ben optimális akciók halmazának nevezzük. Világos, hogy egy olyan politika, amely minden lépésben az aktuális állapothoz tartozó optimális akciók halmazából választ akciót, maga is optimális. Vegyük észre, hogy még v^*-nál is nagyobb fontosságú a

$$Q^* = Qv^* \tag{8}$$

függvény, az ún. optimális állapot-akció költség függvény: ennek ismeretében ugyanis még p és c ismerete nélkül is megkonstruálhatók az optimális stacionér politikák.

4. Adaptív kontroll

Ebben a fejezetben különféle becslő algoritmusok konvergenciáját bizonyítjuk a 2.1. tétel alapján. Megállapodunk, hogy mivel itt maga a politika nem lesz fontos, az alkalmazott politika már olyan, hogy minden (x,a) párt m.b. végte-

len sokszor jár be a kontrollált folyamat. Ezért itt a jelölés egyszerűsítése végett $\{(\xi_t^{(\pi)}, \alpha_t^{(\pi)}, c_t^{(\pi)})\}$-ben elhagyjuk a π-vel való felső indexelést.

4.1. Az optimális költségek becslése

4.1.1. Egy direkt tanulási módszer: a Q-tanulás. Ebben a pontban egy Watkinstól származó iterációs módszert fogunk megadni, mely az irodalomban a Q-tanulásként ismert, mivel itt a $Q^*(x,a)$ optimális állapot-akció költségeket becsüljük. Ennek az algoritmusnak az a különlegessége, hogy ezt anélkül tesszük, hogy p-t vagy c-t megbecsülnénk [20].

Az algoritmus alapgondolata a következő: A Bellman egyenlet miatt $v^*(y) = \min_{b\in\mathcal{A}} Q^*(y,b)$ áll minden $y \in \mathcal{X}$-re és ebből a $\mathcal{Q}$ operátornak az egyenlet mindkét oldalára való alkalmazásával:

$$Q^*(x,a) = (\mathcal{Q}v^*)(x,a) = \Big(\mathcal{Q} \min_{b\in\mathcal{A}} Q^*(\cdot,b)\Big)(x,a),$$

amiből $\mathcal{Q}$ kifejtésével ($\mathcal{Q}$ definíciójára nézve lásd (7))

$$Q^*(x,a) = \sum_{y\in\mathcal{X}} p(x,a,y)\Big\{c(x,a,y) + \gamma \min_{b\in\mathcal{A}} Q^*(y,b)\Big\}, \quad x \in \mathcal{X} \tag{9}$$

adódik. A jobb oldalon $E\big[c(\xi_t,\alpha_t,\xi_{t+1}) + \gamma \min_{b\in\mathcal{A}} Q^*(\xi_{t+1},b) | \xi_t = x,\ \alpha_t = a,\ \mathcal{F}_t\big]$ áll, így a

$$Q_{t+1}(x,a) = \begin{cases} \big(1-\eta_t(x,a)\big)Q_t(x,a) + \eta_t(x,a)\big(c_t + \gamma \min_{b\in\mathcal{A}} Q^*(\xi_{t+1},b)\big), \\ \qquad\qquad\qquad\qquad\qquad\qquad \text{ha } (x,a) = (\xi_t,\alpha_t) \\ Q_t(x,a), \qquad\qquad\qquad\qquad\qquad\qquad \text{különben} \end{cases} \tag{10}$$

iteráció, a nagy számok erős törvényének egy változata szerint (lásd a 4.2. lemmát alább) m.m. Q^*-hoz tart, ha

$$\sum_{t=0}^{\infty} \eta_t(x,a)\chi(\xi_t = x, \alpha_t = a) = \infty$$

$$\sum_{t=0}^{\infty} \eta_t^2(x,a)\chi(\xi_t = x, \alpha_t = a) < \infty$$

(pl. $\eta_t(x,a) = 1/\big(1 + n_t(x,a)\big)$). A rövidség kedvéért vezessük be az $\hat{\eta}_t(x,a) = \eta_t(x,a)\chi(\xi_t = x, \alpha_t = a)$ „tanulási rátákat". Ha a véletlen $T_t : B(\mathcal{X}\times\mathcal{A}) \times B(\mathcal{X}\times\mathcal{A}) \to B(\mathcal{X}\times\mathcal{A})$ operátorokat úgy definiáljuk, hogy a (10) iterációt a tömör

$$Q_{t+1} = T_t(Q_t, Q^*)$$

alakban írhassuk, akkor a

$$T_t(Q,Q')(x,a) = \big(1-\hat{\eta}_t(x,a)\big)Q(x,a) + \\ +\hat{\eta}_t(x,a)\Big(c_t + \gamma \min_{b\in\mathcal{A}} Q'(\xi_{t+1},b)\Big) \tag{11}$$

definíció adódik és azt találjuk, hogy a 2.1. definíció értelmében a $\mathcal{T} = (T_0, T_1, \ldots)$ operátor sorozat approximálja a $T : B(\mathcal{X}\times\mathcal{A}) \to B(\mathcal{X}\times\mathcal{A})$,

$$(TQ)(x,a) = \sum_{y\in\mathcal{X}} p(x,a,y)\Big\{c(x,a,y) + \gamma \min_{b\in\mathcal{A}} Q(y,b)\Big\}, \quad (x,a)\in\mathcal{X}\times\mathcal{A}$$

operátort tetszőleges $Q \in B(\mathcal{X}\times\mathcal{A})$-ra. Azaz $Q_{t+1} = T_t(Q_t, Q) \to TQ$ minden $Q \in B(\mathcal{X}\times\mathcal{A})$-ra. Továbbá (9) értelmében Q^* a T fixpontja. A Q-tanulás alapegyenletéhez úgy jutunk, ha (10)-ben, a $Q_{t+1} = TQ_t$ szukcesszív appoximáció mintájára, Q^*-ot kicseréljük Q_t-re: $Q_{t+1} = T_t(Q_t, Q_t)$ vagy részletesen

$$Q_{t+1}(x,a) = \begin{cases} \big(1-\eta_t(x,a)\big)Q_t(x,a) + \eta_t(x,a)\big\{c_t + \gamma \min_{b\in\mathcal{A}} Q_t(\xi_{t+1},b)\big\}, \\ \qquad\qquad\qquad\qquad\qquad\qquad \text{ha } (x,a) = (\xi_t,\alpha_t) \\ Q_t(x,a), \qquad\qquad\qquad\qquad\qquad\qquad \text{különben.} \end{cases} \tag{12}$$

Ha Q_t-t mint $\mathcal{X}\times\mathcal{A}$-feletti vektort tekintjük, akkor mondhatjuk, hogy Q_t-nek minden lépésben csak egy komponensét változtatjuk (értelmesen nem is lehetne több komponenst módosítani): ezért ezt az iterációt jogosan nevezhetjük aszinkronnak. Watkins eredeti konvergencia bizonyítása hosszadalmas és körülményes volt (szimulációs módszert dolgozott ki és így redukálta a konvergencia kérdését mesterséges Markov döntési problémák megoldására). A Q-tanulást a sztochasztikus approximációval először Jaakkola és munkatársai, valamint Tsitsiklis hozta kapcsolatba [8, 19]. A mi megközelítésünk szerint a Q-tanulás konvergenciája egyszerű következménye a 2.1. általános aszinkron approximációs tételnek.

Mielőtt azonban erre rátérnénk először bebizonyítjuk a nagy számok erős törvényének már emlegetett változatát. A bizonyítás a következő, Robbinstól és Siegmundtól származó szuper-martingál konvergencia típusú lemmán alapszik [17].

4.1. LEMMA. *Legyen $\mathcal{F}_t$ σ-algebrák növekvő sorozata és legyenek Z_t, B_t, C_t, D_t véges, nemnegatív valószínűségi változók úgy, hogy Z_t, B_t, C_t, D_t $\mathcal{F}_t$-mérhetőek. Tegyük fel, hogy*

$$E[Z_{t+1} \mid \mathcal{F}_t] \le (1+B_t)Z_t + C_t - D_t. \tag{13}$$

Akkor a $\big\{\sum_{t=0}^{\infty} B_t < \infty, \sum_{t=0}^{\infty} C_t < \infty\big\}$ halmazon $\sum_{t=0}^{\infty} D_t < \infty$ és $Z_t \to Z < \infty$ m.m.

A lemma bizonyítását nem közöljük mivel az megtalalálható [17]-ben vagy Benveniste és munkatársai [2] könyvében is.

4.2. LEMMA. *Legyen $\mathcal{F}_t$ σ-algebrák növekvő sorozata és legyenek α_t, w_t valószínűségi változók úgy, hogy α_t nemnegatív és α_{t+1}, w_t $\mathcal{F}_{t+1}$-mérhetőek, $t \geq 0$ és α_0 $\mathcal{F}_0$-mérhető. Tegyük fel, hogy $E[w_t \mid \mathcal{F}_t, \alpha_t \neq 0] = A$, $E[w_t^2 \mid \mathcal{F}_t, \alpha_t \neq 0] < B < \infty$, $\sum_{t=0}^{\infty} \alpha_t = \infty$ m.m. és $\sum_{t=0}^{\infty} \alpha_t^2 < \infty$ m.m., ahol $B > 0$. Ekkor a*

$$Q_{t+1} = (1 - \alpha_t)Q_t + \alpha_t w_t, \quad t \geq 0 \tag{14}$$

iteráció m.m. A-hoz tart.

Bizonyítás. Az általánosság megszorítása nélkül feltehető, hogy $E[w_t \mid \mathcal{F}_t] = A$ és $E[w_t^2 \mid \mathcal{F}_t] < B < \infty$. Írjuk át (14)-et $(Q_{t+1} - A) = (Q_t - A) + \alpha_t(w_t - Q_t)$ alakba és legyen $Z_t = (Q_t - A)^2$. Ekkor elemi átalakításokkal, kihasználva $\mathcal{F}_t$ definícióját is kapjuk, hogy

$$E[Z_{t+1} \mid \mathcal{F}_t] \leq (1 + \alpha_t^2)Z_t + \alpha_t^2 \max(0, B - A^2) - 2\alpha_t Z_t.$$

A 4.1. lemmát alkalmazva a $B_t = \alpha_t^2$, $C_t = \alpha_t^2 \max(0, B - A^2)$ és $D_t = 2\alpha_t Z_t$ szereposztással kapjuk, hogy van olyan Z valószínűségi változó, hogy $Z_t \to Z$ m.m. és, hogy

$$\sum_{t=0}^{\infty} 2\alpha_t Z_t < \infty \quad \text{m.m.} \tag{15}$$

Tekintsük azon ω-kat amelyekre $Z_t(\omega) \to Z(\omega)$ és $Z(\omega) \neq 0$. Ekkor van olyan t_0, hogy ha $t \geq t_0$, akkor $Z_t(\omega) > Z(\omega)/2 > 0$ és így

$$\sum_{t=0}^{\infty} 2\alpha_t(\omega)Z_t(\omega) \geq \sum_{t=t_0}^{\infty} 2\alpha_t(\omega)Z_t(\omega) > 2\sum_{t=0}^{\infty} \alpha_t(\omega)Z(\omega)/2 =$$

$$= Z(\omega)\sum_{t=0}^{\infty} \alpha_t(\omega) = \infty,$$

ami ellentmont (15)-nek. Így $Z(\omega) = 0$ m.m. □

Ezekután következzék a Q-tanulás konvergenciáját kimondó tétel:

4.3. TÉTEL. *Legyen $(\mathcal{X}, \mathcal{A}, p, c)$ egy véges állapot és akcióterű MDP és tegyük fel, hogy a $\{(\xi_t, \alpha_t, c_t)\}$ folyamat teljesíti az* 1.1. *feltevésben foglaltakat. Tekintsük a*

$$Q_{t+1}(x, a) = \big(1 - \eta_t(x, a)\big)Q_t(x, a) + \eta_t(x, a)\Big(c_t + \gamma \min_{b \in \mathcal{A}} Q_t(\xi_{t+1}, b)\Big)$$

folyamatot, ahol $\eta_t(x,a)$ *nemnegatív valószínűségi változó és* $\eta_t(x,a) = 0$ *ha* $(x,a) \neq (\xi_t, \alpha_t)$,

$$\sum_{t=0}^{\infty} \eta_t(x,a) = \infty \quad m.m., \textit{ valamint} \tag{16}$$

$$\sum_{t=0}^{\infty} \eta_t^2(x,a) < \infty \quad m.m. \tag{17}$$

Ekkor $Q_t \to Q^*$ *m.m.*

Bizonyítás. A 2.1. tételt használva bizonyítunk. $\mathcal{X} \times \mathcal{A}$-t azonosítjuk a 2.1. tételbeli $\mathcal{X}$ térrel. Legyen $T_t : B(\mathcal{X} \times \mathcal{A}) \times B(\mathcal{X} \times \mathcal{A})$:

$$T_t(Q,Q')(x,a) = \big(1 - \eta_t(x,a)\big)Q(x,a) + \eta_t(x,a)\Big(c_t + \gamma \min_{b \in \mathcal{A}} Q'(\xi_{t+1}, b)\Big).$$

Ekkor $Q_{t+1} = T_t(Q_t, Q_t)$. Feltevéseink és a 4.2. lemma alapján némi számolással belátható, hogy a $\mathcal{T} = (T_0, T_1, \ldots)$ sorozat a 2.1. definíció értelmében bármely $Q \in B(\mathcal{X} \times \mathcal{A})$-nál approximálja T-t. (A lemmabeli $\mathcal{F}_t$-t például a

$$(\xi_t, \alpha_t, \alpha_t(x,a), c_{t-1}, \xi_{t-1}, \alpha_{t-1}, \alpha_{t-1}(x,a), c_{t-2}, \ldots, \xi_0, \alpha_0)$$

által generált σ-algebrának választhatjuk, $t \geq 0$.) A T_t operátor a $g_t(x,a) = 1 - \eta_t(x,a)$ választással triviálisan megfelel a 2.1. tétel 1–2. feltételeinek, így a tétel szerint valóban $\|Q_t - Q^*\| \to 0$ m.m. ha 3. feltételt, mely $\left\| \prod_{i=n}^{t} g_i(\cdot,\cdot) \right\| \to 0$ m.m. kívánja is igazoljuk. Mivel $\mathcal{X} \times \mathcal{A}$ véges így elég adott (x,a) párra bizonyítani, hogy $\prod_{i=n}^{t} g_i(x,a) \to 0$. Az általánosság megszorítása nélkül feltehető, hogy n olyan nagy, hogy $\eta_i(x,a) < 1$ m.m. ($\eta_i(x,a) \to 0$ m.m. (17) miatt). A $\log(1+x) \leq x$, $x > 0$ azonosság és (16) felhasználásával kapjuk, hogy

$$\prod_{i=n}^{t} g_i(x,a) \to 0 \quad \text{m.m.}$$

és ezzel a bizonyítás kész is. □

Könnyű látni, hogy a Q-tanulás konvergenciája érvényben marad akkor is, ha az iterációt Monte-Carlo szimulációban használjuk, azaz ha az algoritmust olyan $\langle \xi_t, \alpha_t, c_t, \phi_t \rangle$ négyesekre hajtjuk végre, ahol ϕ_t ξ_{t+1}-t helyettesíti (a tanulás egyenletében is) — de a szereplő mennyiségek tulajdonságai változatlanak. A Monte-Carlo szimuláció akkor lehet hasznos, ha a rendszer ismert, de túl nagy ahhoz, hogy explicite megoldjuk. A tétel bizonyításához teljesen hasonlóan belátható sok a Q-tanulással rokon tanuló algoritmus konvergenciája is [16, 15].

4.1.2. Az indirekt módszer. Az adaptív kontrollt bevezető fejezetben már megemlítettük, hogy aszimptotikusan optimális politika konstruálására egy lehetséges megoldás, ha p-t és c-t becsüljük a kontroll közben és a becsült értékeket használjuk fel a Bellman egyenletet használva a v^* becslésére. Azonban nagy állapot és/vagy akcióterekben a Bellman egyenlet megoldása költséges lehet. A p és c paraméterek átlagolással kapott becslései inkrementálisan is számolhatók az $(s_1 + \ldots + s_n)/n = (1 - 1/n)(s_1 + \ldots + s_{n-1})/(n-1) + (1/n)s_n$ összefüggés felhasználásával. Miért ne tehetnénk valami hasonlót v^* becslésével? Tekintsük a következő iterációt:

$$v_{t+1}(x) = \begin{cases} (T_t v_t)(x), & \text{ha } x \in U_t \\ v_t(x), & \text{különben,} \end{cases} \tag{18}$$

ahol $T_t : B(\mathcal{X}) \to B(\mathcal{X})$ a p_t, c_t paraméterek által meghatározott operátor:

$$(T_t v)(x) = \min_{a \in \mathcal{A}} \sum_{y \in \mathcal{X}} p_t(x,a,y)\big\{c_t(x,a,y) + \gamma v(y)\big\}$$

és $U_t \subseteq \mathcal{X}$ a t-edik lépésben felfrissített állapotok halmaza. Általában megengedjük, hogy U_t függjön a véletlentől. Számítógépes tapasztalatok szerint például meggyorsíthatja a konvergenciát, ha $\xi_t, \xi_{t-1}, \ldots, \xi_{t-k} \in U_t$ adott $k > 0$-ra. Ennek egy lehetséges magyarázata az, hogy így a ξ_t költségbecslésének változásának esetleges hatásai gyorsan visszagyűrűznek a ξ_t-t megelőző állapotokra. A (18) iteráció is futtatható Monte-Carlo szimulációban. Különösen jó hatásfokot lehet elérni ún. start-cél keresési feladatok esetén, mikoris van egy (vagy több) kitüntetett $x^* \in \mathcal{X}$ cél állapot, mely nyelő tulajdonságú és $c(x^*, a, x^*) = 0$, minden $a \in \mathcal{A}$-ra és vannak kezdőállapotok, melyekből megadják a lehetséges kiindulási állapotokat (ekkor definíció szerint a kezdőállapotoktól különböző állapotokból nem indulhat a rendszer). Ekkor bebizonyítható, hogy az olyan Monte-Carlo szimulációban amit újraindítunk a célállapot elérésekor a lényeges állapotok mentén $v_t(x) \to v^*(x)$ m.m. Egy állapotot akkor nevezünk lényegesnek, ha elérhető valamely kezdőállapotból valamely optimális stacionér politika végrehajtásával [1]. Az alábbi tétel magában foglalja ezt az eredményt és a fenti (18) iteráció konvergenciáját is:

4.4. TÉTEL. *Legyen $M = (\mathcal{X}, \mathcal{A}, p, c)$ egy véges állapot- és akcióterű MDP, legyenek $S^{(x,a)}, S_t^{(x,a)} : B(\mathcal{X}) \to \mathbf{R}$ nemnyújtó operátorok ($S : \mathcal{B}_1 \to \mathcal{B}_2$ nemnyújtó, ha bármely $u, v \in \mathcal{B}_1$ párra $\|Su - Sv\| \le \|u - v\|$) és legyen $T : B(\mathcal{X}) \to B(\mathcal{X})$ a következő operátor:*

$$(Tv)(x) = \min_{a \in \mathcal{A}} S^{(x,a)}\big\{c(x,a,\cdot) + \gamma v(\cdot)\big\},$$

ahol $0 < \gamma < 1$. Hasonlóan, legyen $T_t : B(\mathcal{X}) \to B(\mathcal{X})$ a következő operátor:

$$(T_t v)(x) = \min_{a \in \mathcal{A}} S_t^{(x,a)}\big\{c_t(x,a,\cdot) + \gamma v(\cdot)\big\},$$

ahol $c_t \to c$ *m.m. Tekintsük a*

$$v_{t+1}(x) = \begin{cases} (T_t v_t)(x), & \text{ha } x \in U_t \\ v_t(x), & \text{különben} \end{cases} \tag{19}$$

iterációt. Tegyük fel, hogy

$$\lim_{t\to\infty} \max_{(x,a)\in\mathcal{X}\times\mathcal{A}} \left| S_t^{(x,a)} v - Sv \right| = 0 \quad \text{m.m.} \tag{20}$$

áll minden $v \in B(\mathcal{X})$*-re és minden* $x \in \mathcal{X}$*-re és* $x \in U_t$ *végtelen sokszor m.m. Ekkor* T *kontrakció és fixpontját* v^**-val jelölve* $v_t \to v^*$ *m.m.*

Bizonyítás. Ismét a 2.1. tételt alkalmazzuk. Természetesen adódik, hogy legyen

$$T_t(u,v)(x) = \begin{cases} (T_t v)(x), & \text{ha } x \in U_t \\ u(x), & \text{különben} \end{cases}$$

Legyen $x \in \mathcal{X}$ és legyen $u_{t+1} = T_t(u_t, v)$, $v \in B(\mathcal{X})$. Mivel $u_{t+1}(x) = u_t(x)$, ha $x \notin U_t$, és ha $x \in U_t$, akkor $u_{t+1}(x)$ nem függ u_t-től, és mivel $x \in U_t$ végtelen sokszor m.m., így ahhoz, hogy megmutassuk, hogy $T = (T_0, T_1, \ldots)$ approximálja T-t elég ha megmutatjuk, hogy a $D_t = \left|(T_t v)(x) - (Tv)(x)\right| \to 0$ m.m. Mivel

$$\begin{aligned} D_t &= \left| \min_{a\in\mathcal{A}} S_t^{(x,a)}\{c_t(x,a,\cdot) + \gamma v(\cdot)\} - \min_{a\in\mathcal{A}} S^{(x,a)}\{c(x,a,\cdot) + \gamma v(\cdot)\} \right| \le \\ &\le \max_{a\in\mathcal{A}} \left| S_t^{(x,a)}\{c_t(x,a,\cdot) + \gamma v(\cdot)\} - S^{(x,a)}\{c(x,a,\cdot) + \gamma v(\cdot)\} \right| \le \\ &\le \max_{a\in\mathcal{A}} \left| S_t^{(x,a)}\{c_t(x,a,\cdot) + \gamma v(\cdot)\} - S_t^{(x,a)}\{c(x,a,\cdot) + \gamma v(\cdot)\} \right| + \\ &+ \max_{a\in\mathcal{A}} \left| S_t^{(x,a)}\{c(x,a,\cdot) + \gamma v(\cdot)\} - S^{(x,a)}\{c(x,a,\cdot) + \gamma v(\cdot)\} \right| \le \\ &\le \max_{a\in\mathcal{A}} \max_{y\in\mathcal{X}} \left| c_t(x,a,y) - c(x,a,y) \right| + \\ &+ \max_{a\in\mathcal{A}} \left| S_t^{(x,a)}\{c(x,a,\cdot) + \gamma v(\cdot)\} - S^{(x,a)}\{c(x,a,\cdot) + \gamma v(\cdot)\} \right|. \end{aligned} \tag{21}$$

Mivel feltevés szerint $c_t \to c$ m.m., így $\max_{a\in\mathcal{A}} \max_{y\in\mathcal{X}} \left|c_t(x,a,y) - c(x,a,y)\right| \to 0$ m.m. Másrészről (20) miatt (21) második tagja is nullához tart m.m., így $D_t \to 0$ m.m. is áll.

A 2.1. tétel maradék feltételeit a

$$g_t(x) = \begin{cases} 0, & \text{ha } x \in U_t \\ 1, & \text{különben} \end{cases}$$

függvény triviálisan kielégíti, mivel $S_t^{(x,a)}$ nemnyújtó, $0 < \gamma < 1$, és minden x végtelen sokszor van U_t-ben. □

Jegyezzük meg, hogy a szokásos várható lecsengetett összköltség kritériumhoz az

$$S^{(x,a)}v = \sum_{y\in\mathcal{X}} p(x,a,y)v(y)$$

operátorok tartoznak és ha $p_t \to p$ m.m., akkor a megfelelő

$$S_t^{(x,a)}v = \sum_{y\in\mathcal{X}} p_t(x,a,y)v(y)$$

operátorokra $S_t^{(x,a)} \to S^{(x,a)}$ a (20) értelmében. Érdekességként megjegyezzük, hogy ha $p_t = p$ és $c_t = c$, akkor az U_t halmazok megfelelő választásával visszszakaphatjuk pl. a szukcesszív approximáció különböző változatait, mint pl. a Jacobi-iterációt.

A tételt azért mondtuk ki az $S^{(x,a)}$ operátorok segítségével, hogy közvetlenül alkalmazható legyen egyéb kritériumú döntési problémákra is. Vegyük például a pesszimista összköltség kritériumot. Ekkor egy π politika összköltsége, mint már említettük

$$v_\pi(x) = \operatorname{ess\,sup}\left\{\sum_{t=0}^{\infty}\gamma^t c\big(\xi_t^{(x,\pi)}, \alpha_t^{(x,\pi)}, \xi_{t+1}^{(x,\pi)}\big)\right\}.$$

Hasonlóan a várható összköltség kritériumhoz tartozó levezetésekhez megmutatható, hogy ha a $\mathcal{Q} : B(\mathcal{X}) \to B(\mathcal{X}\times\mathcal{A})$ operátor

$$(\mathcal{Q}v)(x,a) = \max_{y\in\mathcal{X}\,:\,p(x,a,y)>0}\big\{c(x,a,y) + \gamma v(y)\big\}$$

egyenlet által definiált, és $T : B(\mathcal{X}) \to B(\mathcal{X})$, $(Tv)(x) = \min_{a\in\mathcal{A}}(\mathcal{Q}v)(x,a)$ által definiált, akkor a $Tv^* = v^*$ Bellman egyenlet igaz (itt $v^*(x)$ a pesszimistán értékelt optimális összköltség) és a v^*-ra mohó μ stacionér politikák optimálisak is. Természetesen most akkor nevezzük μ-t v-re mohónak, ha

$$\max_{y\in\mathcal{X}\,:\,p(x,\mu(x),y)>0}\big\{c\big(x,\mu(x),y\big) + \gamma v(y)\big\} =$$

$$= \min_{a\in\mathcal{A}}\max_{y\in\mathcal{X}\,:\,p(x,a,y)>0}\big\{c(x,a,y) + \gamma v(y)\big\}$$

vagy tömören $\big(\mathcal{Q}v\big)\big(x,\mu(x)\big) = (Tv)(x)$ áll minden $x\in\mathcal{X}$-re.

Világos, hogy a fenti tételben a pesszimista összköltség kritériumhoz az $S^{(x,a)}v = \max_{y\in\mathcal{X}} p(x,a,y)v(y)$ operátor tartozik és az

$$L_t(x,a) = \{\xi_{i+1} \,|\, \xi_i = x,\ \alpha_i = a,\ 0 \le i+1 \le t\}$$

definícióval és az $S_t^{(x,a)}v = \max_{y\in L_t(x,a)} v(y)$ választással szintén $S_t^{(x,a)} \to S^{(x,a)}$.

4.1.3. Q-tanulás a pesszimista értékelésre. Láttuk, hogy a direkt módszer használható a pesszimista értékeléshez tartozó optimális költségfüggvény becslésére. Ebben a fejezetben bebizonyítjuk a megfelelő indirekt módszer konvergenciáját is. Ezt a módszert Heger javasolta [7] és $\hat{Q}$-tanulásnak nevezte el. A 1.1. feltevést most a következővel helyettesítjük:

4.1. Feltevés. Adott egy (ξ_t, α_t) kontrollált folyamat, mely egy véges állapot- és akció-terű $(\mathcal{X}, \mathcal{A}, p, c)$ Markov döntési folyamathoz és valamely rögzített π politikához tartozik. Feltesszük, hogy (ξ_t, α_t) ergodikus abban az értelemben, hogy minden (x, a) párra $(\xi_t, \alpha_t) = (x, a)$ végtelen sokszor teljesül m.m. Adott továbbá véletlen költségek egy c_t sorozata úgy, hogy ha $t_n(x, a, y)$ jelöli azon időpontokat, amikor $(\xi_t, \alpha_t, \xi_{t+1}) = (x, a, y)$, akkor m.m.

$$c_{t_n(x,a,y)} \le c(x, a, y) \tag{22}$$

és minden olyan y-ra, amelyre $p(x, a, y) > 0$ m.m.

$$\limsup_{n\to\infty} c_{t_n(x,a,y)} = c(x, a, y). \tag{23}$$

A következő tételt bizonyítjuk:

4.5. TÉTEL. *Legyen $(\mathcal{X}, \mathcal{A}, p, c)$ egy véges állapot és akcióterű MDP és tegyük fel, hogy a $\{(\xi_t, \alpha_t, c_t)\}$ folyamat teljesíti az* 4.1. *feltevésben foglaltakat. Legyen $T : B(\mathcal{X} \times \mathcal{A}) \to B(\mathcal{X} \times \mathcal{A})$ a következő kontrakciós operátor:*

$$(TQ)(x, a) = \max_{y\in\mathcal{X} : p(x,a,y)>0} \Big\{ c(x, a, y) + \gamma \min_{b\in\mathcal{A}} Q(y, b) \Big\}$$

és jelöljük a fixpontját Q^*-val.

Tegyük fel, hogy $Q_0(x, a) \le Q^*(x, a)$ és tekintsük a

$$Q_{t+1}(x, a) = \begin{cases} \max\{Q_t(x, a),\ c_t + \gamma \min_{b\in\mathcal{A}} Q_t(\xi_{t+1}, b)\}, & \text{ha } (x, a) = (\xi_t, \alpha_t), \\ Q_t(x, a), & \text{különben} \end{cases}$$

rekurziót. Ekkor Q_t konvergál Q^* m.m.

Világos, hogy T kontrakció a γ kontrakciós faktorral.

Bizonyítás. A bizonyítás ismét csak a 2.1. tétel alkalmazása, csakhogy a T_t operátorsorozatot most ügyesebben kell megválasztanunk. Először is adott (x, a) párra legyen a kritikus (rákövetkező) állapotok halmaza a következő:

$$\mathcal{M}(x, a) = \Big\{ y \in \mathcal{X} \mid p(x, a, y) > 0,\ Q^*(x, a) = c(x, a, y) + \gamma \min_{b\in\mathcal{A}} Q^*(y, b) \Big\}. \tag{24}$$

$\mathcal{M}(x,a)$ nem üres, mert $\mathcal{X}$ véges. Mivel a c_t költségekre áll (22) és (23), így az általánosság megszorítása nélkül feltehető t_n definíciójának módosításával, hogy

$$\lim_{n\to\infty} c_{t_n(x,a,y)} = c(x,a,y). \tag{25}$$

Legyen $T(x,a,y) = \{t_k(x,a,y) \mid k \geq 0\}$ és $T(x,a) = \cup_{y\in\mathcal{M}(x,a)} T(x,a,y)$ és legyen

$$T_t(Q',Q)(x,a) = \begin{cases} \max\left(c_t + \gamma \min_{b\in\mathcal{A}} Q(y_t,b), Q'(x,a)\right); & \text{ha } t \in T(x,a), \\ Q'(x,a); & \text{különben.} \end{cases}$$

Tekintsük a $Q'_0 = Q_0$, $Q'_{t+1} = T_t(Q'_t, Q'_t)$ rekurzióval definiált Q'_t sorozatot és legyen

$$\mathcal{F}_0 = \{Q \in B(\mathcal{X}\times\mathcal{A}) \mid Q(x,a) \leq Q^*(x,a) \text{ for all } (x,a) \in \mathcal{X}\times\mathcal{A}\}$$

a lehetséges kezdeti feltételek halmaza. $\mathcal{F}_0$ invariáns T_t-re. Teljes indukcióval Q_t, Q'_t és Q^* definícióinak felhasználásával könnyen látható, hogy $Q'_t \leq Q_t \leq Q^*$ és így elég belátnunk, hogy Q'_t m.m. konvergál Q^*-hoz.

Világos, hogy T_t approximálja T-t Q^*-nál, mert m.b. van végtelen sok olyan t időpont, hogy $t \in T(x,a)$ és az is világos, hogy a

$$g_t(x,a) = \begin{cases} 0; & \text{ha } (x,a) = (x_t,a_t) \text{ és } y_t \in \mathcal{M}(x,a), \\ 1; & \text{különben,} \end{cases}$$

sorozat kielégíti a 2.1. 1. feltételét, ugyanis $T_t(Q,Q^*)(x,a) = Q^*(x,a)$, ha $(x,a) = (x_t,a_t)$ és $y_t \in \mathcal{M}(x,a)$.

Mutassuk meg, hogy a 2. feltételt is kielégíthetjük ezzel a g_t-vel, azaz becsüljük meg $|T_t(Q',Q)(x,a) - T_t(Q',Q^*)(x,a)|$-et felülről, ahol $Q, Q' \in \mathcal{F}_0$, azaz $Q, Q' \leq Q^*$. Elsőként tegyük föl, hogy $t \in T(x,a)$, azaz, hogy $(x,a) = (x_t,a_t)$ és $y_t \in \mathcal{M}(x,a)$. Ekkor

$$\begin{aligned} |T_t(Q',Q)(x,a) - T_t(Q',Q^*)(x,a)| &\leq \left(c(x,a,y_t) + \gamma \min_{b\in\mathcal{A}} Q^*(y_t,b)\right) - \\ &\quad - \max\left(c_t + \gamma \min_{b\in\mathcal{A}} Q(y_t,b), Q'(x,a)\right) \leq \end{aligned} \tag{26}$$

$$\leq \left(c(x,a,y_t) + \gamma \min_{b\in\mathcal{A}} Q^*(y_t,b)\right) - \left(c_t + \gamma \min_{b\in\mathcal{A}} Q(y_t,b)\right) \tag{27}$$

$$\leq \gamma\|Q^* - Q\| + |c(x,a,y_t) - c_t|, \tag{28}$$

aholis kihasználtuk, hogy $T_t(Q',Q^*)(x,a) \geq T_t(Q',Q)(x,a)$ (mivel T_t a második argumentumában monoton) és, hogy

$$T_t(Q',Q^*)(x,a) \leq \max\left(c(x,a,y_t) + \gamma \min_{b\in\mathcal{A}} Q^*(y_t,b), Q'(x,a)\right) \leq$$

$$\leq \max\Big(c(x,a,y_t) + \gamma \min_{b\in\mathcal{A}} Q^*(y_t,b), Q^*(x,a)\Big) = c(x,a,y_t) + \gamma \min_{b\in\mathcal{A}} Q^*(y_t,b)$$

igaz, mivel $Q' \leq Q^*$ és $y_t \in \mathcal{M}(x,a)$.

Legyen $\sigma_t(x,a) = \big|c(x,a,y_t) - c_t\big|$. Ekkor (25) miatt

$$\lim_{t\to\infty, t\in T(x,a)} \sigma_t(x,a) = 0$$

m.m. A másik esetben (amikor $t \notin T(x,a)$), $\big|T_t(Q',Q)(x,a) - T_t(Q',Q^*)(x,a)\big| = 0$. Így

$$\big|T_t(Q',Q)(x,a) - T_t(Q',Q^*)(x,a)\big| \leq \gamma\big(1 - g_t(x,a)\big)\big(\|Q - Q^*\| + \lambda_t\big),$$

ahol $\lambda_t = \sigma_t(x_t,a_t)/\gamma$, ha $t \in T(x,a)$, és $\lambda_t = 0$, egyébként. Azaz a 2. feltételt is kielégíti g_t, mivel λ_t m.m. nullához tart.

A 3. feltétel teljesüléséhez szükséges és elegendő, hogy $t \in T(x,a)$ végtelen sokszor. Ez azonban a ξ_t, α_t-re kirótt feltétel miatt áll és mivel $p(x,a,y) > 0$ minden $y \in \mathcal{M}(x,a)$-re.

Így a 2.1. Tétel szerint Q'_t m.m. tart Q^*-hoz és ezért Q_t is m.m. tart Q^*-hoz. □

5. Aszimptotikusan optimális adaptív politikák

Térjünk most vissza a bevezetőben említett problémára: adjunk meg olyan π politikát, amely aszimptotikusan optimális, azaz teljesíti az (1) egyenletet. Valamennyi korábbi konvergencia eredményünknél kulcsszerepe volt annak a feltevésnek, hogy a tanulás közbeni π politika olyan, hogy a megfelelő $\big(\xi_t^{(\pi)}, \alpha_t^{(\pi)}\big)$ kontrollált Markov folyamat minden (x,a) párt végtelen sokszor jár be. Ekkor hívtuk a π politikát erősen elegendően felfedezőnek. Először ezt a feltevést gyengítjük.

5.0.4. Elégségesen felfedező politikák.

5.1. Definíció. A π politikát *elégségesen felfedezőnek* nevezzük, ha minden (x,a) párra az $\big\{\omega : x = \xi_t^{(\pi)}(\omega)$ v.s.$\big\}$ halmazon m.m. $\big(x = \xi_t^{(\pi,)}, a = \alpha_t^{(\pi)}\big)$ is v.s. (v.s. ≡ végtelen sokszor). A végtelen sokszor látogatott állapotok halmazát X_∞-vel jelöljük: $X_\infty(\omega) = \big\{x \in \mathcal{X} \,|\, x = \xi_t^{(\pi)}\big\}$.

A továbbiakban csak az várható lecsengetett összköltség kritériumot tekintjük és a Q-tanulás algoritmust. Az eredmények kiterjeszthetők a többi algoritmusra és egyéb kritériumokra is. Tegyük fel tehát, hogy (a π-vel való felsőindexelést elhagyva)

(29)

$$Q_{t+1}(x,a) = \begin{cases} (1-\eta_t(x,a))Q_t(x,a) + \eta_t(x,a)\{c_t + \gamma \min_{b\in\mathcal{A}} Q_t(\xi_{t+1},b)\}, \\ \qquad\qquad\qquad\qquad\qquad\qquad \text{ha } (x,a) = (\xi_t,\alpha_t) \\ Q_t(x,a), \qquad\qquad\qquad\qquad\qquad\qquad \text{különben,} \end{cases}$$

ahol (ξ_t, α_t, c_t) teljesíti az 1.1. feltevésben foglaltakat és

$$\eta_t(x,a) = \frac{1}{1+n_t(x,a)},$$

ahol $n_t(x,a) = \#\{t \geq i \geq 0 : (x,a) = (\xi_t,\alpha_t)\}$.

5.1. TÉTEL. *Legyen π egy elégségesen felfedező politika és legyen Q_t (29)-vel, rekurzívan megadva. Ekkor minden (x,a) párra és m.m. ω-ra amelyre $x \in X_\infty(\omega)$, $\lim_{t\to\infty} Q_t(x,a)(\omega) = Q^*(x,a)$, azaz $\lim_{t\to\infty} Q_t|_{X_\infty} \to Q^*|_{X_\infty}$ m.m.*

Bizonyítás. Legyenek $\mathcal{X}_1, \ldots, \mathcal{X}_k \subset \mathcal{X}$ azok a részhalmazai $\mathcal{X}$-nek, amelyekre $P(\mathcal{X}_\infty = \mathcal{X}_i) > 0$, $1 \leq i \leq k$. Rögzítsünk egy i-t és tekintsük azon eseményeket, amelyekre $\{X_\infty = X_i\}$. Ekkor minden $(x,y) \in \mathcal{X}_i \times (\mathcal{X} \setminus \mathcal{X}_i)$ párra és minden $a \in \mathcal{A}$ akcióra $p(x,a,y) = 0$, mivel különben a politika elégségesen felfedező voltából fakadóan $y \in \mathcal{X}_\infty$ is állna m.m. a $\{\mathcal{X}_\infty = \mathcal{X}_i\}$ halmazon, amiből az $y \in \mathcal{X}_i$ ellentmondás következik. Így az eredeti MDP $\mathcal{X}_i$-re való megszorítása értelmezhető a természetes módon. Jelölje Q_i^* a megszorított MDP-nek megfelelő optimális állapot-akció költség függvényt. A Q_i^*-ra vonatkozó Bellman egyenlet szerint

$$Q_i^*(x,a) = \sum_{y\in\mathcal{X}_i} p(x,a,y)\Big(c(x,a,y) + \gamma \min_{b\in A} Q_i^*(y,b)\Big), \quad \forall (x,a) \in \mathcal{X}_i \times \mathcal{A}.$$

Hasonlóan, Q^* kielégíti a

$$Q^*(x,a) = \sum_{y\in\mathcal{X}} p(x,a,y)\Big(c(x,a,y) + \gamma \min_{b\in A} Q^*(y,b)\Big) =$$

$$= \sum_{y\in\mathcal{X}_i} p(x,a,y)\Big(c(x,a,y) + \gamma \min_{b\in A} Q^*(y,b)\Big), \quad \forall (x,a) \in \mathcal{X}_i \times \mathcal{A}$$

egyenletet. Így mindketten kielégítik ugyanazt a fixpont egyenletet. Mivel $\gamma < 1$ a fixpont egyenletnek csak egy megoldása lehet és így

$$Q_i^* = Q^*|_{\mathcal{X}_i}. \tag{30}$$

Legyen $\tau \in \mathbf{N}$ és legyen $\Omega_{i,\tau} = \{\omega : x_t \in \mathcal{X}_i, t \geq \tau\}$. Ekkor Q_t m.m. tart Q_i^*-hoz $\Omega_{i,\tau}$-n a 4.3. Tétel miatt, mivel τ után és az $\Omega_{i,\tau}$-n a Q-tanulás algoritmus úgy működik, mint egy az $\mathcal{X}_i$-re megszorított MDP-hoz tartozó Q-tanulás algoritmus és π erősen elegendő felfedező $\Omega_{i,\tau}$-n és a megszorított MDP-t tekintve. Mivel m.m. $\cup_{\tau\in\mathbf{N}}\Omega_{i,\tau} = \{\omega : \mathcal{X}_\infty(\omega) = \mathcal{X}_i\}$, így $Q_t(\omega)\big|_{\mathcal{X}_i} \to Q_i^*$, $t \to \infty$ m.m. tart $\mathcal{X}_i = \mathcal{X}_\infty$-n és a tétel következik (30)-ból. □

5.0.5. Optimális adaptív politikák konstrukciója. Mivel a t-edik pillanatban a politika Q_t-től is függ majd, a politika által előírt akció a t-edik lépésben nem csak a kontroll folyamat múltjától, hanem az eddigi költségektől, $c_0, c_1, \ldots, c_{t-1}$-től is függ majd. Azonban mivel c_i a $c(\xi_i, \alpha_i, \xi_{i+1}) + n_i$ alakban írható, ahol n_i véges szórású, zérus várható értékű „zaj" az ilyen politikákkal elérhető legkisebb összköltség sem lehet kisebb mint azon politikák által elérhető, melyek a kontroll folyamat múltjától (és a c költségfüggvénytől) függnek. Így az optimális politikák alakja változatlan marad ezen kiterjesztett politikákra is. A továbbiakban a rövidség kedvéért jelöljük a π politika által az a akció a t-edik időpillanatbeli választására előírt valószínűséget $\pi(a \mid \mathcal{F}_t)$-vel, ahol $\mathcal{F}_t$ jelöli a kontroll folyamat múltját az eddigi költségekkel együtt.

A következő, ún. kiterjesztett Borel-Cantelli lemmára a következő lemmánk bizonyításánál szükség lesz.

5.2. LEMMA. *Legyen $\hat{\mathcal{F}}_k$ σ-algebrák egy növekvő sorozata és legyenek az A_k halmazok $\hat{\mathcal{F}}_k$-mérhetőek. Ekkor*

$$\left\{\omega : \sum_{k=1}^{\infty} P(A_k \mid \hat{\mathcal{F}}_{k-1}) = \infty\right\} = \{\omega : \omega \in A_k \text{ i.o.}\} \quad m.m.$$

A lemma bizonyítása a [4] könyvben található 5.29-es következmény bizonyításához teljesen hasonló módon történhet.

5.3. LEMMA. *Tegyük fel, hogy egy politikánál az akció választás valószínűsége egy adott állapotban attól függ, hogy az állapotot eddig hányszor látogatta meg a kontroll folyamat: Jelölje $\pi\big(a \mid \mathcal{F}_t,\ t = t_k(x)\big)$ annak a valószínűségét, hogy a az a akciót választjuk feltéve, hogy a t-edik lépésben épp k-adszor látogatja meg a kontroll folyamat x-et. (Itt $t_k(x)$ azt az időpontot jelöli, amikor a kontrollált folyamat k-adszor látogatja meg az x állapotot.) Ha minden x-re*

$$\sum_{k=1}^{\infty} \pi\big(a \mid \mathcal{F}_t, t_k(x) = t\big) = \infty, \tag{31}$$

akkor a π politika elegségesen felfedező.

Bizonyítás. Rögzítsük a π politikát és egy $x \in \mathcal{X}$ állapotot. Ekkor minden olyan ω-ra, amelyre $x \in X_\infty(\omega)$ a $t_k(x)$ sorozat jól definiált és a végtelenig folytatható és konstrukció szerint

$$P\big(a_{t_k(x)} = a \mid \mathcal{F}_t, t_k(x) = t\big) = \pi\big(a \mid \mathcal{F}_t, t_k(x) = t\big)$$

áll minden $k \geq 1$-re. Alkalmazzuk az 5.2. lemmát. Legyen $A_k = \left\{a_{t_k(x)} = a\right\}$ és $\hat{\mathcal{F}}_k = \sigma\left\{\mathcal{F}_t,\ t = t_{k+1}(x)\right\} = \cup_{t=t_k(x)} \mathcal{F}_t$. Ekkor A_k $\hat{\mathcal{F}}_k$-mérhető és így (31) miatt

$$P(A_k \mid \hat{\mathcal{F}}_{k-1}) = P\big(a_{t_k(x)} = a \mid \mathcal{F}_t,\ t_k(x) = t\big) = \pi\big(a \mid \mathcal{F}_t, t_k(x) = t\big).$$

Következésképpen m.m. $\{x \in \mathcal{X}^\infty\}$-n $\sum_{k=1}^{\infty} P(A_k \,|\, \hat{\mathcal{F}}_{k-1}) = \infty$ és így az 5.2. lemma szerint $\left\{a = a_{t_k(x)}\right\}$ v.s. is m.m. áll $\{x \in \mathcal{X}^\infty\}$-en. □

5.1. Definíció. Egy π politikát *aszimptotikusan optimálisnak* nevezünk, ha minden $x \in X$-re

$$\lim_{t\to\infty} P(a_t \in \operatorname{Argmin} Q^*(x,a) \,|\, \mathcal{F}_t,\ x = x_t) = 1 \quad \text{m.m.}$$

5.4. TÉTEL. *Tekintsük a (29)-vel megadott Q_t sorozatot és tegyük fel, hogy az 1.1. feltevésben foglaltak teljesülnek. Legyen π egy politika és legyen*

$$\pi\Big(a_t \in \operatorname*{Argmin}_{a\in\mathcal{A}} Q_t(x,a) \,|\, x = x_t,\ \mathcal{F}_t\Big) \stackrel{\text{def}}{=} \sum_{a\in\operatorname{Argmin}_{b\in\mathcal{A}} Q_t(x,b)} \pi(a \,|\, x = x_t,\ \mathcal{F}_t).$$

Tegyük fel, hogy π konstrukciójánál fogva kielégíti a

$$\sum_{k=1}^{\infty} \pi\big(a \,|\, \mathcal{F}_t,\ t_k(x) = t\big) = \infty \tag{32}$$

$$\lim_{t\to\infty} \pi\Big(a_t \in \operatorname*{Argmin}_{a\in\mathcal{A}} Q_t(x,a) \,|\, x = x_t,\ \mathcal{F}_t\Big) = 1. \tag{33}$$

egyenleteket. Ekkor π aszimptotikusan optimális.

Megjegyezzük, hogy $\pi\big(a_t \in \operatorname{Argmin}_{a\in\mathcal{A}} Q_t(x,a) \,|\, x = x_t, \mathcal{F}_t\big)$ jól definiált, mert Q_t $\mathcal{F}_t$ mérhető.

Bizonyítás. (32) szerint az 5.3. Lemma feltételei teljesülnek és így a 5.1. tétel is alkalmazható. Eszerint $\lim_{t\to\infty} Q_t|_{\mathcal{X}_\infty} \to Q^*|_{\mathcal{X}_\infty}$ m.m. és így, mivel $\mathcal{A}$ véges és (33) miatt már következik is π aszimptotikus optimalitása. □

A tételből azonnal adódik, hogy a

$$\pi\big(a \,|\, \mathcal{F}_t, t_k(x) = t\big) = \begin{cases} \frac{1}{C_{t_k}}\left(1 - \frac{1}{k+1}\right); & \text{ha } a \in \operatorname{Argmin}_{a\in\mathcal{A}} Q_{t_k(x)}(x,a), \\ \dfrac{1 - \frac{1}{C_{t_k}}\left(1 - \frac{1}{k+1}\right)}{|\mathcal{A} - C_{t_k}|}; & \text{különben,} \end{cases}$$

politika, ahol $C_t(x)$ az $\operatorname{Argmin}_{a\in\mathcal{A}} Q_t(x,a) = \{a \in \mathcal{A} \,|\, Q_t(x,a) = \min_{b\in\mathcal{A}} Q_t(x,b)\}$ halmaz számosságát jelöli kielégíti az 5.4. tétel feltételeit. Vegyük észre, hogy ez a politika a múlttól csak a Q_t-n keresztül függ és így valóban kiszámítható. Természetesen aszimptotikusan optimális politikák más, nem-egyenletes valószínűségekkel is megadhatók (a fenti politika egyenletes valószínűségekkel választ az mohó és nem mohó akciók közül). Nyitott kérdés, hogy az ily módon előállított aszimptotikusan

optimális politikák közül mely valószínűségeloszlások adnak gyorsabb optimalitáshoz való tartást. Megjegyezzük, hogy a 2.1. tétel bizonyításánál felhasznált módszerrel belátható, hogy a Q_t Q^*-hoz tartásának rátája tetszőleges aszimptotikusan optimális politika esetére $\sqrt{\log\log t/t}$ [14] és ezt (32)-vel ötvözve lehetségesnek tűnik az optimalitáshoz tartás aszimptotikus konvergencia sebességének becslése. Ha egy politikáról tudjuk, hogy aszimptotikusan optimális, akkor a sztochasztikus approximáció közönséges differenciálegyenletekre való visszavezetésének módszere (az „ODE" módszer, lásd [11, 9, 2]) már használható centrális határeloszlás típusú tételek levezetésére.

Arra az esetre, amikor a MDP kritériuma az egy lépésre jutó átlagos költség minimamizálása ismertek a konvergencia sebességre vonatkozó alsó korlátok, pontosabban alsó korlátok arra nézve, hogy a tanulásból eredő az optimális politika végrehajtásának költségéhez képesti veszteség (a „tanuló pénz") legalább mekkora aszimptotikusan. Erre az esetre ismertek továbbá olyan politikák, amelyek ezt az alsó korlátot elérik és ilyen értelemben optimálisak, nem javíthatók [5, 6]. Fontos azonban megjegyezni, hogy ezek az eredmények csak a tranziens teljesítmény aszimptotikájáról szólnak és nem magáról a tranziens teljesítményről, melyről feltehetőleg semmi sem mondható.

Érdekességként megemlítjük, hogy ha a MDP kritériuma a pesszimista összköltség minimalizálásaként van megadva, akkor a minden lépésben a mohó akciót választó politika is optimális lesz [13, 12].

A. Egy aszinkron sztochasztikus approximációs tétel

Itt ismertetjük a 2.1. tétel bizonyításának vázlatát. A bizonyítás az alábbi A.1.–A.4. lemmákon alapszik.

A.1. LEMMA. *Legyen x_t egy véletlen folyamat és tegyük fel, hogy minden η, δ pozitív számokhoz van olyan $M \in \mathbf{N}$, m.m. korlátos véletlen index, hogy*

$$P\Big(\sup_{t\geq M} |x_t| \geq \delta\Big) < \eta. \tag{34}$$

Ekkor x_t m.m. 0-hoz tart.

Bizonyítás. A fenti feltételek mindössze abban különböznek a szokásos definíciótól, hogy itt megengedjük, hogy M véletlen legyen. Azonban elemien megmutatható, hogy $P(\sup_{t\geq k} |x_t| \geq \delta) \leq P(\sup_{t\geq M} |x_t| \geq \delta) + P(M > k)$, ahol k tetszőleges (nem-véletlen) természetes szám és így (34)-ből és M m.m. korlátosságából következik az eredmény. □

A.2. LEMMA. *Legyen $\mathcal{X}$ egy tetszőleges halmaz és tekintsünk egy a*

$$v_{t+1} \leq g_t v_t + \gamma(1 - g_t)\|v_t\| \tag{35}$$

egyenlőtenségeket kielégítő $\{v_t\}$ *véletlen folyamatot, ahol* v_0, g_t *nem-negatív véletlen függvények, és* $\|v_0\|$ m.m. *véges. Ha minden* $k \geq 0$*-ra*

$$\lim_{n\to\infty} \left\| \prod_{t=k}^{n} g_t(\cdot) \right\| = 0$$

m.m., *akkor* $\|v_t\|$ *is* m.m. *nullához tart.*

Bizonyítás. A bizonyítás teljes indukcióval történik. A k-adik lépésben azt bizonyítjuk, hogy van olyan véletlen m.m. korlátos M_k index, hogy ha $t > M_k$, akkor $\|v_t\| \leq \big((1+\gamma)/2\big)^k C$. Ez az A.1. lemma miatt és mivel $\big((1+\gamma)/2\big)^k C \to 0$, m.m. $k \to \infty$, elegendő is. A kérdéses egyenlőtlenség $k = 0$-ra a v_0-ra kirótt feltevésünk miatt áll az $M_0 = 0$ választással. Tegyük fel, hogy az egyenlőtlenséget már valamely $k \geq 0$ egész számig beláttuk. Ekkor, mivel az $u_{M_k} = v_{M_k}$, $u_{t+1} = g_t u_t + \gamma(1 - g_t)\big((1+\gamma)/2\big)^k C$, $t \geq M_k$ sorozatra $0 \leq v_t \leq u_t$ és $\left\|u_t - \gamma\big((1+\gamma)/2\big)^k C\right\| \to 0$ m.m. megmutatható, így $\limsup_{t\to\infty} \|v_t\| \leq \gamma\big((1+\gamma)/2\big)^k C < \big((1+\gamma)/2\big)^{k+1} C$, amiből már következik is a m.m. korlátos M_{k+1} létezése. □

A.1. Definíció. Legyen $G : \mathcal{B}_1 \times \mathcal{B}_2 \to \mathcal{B}_1$ egy véletlentől függő leképezés, ahol $\mathcal{B}_1, \mathcal{B}_2$ normált vektorterek. G-t homogénnak nevezzük, ha minden pozitív β-ra és $v \in \mathcal{B}_1, \varepsilon \in \mathcal{B}_2$-re $\beta G(v, \varepsilon) = G(\beta v, \beta\varepsilon)$.

A.2. Jelölés. Legyen $\varepsilon = (\varepsilon_0, \varepsilon_1, \ldots)$ $\mathcal{B}_2$-értékű véletlen sorozat és $G_t : \mathcal{B}_1 \times \mathcal{B}_2 \to \mathcal{B}_1$ véletlen leképezések sorozata. A továbbiakban a $v_{t+1} = G_t(v_t, \varepsilon_t)$ véletlen folyamat t-edik elemét, v_t-t, $v_t(v_0; \varepsilon)$-val jelöljük, ahol $\|v_0\| < \infty$ m.m.

A.3. Definíció. A $v_t(v_0; \varepsilon)$ véletlen folyamatot az ε sorozat véges perturbációira érzéketlennek nevezzük, ha minden ε' véletlen sorozatra amely ε-tól legfeljebb véges sok tagban tér el úgy, hogy az eltérések maximális száma a véletlentől független, áll, hogy ha $\|v_t(v_0; \varepsilon)\|$ m.m. nullához tart, akkor $\|v_t(v_0; \varepsilon')\|$ is m.m. nullához tart.

A.4. Definíció. A $v_t(v_0; \varepsilon)$ véletlen folyamatot érzéketlennek nevezzük az ε sorozat kicsinyítéseire, ha minden $0 < c \leq 1$ véletlen számra áll, hogy ha $\|v_t(v_0; \varepsilon)\|$ m.m. tart nullához, akkor $\|v_t(v_0; \varepsilon')\|$ is m.m. tart nullához.

A.3. LEMMA (Újraskálázási Lemma). *Tegyük fel, hogy a* $v_t(w; \varepsilon)$ *véletlen sorozatot egy véletlen, homogén* $G_t : \mathcal{B}_1 \times \mathcal{B}_2 \to \mathcal{B}_1$ *függvénysorozat indukálja. Ekkor* $\|v_t(w; \varepsilon)\|$ m.m. *nullához tart, ha*

(i) v_t *érzéketlen az* ε *véges perturbációira,*

(ii) v_t *érzéketlen az* ε *kicsinyítéseire és*

(iii) $\|u_t(w; \varepsilon)\|$ m.m. *nullához tart, ahol* u_t *a* v_t *korlátosan tartott változata:* $u_0 = v_0$ *és* $u_{t+1} = S_t G_t(u_t, \varepsilon_t)$, *ahol* S_t *a* t*-edik ún. átskálázási faktor, mely a kö-*

vetkezőképp van definiálva:

$$S_t = \begin{cases} 1/\|G_t(u_t,\varepsilon_t)\|, & \text{ha } \|G_t(u_t,\varepsilon_t)\| > 1; \\ 1, & \text{különben.} \end{cases}$$

Bizonyítás. Először is jegyezzük meg, hogy $0 < S_t \le 1$, $t \ge 0$. Teljes indukcióval belátható, hogy tetszőleges pozitív S-ra áll az $Sv_t(u;\varepsilon) = v_t(Su, S\varepsilon)$ azonosság, itt u, ε tetszőlegesek. Ebből, szintén teljes indukcióval következik, hogy a $c_{t+1} = \prod_{j=t}^{\infty} S_j$ és $d_{t+1} = \prod_{i=1}^{t} S_i$ sorozatokkal $u_t(w;\varepsilon) = v_t(d_t w; c\varepsilon)$, ahol $c = (c_0, c_1, \ldots)$ és ahol a $c\varepsilon$ szorzat tagonként értendő: $c\varepsilon = (c_0\varepsilon_0, c_1\varepsilon_1, \ldots)$.

Mivel feltevés szerint $\|u_t\| \to 0$ m.m., így minden $1 > \delta > 0$-hoz van olyan $M = M(\delta)$ véges index, hogy ha $t > M$, akkor $P(\|u_t\| < \delta) > 1 - \delta$. Tekintsük az $A_\delta = \{\|u_t\| < \delta\}$ halmaz ω eseményeit. Mivel itt $S_t(\omega) = 1$, ha $t > M$, így $c_{t+1} = \prod_{j=t}^{M} S_j$, ha $t < M$ és $c_{t+1} = 1$, ha $t \ge M$. Hasonlóképp, $d_{t+1} = d_{M+1}$, ha $t \ge M$. Az $u_t(w;\varepsilon) = v_t(d_t w; c\varepsilon)$ azonosságból következően pedig $\|v_t(d_{M+1}w; c\varepsilon)\|$ m.m. nullához tart A_δ-n. Homogenitás miatt $v_t(d_{M+1}w; c\varepsilon) = d_{M+1}v_t(w; c\varepsilon/d_{M+1})$, és így $\|v_t(w; c\varepsilon/d_{M+1})\|$ is m.m. nullához tart A_δ-n. Mivel $v_t(w; c\varepsilon) = v_t\big(w; d_{M+1}(c\varepsilon/d_{M+1})\big)$, $0 < d_{M+1} \le 1$ és mivel v_t érzéketlen ε kicsinyítéseire, így $\|v_t(w; c\varepsilon)\|$ is m.m. nullához tart A_δ-n. Végül, mivel A_δ-n $c\varepsilon$ az ε véges perturbációja kapjuk, hogy $\|v_t(w;\varepsilon)\|$ is m.m. nullához tart A_δ-n. Mármost az állítás ebből és abból, hogy δ tetszőleges volt és $\lim_{\delta\to 0} P(A_\delta) \to 1$ már következik is. □

A következő lemma az A.2. lemma folyamatának perturbált változatára vonatkozik:

A.4. LEMMA. *Legyen $\mathcal{X}$ egy tetszőleges halmaz és tekintsünk egy a*

((36)) $$v_{t+1} \le g_t v_t + \gamma(1 - g_t)\big(\|v_t\| + \varepsilon_t\big)$$

egyenlőtlenségeket kielégítő $\{v_t\}$ véletlen folyamatot, ahol v_0, g_t nem-negatív véletlen függvények, $\|v_0\| < \infty$ m.m. és $0 \le \varepsilon_t \to 0$ m.m. Ha minden $k \ge 0$-ra

$$\lim_{n\to\infty} \left\| \prod_{t=k}^{n} g_t(\cdot) \right\| = 0$$

m.m., akkor $\|v_t\|$ is m.m. nullához tart.

Bizonyítás. Figyeljük meg, hogy $v_t = v_t(v_0;\varepsilon)$, a $G_t : B(\mathcal{X}) \times \mathbf{R} \to B(\mathcal{X})$, $G_t(v,\varepsilon) = g_t v + f_t\big(\|v\| + \varepsilon_t\big)$ homogén függvényekkel. Teljes indukcióval könnyen látszik, hogy $v_t(v_0;\varepsilon)$ ε kicsinyítéseire érzéketlen, mivel ha $0 < c \le 1$, akkor $0 \le v_t(w; c\varepsilon) \le v_t(w;\varepsilon)$, $w \ge 0$. Hasonlóképp, $v_t(v_0;\varepsilon)$ érzéketlen ε véges perturbációira. Ezt az A.2. lemma $\delta_t = \big|v_t(v_0;\varepsilon) - v_t(v_0;\varepsilon')\big|$ folyamatra való alkalmazásával bizonyíthatjuk, ugyanis elég nagy t-re $\varepsilon_t = \varepsilon'_t$ és ekkor δ_t m.m. kielégíti a $\delta_{t+1} \le g_t\delta_t + \gamma(1 - g_t)\|\delta_t\|$ egyenlőtlenséget. Így az újraskálázási lemma szerint

elegendő ha v_t korlátosan tartott verziójának m.m. nullához tartását belátjuk. Ez azonban az A.2. lemma bizonyításához teljesen hasonlóan következik. □

A 2.1. tétel bizonyítása ezek után már könnyen adódik. A tétel szövegét az olvasó kényelme érdekében megismételjük:

A.5. TÉTEL. *Legyen $\mathcal{X}$ egy tetszőleges halmaz, $\mathcal{B} = \big(B(\mathcal{X}), \|\cdot\|\big)$, ahol $\|v\| = \sup_{x\in\mathcal{X}} |v(x)|$ és legyen $T : B(\mathcal{X}) \to B(\mathcal{X})$ egy fixponttal rendelkező operátor. Jelöljük v^*-val T egy fixpontját és tegyük fel, hogy a $\mathcal{T} = (T_0, T_1, \ldots)$ véletlen operátorok sorozata approximálja T-t a v^*-nál, az $F_0 \subseteq B(\mathcal{X})$ kezdeti értékek mellett. Tegyük fel továbbá, hogy $v^* \in F_0$, F_0 invariáns $\mathcal{T}$-re és hogy léteznek olyan $g_t : \mathcal{X} \to [0,1]$ mérhető függvények és egy olyan $0 < \gamma < 1$ konstans, hogy az alábbi feltételek elegendően nagy t-re m.m. teljesülnek:*

1. $\big|T_t(u_1, v^*)(x) - T_t(u_2, v^*)(x)\big| \le g_t(x)\big|u_1(x) - u_2(x)\big|$, *ahol* $t \in \mathbf{N}$, $x \in \mathcal{X}$ *és* $u_1, u_2 \in F_0$.
2. $\big|T_t(u, v)(x) - T_t(u, v^*)(x)\big| \le \gamma\big(1 - g_t(x)\big)\big(\|v - v^*\| + \lambda_t\big)$, *ahol* $t \in \mathbf{N}$, $x \in \mathcal{X}$ *és* $u, v \in F_0$ *és* $\lambda_t \to 0$ m.m.
3. $\lim_{n\to\infty} \big\| \prod_{t=k}^{n} g_t(\cdot) \big\| = 0$, $k \ge 0$.

Ekkor tetszőleges $v_0 \in F_0$-ra, a $v_{t+1} = T_t(v_t, v_t)$ függvénysorozat v^*-hoz tart m.m. a $B(\mathcal{X})$ normájában.

Bizonyítás. Mivel v^* a T fixpontja, elég a $\delta_t = v_t - u_t$ sorozat egyenletes nullához tartását megvizsgálni, ahol $u_{t+1} = T_t(u_t, v^*)$ és $u_0 \in F_0$. Mivel $v^*, u_0 \in F_0$ és F_0 invariáns $\mathcal{T}$-re, így $v_t, u_t \in F_0$ is áll, a tétel 1–2 feltételeinek egyenlőtlenségei tehát alkalmazhatók. Így a háromszög egyenlőtlenség ismételt alkalmazásával és elegendően nagy t-kre adódik, hogy

$$\delta_{t+1} \le g_t\delta_t + \gamma(1 - g_t)\big(\|\delta_t\| + \|u_t - v^*\| + \lambda_t\big),$$

amiből az A.4. lemma szerint $\|\delta_t\| \to 0$ m.m., azaz $\|v_t - v^*\| \to 0$ m.m. □

IRODALOM

[1] A. G. Barto, S. J. Bradtke, S. P. Singh, „Real-time Learning and Control using Asynchronous Dynamic Programming", Technical Report 91–57, Computer Science Department, University of Massachusetts, 1991.

[2] A. Benveniste, M. Métivier, P. Priouret, *Adaptive Algorithms and Stochastic Approximations* (Springer Verlag, New York, 1990).

[3] D. Blackwell, Discounted dynamic programming, *Annals of Math. Statistics* **36** (1965) 226–235.

[4] L. Breiman, *Probability* (Addison-Wesley Publishing Company, Reading, 1968).

[5] A. N. Burnetas, M. N. Katehakis, Optimal adaptive policies for Markov Decision Processes, *Mathematics of Operations Research* **22(1)** (1997).

[6] T. L. Graves, T. L. Lai, Asymptotically efficient adaptive choice of control laws in controlled Markov chains, *SIAM J. Contr. and Opt.* **35(3)** (1997) 715–743.

[7] M. Heger, Consideration of risk in reinforcement learning, 1994. Revised submission to the 11th International Machine Learning Conference ML-94.

[8] Tommi Jaakkola, Michael I. Jordan, Satinder P. Singh, On the convergence of stochastic iterative dynamic programming algorithms, *Neural Computation* **6(6)** (November 1994) 1185–1201.

[9] H. J. Kushner, D. S. Clark, *Stochastic approximation methods for constrained and unconstrained systems,* (Springer-Verlag, Berlin, Heidelberg, New York, 1978).

[10] M. L. Littman, Cs. Szepesvári, „A Generalized Reinforcement Learning Model: convergence and applications", in: *Int. Conf. on Machine Learning,* 1996. http://iserv.iki.kfki.hu/asl-publs.html.

[11] L. Ljung, Analysis of recursive stochastic algorithms, *IEEE Tran. Automat. Control* **22** (1977), 551–575.

[12] Cs. Szepesvári, „Certainty equivalence policies are self-optimizing under minimax optimality", Technical Report 96-101, Research Group on Artificial Intelligence, JATE-MTA, Szeged 6720, Aradi vrt tere 1., HUNGARY, August 1996. e-mail: szepes@math.u-szeged.hu, http://www.inf.u-szeged.hu/ rgai.

[13] Cs. Szepesvári, „Learning and exploitation do not conflict under minimax optimality", in: *Machine Learning: ECML'97* (9th European Conf. on Machine Learning, Proceedings), M. van Someren and G. Widmer, editors (Springer, Berlin, Prague, Czech Republic, April 1997) 242–249.

[14] Cs. Szepesvári, „On the asymptotic convergence rate of Q-learning", in: *Proc. of Neural Information Processing Systems* (1997), in press.

[15] Cs. Szepesvári, M. L. Littman, Generalized Markov Decision Processes: Dynamic programming and reinforcement learning algorithms, *Neural Computation* (1997), in preparation.

[16] C. H. C. Ribeiro, Cs. Szepesvári, „Q-learning combined with spreading: convergence and results", in: *Proceedings of ISRF-IEE International Conference: Intelligent and Cognitive Systems, Neural Networks Symposium,* (Tehran, Iran, 1996) 32–36.

[17] H. Robbins, D. Siegmund, „A convergence theorem for non-negative almost supermartingales and some applications", in: *Optimizing Methods in Statistics,* J. Rustagi, editor (Academic Press, New York, 1971) 235–257.

[18] S. M. Ross, *Applied Probability Models with Optimization Applications* (Holden Day, San Francisco, California, 1970).

[19] J. N. Tsitsiklis, Asynchronous stochastic approximation and Q-learning, *Machine Learning* **8(3–4)** (1994), 257–277.

[20] C. J. C. H. Watkins, „Learning from Delayed Rewards", PhD thesis. King's College, Cambridge, 1990. QLEARNING.

(Beérkezett: 1997. október 29.)

SZEPESVÁRI CSABA
MINDMAKER KFT.
1112 BUDAPEST
KONKOLY TH. M. U. 29–33.
E-mail: szepes@mindmaker.hu

AN ASYNCHRONOUS STOCHASTIC APPROXIMATION THEOREM AND SOME APPLICATIONS

CSABA SZEPESVÁRI

In the paper we are concerned with the adaptive optimal control of Markov decision problems. An asynchronous stochastic approximation theorem is proved. Then the theory of Markov decision problems is reviewed and three applications of the main theorem are given for the estimation of the optimal cost-function. Finally, a class of adaptive optimal policies is constructed.

Alkalmazott Matematikai Lapok **19** (1999) 69–85

HIPERCSERESZNYEFÁKKAL ADOTT VALÓSZÍNŰSÉGI KORLÁTOK

SZÁNTAI TAMÁS ÉS BUKSZÁR JÓZSEF

Budapest, Miskolc

Ebben a dolgozatban új típusú alsó és felső korlátokat adunk meg véges számú esemény uniójának valószínűségére. Ehhez segédeszközként a hipergráfok körében egy új fogalmat vezetünk be, melyet hipercseresznyefának fogunk nevezni. Ezeknek az új eredményeknek közvetlen előzménye a Prékopa András és Bukszár József által korábban bevezetett, cseresznyefának, illetve speciális esetben t-cseresznyefának nevezett (valójában ugyancsak hipergráfnak tekinthető) hipergráfstruktúrák által származtatott felső korlátok, valamint a Bukszár József által definiált m-multifák által származtatott felső korlátok. Ezek mind a D. Hunter által közönséges gráfok maximális súlyú feszítőfája segítségével származtatott felső korlát javításainak tekinthetők, s mint ilyenek értelemszerűen csak felső korlátokat szolgáltathatnak. I. Tomescu a hipergráfok körében bevezetett hiperfák segítségével úgy tudta általánosítani D. Hunter eredményét, hogy nemcsak további felső korlátokat nyert, hanem hasonlóan jó alsó korlátokat is. A dolgozatban közölt új típusú alsó és felső korlátok ugyanolyan értelemben általánosítják, illetve javítják az I. Tomescu féle alsó és felső korlátokat, mint ahogyan a Prékopa András és Bukszár József által bevezetett felső korlátok javították a D. Hunter féle felső korlátot. Az új korlátok hatékonyságát egy speciális megbízhatósági rendszer megbízhatóságára számított alsó és felső korlátokkal illusztráljuk.

1. Bevezetés

Legyenek $A_1, \ldots, A_n$ az $(\Omega, \mathcal{A}, P)$ valószínűségi mező tetszőleges eseményei. Ezek uniójának a valószínűségére legkorábban G. Boole [1] adott meg egy felső korlátot, melyben csak a $P(A_i)$, $i = 1, \ldots, n$ valószínűségek összegét, azaz az úgynevezett első binomiális momentumot, S_1-et használta. Ezt a felső korlátot általánosították a C. E. Bonferroni által adott egyenlőtlenségek (lásd [2]), melyek szerint az $S_1, S_2, \ldots, S_h, h < n$ binomiális momentumok váltakozó előjelű összege páratlan h esetén mindig felső korlátot, páros h esetén pedig alsó korlátot ad az $A_1, \ldots, A_n$ események uniójának a valószínűségére. Ezt követően a Bonferroni egyenlőtlenségek általánosításainak egy sora jelent meg, míg végül Prékopa András dolgozatai zárták le a kutatások ezen irányát azzal, hogy bebizonyították az addig megtalált, első három binomiális momentumot használó Bonferroni típusú egyenlőtlensé-

gekről, hogy azok élesek, valamint az első négy binomiális momentumot használó éles felső korlát explicit megadása mellett felírták azt a lineáris programozási feladatpárt, amely megoldásai akármennyi $S_1, S_2, \ldots, S_h, h < n$ binomiális momentum ismeretében szolgáltatják az éles Bonferroni típusú alsó és felső korlátokat.

Mivel a binomiális momentumok számítása a gyakorlatban legtöbbször az

$$S_k = \sum_{1 \le i_1 < \ldots < i_k \le n} P(A_{i_1} \cap \cdots \cap A_{i_k}), \quad k = 1, \ldots, n$$

képlettel történik, azért az első h binomiális momentum ismerete feltételezi azt, hogy a bennük foglalt $P(A_{i_1})$, $P(A_{i_1} \cap A_{i_2}), \ldots, P(A_{i_1} \cap \ldots \cap A_{i_h})$, $1 \le i_1 < \ldots < i_h \le n$ szorzat esemény valószínűségeket is mind külön-külön számítani tudjuk, illetve ki is számítottuk. Ezért logikus az a törekvés, hogy ne csupán az első h binomiális momentumba aggregált, hanem az egyes szorzat eseményekben különkülön meglévő információt is próbáljuk meg a $P(A_1 \cup \cdots \cup A_n)$ valószínűség jobb korlátainak a készítésére felhasználni. Ebben az irányban az első lépéseket már G. Boole is megtette, hiszen voltaképpen megfogalmazta az úgynevezett diszaggregált lineáris programozási feladatpár duál feladatát. Mivel azonban az ő idejében még nem volt ismert annak megoldására egy általános megoldó módszer, azért csupán speciális, egyedi becsléseket tudott a módszerével megadni. G. Boole munkáját Th. Hailperin elevenítette fel (lásd [7]) először, majd ennek, és egymás munkájának ismerete nélkül D. Hunter [8] és K. J. Worsley [14] adott meg olyan felső korlátot a $P(A_1 \cup \cdots \cup A_n)$ valószínűségre, amely az S_1 első binomiális momentumon túl a $P(A_{i_1} \cap A_{i_2}), 1 \le i_1 < i_2 \le n$ szorzat esemény valószínűségek közül csak azokat használja, amelyek egy n csúcsú, az éleket a $P(A_{i_1} \cap A_{i_2}), 1 \le i_1 < i_2 \le n$ valószínűségekkel súlyozó teljes gráf maximális súlyú feszítőfájának élei mentén találhatók. Ugyanakkor az így kapható felső korlátról könnyű belátni, hogy jobb, mint az első két binomiális momentumot használó legjobb Bonferroni típusú felső korlát.

Ezt követően I. Tomescu [13] általánosította a Hunter–Worsley féle felső korlátot úgy, hogy páros h esetén az első $h+1$ binomiális momentum váltakozó előjelű összegéből levonta, illetve páratlan h esetén az első $h+1$ binomiális momentum váltakozó előjelű összegéhez hozzáadta egy speciális hipergráfstruktúra, az úgynevezett $(h+2)$-*hiperfa* (egy speciális $(h+2)$-uniform, azaz olyan hipergráf, mely minden éle $(h+2)$ csúcsból áll) éleinek megfelelő szorzat események valószínűségeinek az összegét. Ezzel nyilvánvaló módon megjavította a Bonferroni-féle felső, illetve alsó korlátokat. Megjegyezzzük, hogy $h = 0$ esetén a Tomescu féle felső korlát azonos a Hunter–Worsley féle felső korláttal. A Tomescu féle korlátok előnye, hogy alsó korlátok is vannak közöttük, sajnos azonban az éles Tomescu féle korlátok meghatározásához a $(h+2)$-hiperfák körében kellene maximális súlyút keresni, amely feladat megoldására általánosan nem ismert hatékony algoritmus, csupán a $h = 0$ esetben tudjuk, hogy az megoldható mohó típusú algoritmusokkal.

A. Prékopa, B. Vizvári és G. Regős [11] abból a feltételezésből indultak ki, hogy egy valószínűségi mező tetszőleges $A_1, \ldots, A_n$ eseményére mind a $P(A_i)$, $i = 1, \ldots, n$ egyedi, mind a $P(A_{i_1} \cap A_{i_2}), \ldots, P(A_{i_1} \cap \cdots \cap A_{i_h})$, $1 \le i_1 < \ldots < i_h \le n$

metszet valószínűségek rendelkezésre állnak. Az ezen információ birtokában megfogalmazott lineáris egyenlőség-egyenlőtlenség rendszerhez lineáris célfüggvényeket hozzávéve és a keletkezett lineáris programozási feladatokra duálmegengedett megoldásokat szerkesztve az n esemény különféle Boole-függvényeire tudtak korlátokat megadni. Sajnos az így felírt lineáris programozási feladatok mérete túl nagy ahhoz, hogy az optimális megoldását, és így az adott információ birtokában lehetséges éles korlát megtalálását 15–20-nál nagyobb n érték esetén egyáltalán remélni lehessen. Speciálisan az n esemény uniójának a valószínűségére azonban meg tudtak adni néhány speciális hipergráfstruktúrát, melyekhez a lineáris programozási feladat egy-egy olyan duálmegengedett megoldását lehetett rendelni, hogy az a Hunter–Worsley korlátnál bizonyíthatóan jobb felső korlátot eredményezett.

J. Bukszár [3] PhD értekezésében ugyancsak speciális, szemléletesen *cseresznyefának* elnevezett hipergráf struktúrát definiált, amely segítségével meglepően jó felső korlátot tudott adni n esemény uniójára. Később J. Bukszár és A. Prékopa [4] azt is megmutatták, hogy a cseresznyefák halmazában található olyan részhalmaz, amely elemeihez szintén tartozik az előbbi lineáris programozási feladat egy duálmegengedett megoldása. Ezeket a speciális cseresznyefákat *t-cseresznyefának* nevezték el, és arra javasolták felhasználni, hogy a hozzátartozó duálmegengedett megoldásból duálmegengedett bázismegoldást készítve, néhány duálszimplex iterációt végrehajtva még jobb felső korlátot nyerjenek.

A jelen dolgozat fő célkitűzése az, hogy a *hipercseresznyefa* fogalmának bevezetésével általánosítsa a cseresznyefa fogalmát, és megmutassa, hogy míg a cseresznyefák segítségével a $h = 0$-ra vett Tomescu-féle felső korlátot lehetett csak javítani, addig ezen új gráfstruktúra alkalmas arra, hogy bármely $h > 0$-ra vett Tomescu-féle, tehát akár alsó, akár felső korlát javítását megadjuk. Megjegyezzük, hogy J. Bukszár a [3]-ban mind a cseresznyefa, mind a hipercseresznyefa fogalmát tovább általánosította az *m-multifa*, illetve a *(h,m)-hipermultifa* fogalmává, amely gráfstruktúrák további, még jobb alsó és felső korlátok szerkesztésére adnak lehetőséget. A (h,m)-hipermultifa hipergráfstruktúra speciális esetként magába foglalja a közönséges cseresznyefa és a hipercseresznyefa fogalmát is.

A 2. fejezetben definiáljuk a hipercseresznyefákat és módszert adunk arra, hogyan származtathatók egyszerűbb hipergráf struktúrákból. A 3. fejezetben leírjuk, hogyan számíthatunk alsó és felső korlátokat a hipercseresznyefák alapján és algoritmusokat adunk meg azon hipercseresznyefák keresésére, melyekből „jó" korlátok nyerhetők. Itt a korlátok pontossága mellett figyelembe kell vennünk a korlátok számításainak költségeit. Számos alkalmazásban ugyanis, mint amilyen például a k-out-of-r-from-n: F rendszer megbízhatóságának becslése, a többváltozós eloszlásfüggvény értékeinek becslése stb., a $P(A_{i_1})$, $P(A_{i_1} \cap A_{i_2})$,... valószínűségeket ki kell értékelnünk és a kiértékelés meglehetősen költséges. Ezekben az alkalmazásokban célszerű olyan algoritmusokat választani, melyek csak kevés $P(A_{i_1})$, $P(A_{i_1} \cap A_{i_2})$,... valószínűségre támaszkodnak, mégis „jó" korlátokat szolgáltatnak. Ebben a fejezetben ilyen algoritmusokat is fogunk látni. Végül a 4. fejezetben olyan numerikus eredményeket közlünk, amelyekben összehasonlítjuk a hipercseresznyefák segítségével számított korlátokat más ismert korlátokkal.

2. Hipercseresznyefák

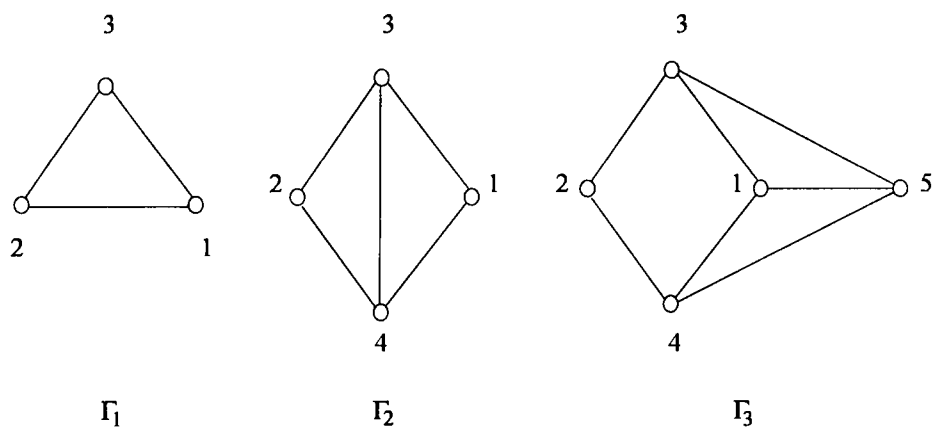
Mielőtt definiálnánk a h-hipercseresznyefákat, felelevenítjük a cseresznyefa fogalmát (lásd [4]).

1. Definíció. Tekintsük azokat a hipergráfokat, amelyeket a $(V, \mathcal{E}, \mathcal{C})$ hármas ír le, ahol V a csúcsokat, $\mathcal{E}$ az éleket (két csúcsból álló halmazok), $\mathcal{C}$ pedig az úgynevezett cseresznyéket (három csúcsból álló halmazok) azonosítja. Ezen nemuniform hipergráfok körében a cseresznyefát az alábbi rekurzióval definiáljuk:

i) A legkisebb cseresznyefa két csúcsból és az őket összekötő élből áll, cseresznyéje nincs.

ii) Egy cseresznyefából egy újabb cseresznyefát nyerhetünk, ha hozzávesszünk egy új csúcsot, két olyan élt, melyek az új csúcsot egy-egy már meglévő csúccsal kötik össze, valamint az új csúcsból és két szomszédjából álló cseresznyét.

iii) Ha az ily módon nyert csúcsok, élek és cseresznyék halmaza rendre V, $\mathcal{E}$ és $\mathcal{C}$, akkor a $\Gamma = (V, \mathcal{E}, \mathcal{C})$ hármast cseresznyefának nevezzük.



1. ábra

1. Példa. Az 1. ábrán három cseresznyefa látható.

- $\Gamma_1 = (V_1, \mathcal{E}_1, \mathcal{C}_1)$, ahol $V_1 = \{1,2,3\}$, $\mathcal{E}_1 = \{\{1,2\},\{1,3\},\{2,3\}\}$, $\mathcal{C}_1 = \{\{1,2,3\}\}$, és például az 1, 2 csúcsokból és az őket összekötő $\{1,2\}$ élből kiindulva a 3 csúcs, valamint a megfelelő $\{1,3\}$, $\{2,3\}$ élek és az $\{1,2,3\}$ cseresznye hozzávételével kapható meg.
- $\Gamma_2 = (V_2, \mathcal{E}_2, \mathcal{C}_2)$, ahol $V_2 = \{1,2,3,4\}$, $\mathcal{E}_2 = \{\{1,3\},\{1,4\},\{2,3\},\{2,4\},\{3,4\}\}$, $\mathcal{C}_2 = \{\{1,3,4\},\{2,3,4\}\}$, és például a 3, 4 csúcsokból és az őket összekötő $\{3,4\}$ élből kiindulva rendre az 1 csúcs, valamint a megfelelő $\{1,3\}$, $\{1,4\}$ élek és az $\{1,3,4\}$ cseresznye, illetve a 2 csúcs, valamint a megfelelő $\{2,3\}$, $\{2,4\}$ élek és a $\{2,3,4\}$ cseresznye hozzávételével kapható meg.

- $\Gamma_3 = (V_3, \mathcal{E}_3, \mathcal{C}_3)$, ahol $V_3 = \{1,2,3,4,5\}$, $\mathcal{E}_3 = \big\{\{1,3\},\{1,4\},\{1,5\},\{2,3\},\{2,4\},\{3,5\},\{4,5\}\big\}$, $\mathcal{C}_3 = \big\{\{1,3,5\},\{1,4,5\},\{2,3,4\}\big\}$, és például az 1, 5 csúcsokból és az őket összekötő $\{1,5\}$ élből kiindulva rendre a 3, 4 és 2 csúcsok valamint a megfelelő élek és cseresznyék hozzávételével kapható meg.

Ezek után tetszőleges $h \geq 0$ egész számra bevezetjük a h-hipercseresznyefa fogalmát. A $\triangle = (V, \mathcal{E}, \mathcal{C})$ h-hipercseresznyefa egy rekurzív módon megadott hipergráf, ahol V a csúcsok, $\mathcal{E}$ $h+2$ csúcsból, $\mathcal{C}$ pedig $h+3$ csúcsból álló halmazok családja, ahol $\mathcal{E}$ elemeit hiperéleknek, $\mathcal{C}$ elemeit pedig hipercseresznyéknek nevezzük. Megjegyezzük, hogy $h = 0$ esetén a hiperélek és hipercseresznyék halmaza közönséges élek és cseresznyék halmaza lesz, és amint azt az alábbi rekurzív definícióból láthatjuk, maga a h-hipercseresznyefa is közönséges cseresznyefává válik.

2. Definíció. A h-hipercseresznyefák rekurzív definíciója.

A 0-hipercseresznyefák legyenek azonosak a cseresznyefákkal. Tetszőleges $h \geq 1$ egész szám esetén tegyük fel, hogy a $(h-1)$-hipercseresznyefát már definiáltuk.

i) A legkisebb $\triangle = (V, \mathcal{E}, \mathcal{C})$ h-hipercseresznyefa csúcsainak száma $h+2$, $\mathcal{E}$-nek egyetlen eleme van, mégpedig az összes csúcsot magába foglaló hiperél, $\mathcal{C}$ pedig üres halmaz.

ii) Egy $\triangle = (V, \mathcal{E}, \mathcal{C})$ h-hipercseresznyefából egy eggyel több csúcsú $\triangle' = (V', \mathcal{E}', \mathcal{C}')$ h-hipercseresznyefát a következő módon nyerünk. Tekintsünk a V csúcsok halmazán egy tetszőleges $\Gamma = (V, \mathcal{E}^*, \mathcal{C}^*)$ $(h-1)$-hipercseresznyefát. A V-hez vegyünk hozzá egy új csúcsot, melyet jelöljön v, az $\mathcal{E}$-hez vegyük hozzá a Γ $\mathcal{E}^*$-beli hiperéleit v-vel kiegészítve, a $\mathcal{C}$-hez pedig vegyük hozzá a Γ $\mathcal{C}^*$-beli hipercseresznyéit ugyancsak v-vel kiegészítve. Azaz

$$V' = V \cup \{v\}, \qquad \mathcal{E}' = \mathcal{E} \cup \bigcup_{E \in \mathcal{E}^*} \{E \cup \{v\}\},$$

$$\mathcal{C}' = \mathcal{C} \cup \bigcup_{C \in \mathcal{C}^*} \{C \cup \{v\}\}$$

esetén $\triangle' = (V', \mathcal{E}', \mathcal{C}')$ egy újabb h-hipercseresznyefa.

2. Példa. Az alábbi hipergráfok 1-hipercseresznyefák.

- $\triangle_1 = (V_1, \mathcal{E}_1, \mathcal{C}_1)$, ahol $V_1 = \{1,2,3\}$, $\mathcal{E}_1 = \big\{\{1,2,3\}\big\}$, $\mathcal{C}_1 = \emptyset$ a legkisebb 1-hipercseresznyefa.
- $\triangle_2 = (V_2, \mathcal{E}_2, \mathcal{C}_2)$, ahol $V_2 = \{1,2,3,4\}$, $\mathcal{E}_2 = \mathcal{E}_1 \cup \big\{\{1,2,4\},\{1,3,4\},\{2,3,4\}\big\}$, $\mathcal{C}_2 = \mathcal{C}_1 \cup \big\{\{1,2,3,4\}\big\}$, a $\triangle_1$-ből az 1. ábrán látható Γ_1 cseresznyefa (azaz 0-hipercseresznyefa) alapján képzett 1-hipercseresznyefa.
- $\triangle_3 = (V_3, \mathcal{E}_3, \mathcal{C}_3)$, ahol $V_3 = \{1,2,3,4,5\}$, $\mathcal{E}_3 = \mathcal{E}_2 \cup \big\{\{1,3,5\},\{1,4,5\},\{2,3,5\},\{2,4,5\},\{3,4,5\}\big\}$, $\mathcal{C}_3 = \mathcal{C}_2 \cup \big\{\{1,3,4,5\},\{2,3,4,5\}\big\}$, a $\triangle_2$-ből az 1. ábrán látható Γ_2 cseresznyefa alapján képzett 1-hipercseresznyefa.
- $\triangle_4 = (V_4, \mathcal{E}_4, \mathcal{C}_4)$, ahol $V = \{1,2,3,4,5,6\}$, $\mathcal{E}_4 = \mathcal{E}_3 \cup \big\{\{1,3,6\},\{1,4,6\},\{1,5,6\},\{2,3,6\},\{2,4,6\},\{3,5,6\},\{4,5,6\}\big\}$, $\mathcal{C}_4 = \mathcal{C}_3 \cup \big\{\{1,3,5,6\},\{1,4,5,6\},$

$\{2,3,4,6\}\}$, a $\triangle_3$-ból az 1. ábrán látható Γ_3 cseresznyefa alapján képzett 1-hipercseresznyefa.

Látható, hogy a Γ_1, Γ_2 és Γ_3 cseresznyefák egymástól teljesen függetlenül voltak választhatók.

1. LEMMA. *Egy n-csúcsú h-hipercseresznyefa hiperéleinek száma $2\binom{n-2}{h+1} + \binom{n-2}{h}$, hipercseresznyéinek száma $\binom{n-2}{h+1}$.*

Bizonyítás. Az állítást h-ra való indukcióval igazoljuk. A $h = 0$ esetén a cseresznyefát kapjuk vissza, mely éleinek száma $2n-3$, cseresznyéinek száma $n-2$. Tegyük fel, hogy az állítás h helyére $h-1$-et írva teljesül. A $\triangle$-át előállító rekurzió során a (legkisebb) $h+2$-csúcsú h-hipercseresznyefából indulunk ki, mely hiperéleinek száma 1, hipercseresznyéinek száma pedig 0. A rekurzió során, amikor egy újabb csúcsot veszünk hozzá a k $(k = h+2, \ldots, n-1)$ meglévőhöz, az indukciós állítás értelmében $2\binom{k-2}{h} + \binom{k-2}{h-1}$ új hiperélt és $\binom{k-2}{h}$ új hipercseresznyét veszünk hozzá a meglévőkhöz. A $\triangle$ hiperéleinek száma tehát

$$\sum_{k=h+1}^{n-1} 2\binom{k-2}{h} + \binom{k-2}{h-1} = 2\binom{n-2}{h+1} + \binom{n-2}{h},$$

hipercseresznyéinek száma pedig

$$\sum_{k=h+2}^{n-1} \binom{k-2}{h} = \binom{n-2}{h+1}. \qquad \square$$

2. TÉTEL. *Legyenek g, h és n egész számok, melyekre $0 \le g < h$, $h+2 \le n$. Legyen $\theta = (V, \mathcal{E}, C)$ egy g-hipercseresznyefa, melyre $V = \{1, \ldots, n-h+g\}$ és tegyük fel, hogy a csúcsok egy a θ-át előállító rekurzió lépései szerint vannak számozva. Ha*

$$\mathcal{E}' = \bigcup_{\{i_1,\ldots,i_{g+2}\}\in\mathcal{E}} E_{\{i_1,\ldots,i_{g+2}\}}, \qquad C' = \bigcup_{\{i_1,\ldots,i_{g+3}\}\in C} C_{\{i_1,\ldots,i_{g+3}\}}$$

ahol $E_{\{i_1,\ldots,i_{g+2}\}} = \Big\{H \cup \{i_1,\ldots,i_{g+2}\} \,|\, H \subset \big\{\max\{i_1,\ldots,i_{g+2}\}+1,\ldots,n\big\}, |H| = h-g\Big\}$ *és* $C_{\{i_1,\ldots,i_{g+3}\}} = \Big\{H \cup \{i_1,\ldots,i_{g+3}\} \,|\, H \subset \big\{\max\{i_1,\ldots,i_{g+3}\}+1,\ldots,n\big\}, |H| = h-g\Big\}$, *akkor* $\triangle = (V', \mathcal{E}', C')$ *egy h-hipercseresznyefa, melyre* $V' = \{1,\ldots,n\}$.

Bizonyítás. A h-ra és n-re való indukcióval bizonyítunk. A $h = g$ esetén a tétel állítása triviális, hiszen $\triangle$ megegyezik θ-val. Rögzítsük h-át $(g+1 \le h)$ és tegyük fel, hogy a tétel állítása h helyére $h-1$-et írva teljesül minden $h+2 \le n$-re. Az n-re való indukció értelmében tegyük fel, hogy $n = h+2$. Ekkor a tételben szereplő $\theta = (V, \mathcal{E}, C)$ csúcsainak száma $n-h+g = g+2$. A θ tehát a legkisebb g-hipercseresznyefa, amiből következik, hogy $\mathcal{E}$ egyetlen eleme $\{1,\ldots,g+2\}$ és

$\mathcal{C} = \emptyset$. A θ-ából a tételben leírt módon valóban h-hipercseresznyefát kapunk, mely csúcsainak halmaza $\{1, \ldots, h+2\}$, egyetlen hiperéle $\{1, \ldots, h+2\}$, hipercseresznyéje pedig nincs. Ez tehát egy legkisebb h-hipercseresznyefa. Rögzítsük n-et és tegyük fel, hogy a tétel állítása n helyére $n-1$-et írva teljesül a már rögzített h-ra. Legyen $\theta = (V, \mathcal{E}, C)$ egy g-hipercseresznyefa, melyre $V = \{1, \ldots, n-h+g\}$, és tegyük fel, hogy a csúcsok egy a θ-át előállító rekurzió lépései szerint vannak számozva. Legyen θ^* az a g-hipercseresznyefa, melyet a θ-ából nyerünk az „$n-h+g$" csúcs és a hozzá tartozó hiperélek és hipercseresznyék elhagyásával. (Mivel a θ csúcsai egy a θ-át előállító rekurzió lépései szerint vannak számozva, ezért az „$n-h+g$" csúcsot a θ-át előállító rekurzió utolsó lépésében vettük a meglévőekhez. Tehát az „$n-h+g$" csúcsot a hozzá tartozó hiperélekkel elhagyva valóban g-hipercseresznyefát kapunk.) Az n-re való indukció miatt a θ^*-ból a tételben leírt módon származtathatunk egy $\triangle^*$ h-hipercseresznyefát, mely csúcsainak halmaza $\{1, \ldots, n-1\}$. A h-ra való indukció miatt a θ-ból a tételben leírt módon származtathatunk egy Γ $(h-1)$-hipercseresznyefát, mely csúcsainak halmaza szintén $\{1, \ldots, n-1\}$. Ha $\triangle^*$ csúcsaihoz hozzáveszük az n csúcsot, hiperéleihez a Γ hiperéleit az n csúccsal kiegészítve, akkor éppen a θ-ából származtatható $\triangle$ h-hipercseresznyefát kapjuk. Valóban, a $\triangle^*$ hiperélei a $\triangle$ azon hiperélei, melyek nem tartalmazzák az n csúcsot, hiszen minden egyes a $\theta \backslash \theta^*$-beli $\{i_1, \ldots, i_{g+1}, n-h+g\}$ hiperélhez csak egyetlen olyan H halmaz van, melyre $H \subset \{n-h+g+1, \ldots, n\}$, $|H| = h-g$, mégpedig a $H = \{n-h+g+1, \ldots, n\}$, ami pedig tartalmazza az „n" csúcsot; valamint ha $\triangle$ az „n" csúcsot tartalmazó hiperéleiből elhagyjuk az „n" csúcsot, akkor éppen a Γ hiperélcit kapjuk meg. Ugyanez mondható el a hipercseresznyékről is. A fentiekből kifolyólag $\triangle$ valóban h-hipercseresznyefa. □

3. Valószínűségi korlátok

Először az alábbi definícióval vezessük be a h-hipercseresznyefa súlyának a fogalmát.

3. Definíció. A $\triangle = (V, \mathcal{E}, C)$ h-hipercseresznyefa súlya:

$$w(\triangle) = \sum_{\{i_1, \ldots, i_{h+2}\} \in \mathcal{E}} P(A_{i_1} \cap \ldots \cap A_{i_{h+2}}) - \sum_{\{i_1, \ldots, i_{h+3}\} \in C} P(A_{i_1} \cap \ldots \cap A_{i_{h+3}})$$

Ennek segítségével tetszőleges $A_1, \ldots, A_n$ események uniójának a valószínűségére az alábbi korlátok adhatók.

3. TÉTEL. *Legyenek $A_1, \ldots, A_n$ tetszőleges események és legyen $\triangle = (V, \mathcal{E}, C)$ egy tetszőleges h-hipercseresznyefa, melyre $V = \{1, \ldots, n\}$. Ekkor teljesülnek a következő egyenlőtlenségek.*

- *Ha h páros:*

$$(1) \qquad P\Big(\bigcup_{i=1}^{n} A_i\Big) \le S_1 - S_2 + \ldots + S_{h+1} - w(\triangle) =$$

$$= \sum_{k=1}^{h+1} (-1)^{k-1} S_k - \sum_{\{i_1,\ldots,i_{h+2}\}\in\mathcal{E}} P(A_{i_1} \cap \ldots \cap A_{i_{h+2}}) +$$

$$+ \sum_{\{i_1,\ldots,i_{h+3}\}\in\mathcal{C}} P(A_{i_1} \cap \ldots \cap A_{i_{h+3}}).$$

- *Ha h páratlan:*

$$(2) \qquad P\Big(\bigcup_{i=1}^{n} A_i\Big) \ge S_1 - S_2 + \ldots - S_{h+1} + w(\triangle) =$$

$$= \sum_{k=1}^{h+1} (-1)^{k-1} S_k + \sum_{\{i_1,\ldots,i_{h+2}\}\in\mathcal{E}} P(A_{i_1} \cap \ldots \cap A_{i_{h+2}}) -$$

$$- \sum_{\{i_1,\ldots,i_{h+3}\}\in\mathcal{C}} P(A_{i_1} \cap \ldots \cap A_{i_{h+3}}).$$

Bizonyítás. A bizonyítást h-ra való indukcióval végezzük oly módon, hogy (1) és (2) egyenlőtlenségeket egyszerre igazoljuk. Ha $h = 0$, akkor a szokásos cseresznyefa korlátot kapjuk vissza az (1) egyenlőtlenségben (lásd [4]). Legyen $h \ge 1$ és tegyük fel, hogy (1) és (2) egyenlőtlenségek teljesülnek, ha h helyett kisebb számot írunk. Feltesszük, hogy h páratlan. Az az eset, amikor h páros, analóg bizonyítható. Igazoljuk, hogy (2) egyenlőtlenség teljesül ezzel a h-val. Egy n-re való indukciót is használunk. Ha $n = h+2$ vagy $n = h+3$, akkor a (2) a jól ismert szita formula lesz, mely egyenlőséggel teljesül. Legyen $n \ge h+4$ és tegyük fel, hogy (2) egyenlőtlenség teljesül, ha n helyett kisebb számot írunk. Az általánosság megszorítása nélkül feltehetjük, hogy $\triangle$ rekurzív előállításának utolsó lépésében az n csúcsot vettük hozzá a $\overline{\triangle} = (\overline{V}, \overline{\mathcal{E}}, \overline{\mathcal{C}})$ h-hipercseresznyefa csúcsaihoz és a $\triangle$ $\mathcal{E}$-beli hiperéleit úgy nyertük, hogy az $\overline{\mathcal{E}}$-beli hiperélekhez vettük hozzá az $\mathcal{E}^*$ elemeit az n csúccsal kibővítve, valamint a $\triangle$ $\mathcal{C}$-beli hipercseresznyéit úgy nyertük, hogy a $\overline{\mathcal{C}}$-beli hiperélekhez vettük hozzá a $\mathcal{C}^*$ elemeit az n csúccsal kibővítve, ahol $\mathcal{E}^*$ és $\mathcal{C}^*$ a Γ $(h-1)$-hipercseresznyefa hiperéleinek illetve hipercseresznyéinek halmaza. Az n-re való indukcióból a (2) egyenlőtlenséget az $A_1, \ldots, A_{n-1}$ eseményekre alkalmazva a $\overline{\triangle} = (\overline{V}, \overline{\mathcal{E}}, \overline{\mathcal{C}})$ alapján következik, hogy

$$(3) \qquad P\Big(\bigcup_{i=1}^{n-1} A_i\Big) \ge \sum_{k=1}^{h+1} (-1)^{k-1} \sum_{\{i_1,\ldots,i_k\}\subset\{1,\ldots,n-1\}} P(A_{i_1} \cap \ldots \cap A_{i_k}) +$$

$$+ \sum_{\{i_1,\ldots,i_{h+2}\}\in\overline{\mathcal{E}}} P(A_{i_1}\cap\ldots\cap A_{i_{h+2}})-$$

$$- \sum_{\{i_1,\ldots,i_{h+3}\}\in\overline{\mathcal{C}}} P(A_{i_1}\cap\ldots\cap A_{i_{h+3}}),$$

valamint a h-ra való indukcióból az (1) egyenlőtlenséget az $A_1\cap A_n$, $A_2\cap A_n,\ldots$, $A_{n-1}\cap A_n$ eseményekre alkalmazva a $\Gamma=(V^*,\mathcal{E}^*,\mathcal{C}^*)$ alapján

$$(4)-P\Big(\bigcup_{i=1}^{n-1}(A_i\cap A_n)\Big)\geq\sum_{k=1}^{h}(-1)^k\sum_{\{i_1,\ldots,i_k\}\subset\{1,\ldots,n-1\}}P(A_{i_1}\cap\ldots\cap A_{i_k}\cap A_n)+$$

$$+\sum_{\{i_1,\ldots,i_{h+1}\}\in\mathcal{E}^*}P(A_{i_1}\cap\ldots\cap A_{i_{h+1}}\cap A_n)-$$

$$-\sum_{\{i_1,\ldots,i_{h+2}\}\in\mathcal{C}^*}P(A_{i_1}\cap\ldots\cap A_{i_{h+2}}\cap A_n)$$

Összeadva a (3) és (4) egyenlőtlenségeket, majd mindkét oldalhoz $P(A_n)$-et adva az alábbi egyenlőtlenséget nyerjük:

$$(5)\qquad P\Big(\bigcup_{i=1}^{n}A_i\Big)=P\Big(\bigcup_{i=1}^{n-1}A_i\Big)+P(A_n)-P\Big(\bigcup_{i=1}^{n-1}(A_i\cap A_n)\Big)\geq$$

$$\geq\sum_{k=1}^{h+1}(-1)^{k-1}S_k+\sum_{\{i_1,\ldots,i_{h+2}\}\in\mathcal{E}}P(A_{i_1}\cap\ldots\cap A_{i_{h+2}})-$$

$$-\sum_{\{i_1,\ldots,i_{h+3}\}\in\mathcal{C}}P(A_{i_1}\cap\ldots\cap A_{i_{h+3}})$$

Ezzel a (2) egyenlőtlenséget igazoltuk. Ha h páros, akkor az (1) egyenlőtlenség teljesen analóg módon igazolható. □

1. Megjegyzés. Látható, hogy mind az (1) mind a (2) korlátokat használva az a célunk, hogy minél nagyobb súlyú h-hipercseresznyefát találjunk. Tegyük fel például, hogy az (1) egyenlőtlenségben egy a [4]-ben leírt módszer alapán nyert cseresznyefát alkalmazunk. Hagyjunk el ebből a cseresznyefából egy kettőfokú csúcsot a hozzátartozó élekkel és cseresznyével és az így kapott cseresznyefát az 2. tételben leírtak alapján bővítsük 1-hipercseresznyefává. Ezt az 1-hipercseresznyefát a (2)-ben alkalmazva alsó korlátot kapunk. Heurisztikus meggondolások alapján az várható, hogy egy nagy súlyú cseresznyefából a fenti bővítéssel nagy súlyú 1-hipercseresznyefát nyerünk. Erre világít rá a következő numerikus példa.

3. Példa. Legyenek A_1, A_2, A_3, A_4, A_5 a következő valószínűségekkel rendelkező események:

$$p_1 = p_2 = p_3 = p_4 = p_5 = 0,38;$$

$$p_{1,2} = 0,21; \quad p_{1,3} = 0,20; \quad p_{1,4} = 0,19; \quad p_{1,5} = 0,18; \quad p_{2,3} = 0,17;$$

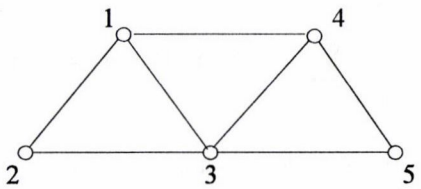
$$p_{2,4} = 0,16; \quad p_{2,5} = 0,17; \quad p_{3,4} = 0,18; \quad p_{3,5} = 0,19; \quad p_{4,5} = 0,20;$$

$$p_{1,2,3} = p_{1,2,4} = p_{1,2,5} = p_{1,3,4} = p_{1,3,5} =$$

$$= p_{1,4,5} = p_{2,3,4} = p_{2,3,5} = p_{2,4,5} = p_{3,4,5} = 0,11;$$

$$p_{1,2,3,4} = 0,07; \quad p_{1,2,3,5} = 0,08; \quad p_{1,2,4,5} = 0,09;$$

$$p_{1,3,4,5} = 0,08; \quad p_{2,3,4,5} = 0,07.$$



2. ábra

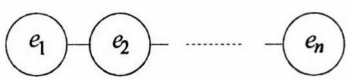
A maximális súlyú cseresznyefa a 2. ábrán látható, súlya 1,01, így az általa szolgáltatott felső korlát $1,9 - 1,01 = 0,89$. (Az éles felső korlát 0,88.) Hagyjuk el az „5" csúcsot a hozzátartozó élekkel és cseresznyével, majd az így kapott cseresznyefát bővítsük ki 1-hipercseresznyefává a 2. tételben leírtak alapján. A kapott 1-hipercseresznyefa a $\big(V = \{1,2,3,4,5\}, \mathcal{E} = \big\{\{1,2,3\},\{1,2,4\},\{1,2,5\},$ $\{1,3,4\},\{1,3,5\},\{1,4,5\},\{2,3,4\},\{2,3,5\},\{3,4,5\}\big\}, \mathcal{C} = \big\{\{1,2,3,4\},\{1,2,3,5\},$ $\{1,3,4,5\}\big\}\big)$, súlya 0,76, így az általa szolgáltatott alsó korlát $1,9 - 1,85 + 0,76 =$ 0,81. Az éles alsó korlát 0,82, mely az imént alkalmazott módszerrel is megkapható a következő módon. Számozzuk át a 2. ábrán látható cseresznyefa pontjait a következő módon: $1 \to 4$, $2 \to 5$, $3 \to 3$, $4 \to 2$, $5 \to 1$. Hagyjuk el az „5" csúcsot a hozzátartozó élekkel és cseresznyével, majd az így kapott cseresznyefát bővítsük ki 1-hipercseresznyefává a 2. tételben leírtak alapján. A kapott 1-hipercseresznyefa $\big(V = \{1,2,3,4,5\}, \quad \mathcal{E} = \big\{\{1,2,3\},\{1,2,4\},\{1,2,5\},\{1,3,4\},\{1,3,5\},\{2,3,4\},$ $\{2,3,5\},\{2,4,5\},\{3,4,5\}\big\}, \mathcal{C} = \big\{\{1,2,3,4\},\{1,2,3,5\},\{2,3,4,5\}\big\}\big)$, mely valóban a 0,82 alsó korlátot szolgáltatja a (2) egyenlőtlenséggel. E példán is látható, hogy más-más kibővítéssel más-más korlátot kapunk.

2. Megjegyzés. Egy cseresznyefa fent leírt módon 1-hipercseresznyefává való bővítéséből nyert alsó korlát kiszámításához nem szükséges az összes $\binom{n}{4}$ eseménynégyes metszetvalószínűségét ismernünk, illetve adott esetben kiértékelnünk, hanem elegendő csak az 1-hipercseresznyefa hipercseresznyéinek megfelelőket, azaz

$\binom{n-2}{2}$-őt. Ugyanez vonatkozik az eseményhármasokra is: elegendő a hiperéleknek megfelelő $2\binom{n-2}{2} + (n-2) = (n-2)^2$-őt ismernünk. Általánosságban pedig egy g-hipercseresznyefa h-hipercseresznyefává való bővítéséből nyert korlát kiszámításához $\mathcal{O}(n^{h+1})$ nagyságrendű legfeljebb $h+3$ esemény metszetéből álló valószínűséget kell ismernünk illetve kiértékelnünk.

4. Numerikus tesztek

Az alábbi példákban a lineáris k-out-of-r-from-n: F rendszer megbízhatóságának becslésére alkalmazzuk módszerünket. Egy k-out-of-r-from-n: F rendszer n egymást követő elemből áll, jelöljük ezeket $e_1, \ldots, e_n$ -nel (3. ábra).



3. ábra

Minden elemnek két lehetséges állapota van: működik vagy nem működik. Az e_i elem működésének a valószínűsége legyen p minden $i = 1, \ldots, n$-re. A rendszer pontosan akkor nem működik, ha valamely r egymást követő eleme közül legalább k nem működik. Ha A_i-vel $(i = 1, \ldots, n-r+1)$ jelöljük azt az eseményt, hogy az $\{e_i, \ldots, e_{i+r-1}\}$ halmazból legalább k elem nem működik, akkor a rendszer működésének a valószínűsége $1 - P(A_1 \cup \ldots \cup A_N)$, ahol $N = n - r + 1$. Az $A_1, \ldots, A_N$ események, azok párjai és hármasai metszeteinek valószínűségeit az M. Sfakianakis, S. Kounias és A. Hillaris [12] által kidolgozott algoritmussal, a négyes metszeteik valószínűségeit pedig az A. Habib és T. Szántai [6] által publikált algoritmussal hatékonyan lehet számítani. Ezért a rendszer működése valószínűségének a becslésére alkalmazni lehetett a 3. szakaszban kidolgozott valószínűségi korlátokat. Néhány tesztfeladatra vonatkozó eredményt a következő táblázatokban foglaltunk össze.

n = 20, r = 13, k = 4, p = 0,80

	alsó korlát	felső korlát
S_1, S_2, S_3	0,4350	0,6033
S_1, S_2, S_3, S_4	0,5193	0,5817
cseresznyefa	**0,5337**	–
1-hipercseresznyefa	–	0,5696
1-hipercseresznyefa*	–	**0,5575**

I. Táblázat

n = 15, r = 8, k = 3, p = 0,75

	alsó korlát	felső korlát
S_1, S_2, S_3	0,2328	0,4639
S_1, S_2, S_3, S_4	0,3421	0,4337
cseresznyefa	**0,3544**	–
1-hipercseresznyefa	–	0,4239
1-hipercseresznyefa*	–	**0,4039**

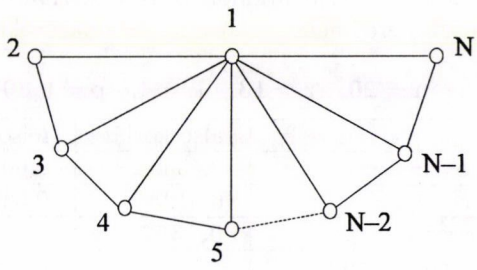
II. Táblázat

n = 30, r = 22, k = 3, p = 0,90

	alsó korlát	felső korlát
S_1, S_2, S_3	0,3254	0,5022
S_1, S_2, S_3, S_4	0,4256	0,4856
cseresznyefa	**0,4551**	–
1-hipercseresznyefa	–	0,4707
1-hipercseresznyefa*	–	**0,4633**

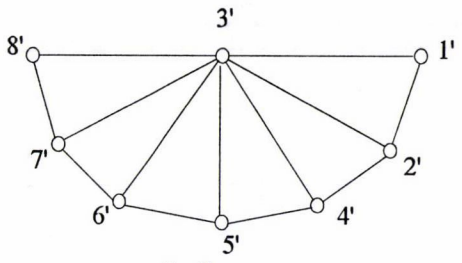
III. Táblázat

A táblázatokban az S_1, S_2, S_3 és S_1, S_2, S_3, S_4 azonosítójú korlátok a felsorolt binomiális momentumok ismeretében adható éles Boole–Bonferroni korlátokat jelentik. A cseresznyefa azonosítójú korlátot a 4. ábrán látható cseresznyefa alapján számítottuk, ahol az I. és II. táblázat példájában $N = 8$, a III. táblázatében pedig $N = 9$. Sejtésünk szerint minden lineáris k-out-of-r-from-n: F rendszer megbízhatósági becslésénél, ha az $e_1, \ldots, e_n$ elemek működésének valószínűsége egyenlő, a 4. ábrán látható szerkezetű cseresznyefa a maximális súlyú.



4. ábra

Tekintsük az I. és II. táblázat példáját. Számozzuk át a 4. ábrán lévő cseresznyefa $N = 8$ csúcsát a $p(1,2,3,4,5,6,7,8) = (3,8,7,6,5,4,2,1)$ permutációnak megfelelően, amivel az 5. ábrán lévő cseresznyefát kapjuk. Ebben a sorrendben véve a csúcsokat a cseresznyefa valóban előállítható. Hagyjuk el a $8'$ csúcsot a hozzá



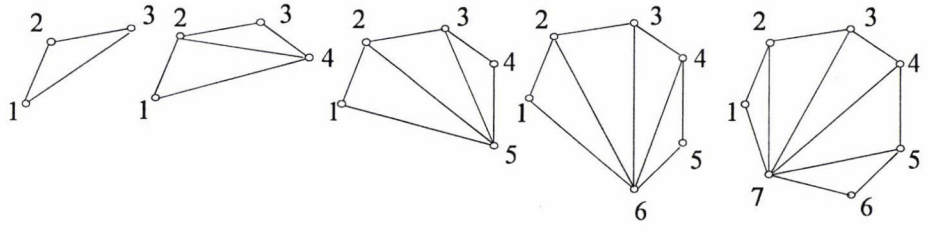
5. ábra

tartozó élekkel és cseresznyével, majd az így kapott cseresznyefából az 2. tételben leírtak szerint szerkesszünk 1-hipercseresznyefát.

A III. táblázat példájához hasonlóképpen szerkeszthetünk 1-hipercseresznyefát a 4. ábrán lévő cseresznyefa $N = 9$ csúcsának a $p(1,2,3,4,5,6,7,8,9) = (3,9,8,7,6,5,4,2,1)$ permutáció szerinti átszámozásával, majd a $9'$ csúcs elhagyásával. Az ily módon nyert 1-hipercseresznyefák alapján számított korlátok a táblázatok 1-hipercseresznyefa azonosítójú soraiban találhatók.

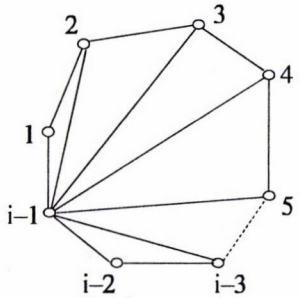
Egy másik módszer „nagy súlyú" 1-hipercseresznyefa keresésére a következő. Tekintsük a csúcsok egy p permutációját. Az 1-hipercseresznyefát — rekurzív definíciójának értelmében — úgy állítjuk elő, hogy a rekurzió i-edik lépésében a $p(i)$ csúcsot veszük hozzá a meglévőkhöz, a hiperéleket és hipercseresznyéket pedig a Γ_i cseresznyefa alapján veszük, melyet mohó algoritmussal konstruálunk a $p(1),\ldots,p(i-1)$ csúcsokon. Az ily módon nyert 1-hipercseresznyefák alapján számított korlátok a táblázatok 1-hipercseresznyefa* azonosítójú sorában találhatók.

Az I. és II. táblázat példáján a permutáció $p(1,2,3,4,5,6,7,8) = (1,2,3,4,5,6,7,8)$, a Γ_i cseresznyefákat a 6. ábrán láthatjuk. (Természetesen Γ_i csak $i = 4,5,6,7,8$-ra létezik, hiszen a legkisebb 1-hipercseresznyefa háromcsúcsú, hiperélei és hipercseresznyéi definíció szerint adottak.)



6. ábra

A III. táblázat példáján a permutáció $p(1,2,3,4,5,6,7,8,9) = (1,2,3,4,5,6,7,8,9)$, a Γ_i cseresznyefákat $i = 4,5,6,7,8$-ra a 6. ábrán láthatjuk, a Γ_9 pedig a 7. ábrán látható i helyére 8-at írva.



7. ábra

Sejtésünk szerint minden lineáris k-out-of-r-from-n: F rendszer megbízhatósági becslésénél, ha az $e_1, \ldots, e_n$ elemek működésének valószínűsége egyenlő, az az 1-hipercseresznyefa a maximális súlyú, melyet az imént említettek szerint állítunk elő az identitás permutációval és azokkal a Γ_i, $i = 4, \ldots, N$ fákkal, melyek általános alakja a 7. ábrán látható.

Ezen módszer alkalmalmazásánál szükségünk van a csúcsok egy megfelelő permutációjára. Ezt a permutációt megkaphatjuk például oly módon, hogy az 1-hipercseresznyefa rekurzív előállítása során mindig azt a csúcsot veszük a már meglévőkhöz, mely hozzávételével a fent említett mohó algoritmusunk szerint a legjobban tudjuk növelni az 1-hipercseresznyefa súlyát. Hasonlóan találhatunk „nagy súlyú" h-hipercseresznyefát úgy, hogy a rekurzió minden lépésében „nagy súlyú" $(h-1)$-hipercseresznyefát konstruálunk.

Temészetesen más módszerrel is találhatunk „nagy súlyú" h-hipercseresznyefát. A kérdés, hogy vajon megtalálhatjuk-e a legjobbat illetve közel a legjobbat elfogadható időn belül, még nyitott.

A IV–VII. táblázatok néhány további lineáris k-out-of-r-from-n: F rendszer megbízhatóságára adott korlátokat tartalmazzák. A példákat A. Habib [5] kandidátusi értekezéséből vettük. A Hunter–Worsley korlátok leírása az [8] és a [14] dolgozatban, a Jun Cai korláté a [9] dolgozatban, az S_1, S_2, S_3 és S_4 értékeken alapuló korlátoké pedig a [10] könyvben található meg.

n = 15, r = 7, k = 5, p = 0,75,
$R(p; k, r, n) = 0{,}943$

módszer	alsó korlát	felső korlát
Hunter Worsley	0,9377	–
Jun Cai	0,9153	0,9559
S_1, S_2	0,9084	0,9592
S_1, S_2, S_3	0,9352	0,9537
S_1, S_2, S_3, S_4	0,9391	0,9466
cseresznyefa	**0,9413**	–
1-hipercseresznyefa*	–	**0,9432**

IV. Táblázat

n = 30, r = 6, k = 3, p = 0,90,
$R(p; k, r, n) = 0{,}849$

módszer	alsó korlát	felső korlát
Hunter Worsley	0,8267	–
Jun Cai	0,8087	0,8725
S_1, S_2	0,6451	0,8880
S_1, S_2, S_3	0,8131	0,8837
S_1, S_2, S_3, S_4	0,8213	0,8637
cseresznyefa	**0,8360**	–
1-hipercseresznyefa*	–	**0,8511**

V. Táblázat

n = 40, r = 7, k = 4, p = 0,90,
$R(p; k, r, n) = 0{,}957$

módszer	alsó korlát	felső korlát
Hunter Worsley	0,9540	–
Jun Cai	0,9464	0,9655
S_1, S_2	0,9129	0,9695
S_1, S_2, S_3	0,9515	0,9687
S_1, S_2, S_3, S_4	0,9522	0,9616
cseresznyefa	**0,9561**	–
1-hipercseresznyefa*	–	**0,9582**

VI. Táblázat

$\mathbf{n = 50,\ r = 40,\ k = 28,\ p = 0{,}50,}$
$R(p; k, r, n) = 0{,}979$

módszer	alsó korlát	felső korlát
Hunter Worsley	0,9739	–
Jun Cai	0,9345	0,9882
S_1, S_2	0,9560	0,9863
S_1, S_2, S_3	0,9697	0,9832
S_1, S_2, S_3, S_4	0,9754	0,9816
cseresznyefa	**0,9766**	–
1-hipercseresznyefa*	–	**0,9794**

VII. Táblázat

IRODALOM

[1] Boole, G., *Laws of Thought* (American reprint of 1854 edition, Dover, New York, 1854).

[2] Bonferroni, C. E., „Teoria Statistica Delle Classi e Calcolo Delle Probabilitá", in: *Volume in onordi Riccardo Dalla Volta* Universitá di Firenze, (1937) 1–62.

[3] Bukszár, J., „Események uniójára adott valószínűségi korlátok", PhD értekezés. Eötvös Loránd Tudományegyetem, 1998 November.

[4] Bukszár, J. és Prékopa, A., Probability bounding with cherry trees, RUTCOR Research Report RRR 04-99, 1999.

[5] Habib, A., „Reliability Problems of Network Systems", PhD Thesis, Hungarian Academy of Sciences, September, 1997.

[6] Habib, A., Szántai, T., Egy speciális hálózati rendszer megbízhatósága, *Alkalmazott Matematikai Lapok* **18** (1998) 39–61.

[7] Hailperin, Th., Best possible inequalities for the probability of a logical function of events, *The American Mathematical Monthly* **72** (1965) 343–359.

[8] Hunter, D., Bounds for the probability of a union, *Journal of Applied Probability* **13** (1976) 597–603.

[9] Jun Cai, Reliability of a large consecutive k-out-of-r-from-n: F system with unequal component-reliability, *IEEE Transactions on Reliability* **R-43** (1994) 107–111.

[10] Prékopa, A., *Stochastic Programming* (Kluwer Academic Publishers, Dordrecht, 1995).

[11] Prékopa, A., Vizvári, B., Regős, G., Lower and upper bounds on probabilities of Boolean functions of events, *RUTCOR Research Report* **RRR 36-95**, December, 1997.

[12] Sfakianakis, M., Kounias, S. and Hillaris, A., Reliability of a consecutive k-out-of-r-from-n: F system, *IEEE Transactions on reliability* **R-41** (1992) 442–447.

[13] Tomescu, I., Hypertrees and Bonferroni inequalities, *Journal of Combinatorial Theory, Series B* **41** (1986) 209–217.

[14] Worsley, K. J., An improved Bonferroni inequality and applications, *Biometrika* **69** (1982) 297–302.

(Beérkezett: 1998. február 19.)

BUKSZÁR JÓZSEF, MISKOLCI EGYETEM
MATEMATIKA INTÉZET, ANALÍZIS TANSZÉK
EGYETEMVÁROS, 3515 MISKOLC
E-mail: matbuk@gold.uni-miskolc.hu

SZÁNTAI TAMÁS, BUDAPESTI MŰSZAKI EGYETEM
MATEMATIKA INTÉZET, DIFFERENCIÁLEGYENLETEK TANSZÉK
MŰEGYETEM RKP. 3, 1111 BUDAPEST
E-mail: szantai@math.bme.hu

PROBABILITY BOUNDS GIVEN BY HYPERCHERRY TREES

JÓZSEF BUKSZÁR AND TAMÁS SZÁNTAI

In the paper new bounds are given for the probability of the union of events. For this purpose new concept of hypercherry trees was introduced. The concept of cherry trees has been introduced earlier by A. Prékopa and J. Bukszár and the concept of m-multitree has been introduced earlier by J. Bukszár. These hypergraph structures were applied succesfully for defining bounds on the union of events, too. All these bounds can be regarded as generalizations of the upper bound introduced by D. Hunter by means of maximum weight spanning tree and so all these bounds were upper bounds. I. Tomescu introduced the concept of hypertrees in the framework of uniform hypergraphs and by means of these new hypergraph structures he was able to define not only upper but also lower bounds on the probability of union of events. The new bounds of the paper are the improvements of Tomescu's lower and upper bounds in the same sense as the upper bounds by A. Prékopa and J. Bukszár were improvements of Hunter's upper bound. The efficiency of the new bounds is illustrated on some test examples according to a special reliability system.

Alkalmazott Matematikai Lapok **19** (1999) 87–111

A KOCKARENDSZERHEZ TARTOZÓ TÉRCSOPORTOK OPTIMÁLIS EGYSZERESEN TRANZITÍV GÖMBKITÖLTÉSEINEK MEGHATÁROZÁSA SZÁMÍTÓGÉPPEL

MÁTÉ CSILLA és SZIRMAI JENŐ*

Budapest

Ajánljuk Szüleinknek

Az általunk ebben a munkában tárgyalt problémát U. Sinogowitz vetette fel 1943-ban a háromdimenziós euklideszi tér ($\mathbf{E}^3$), legsűrűbb gömbkitöltésének keresésével kapcsolatban [11].

Ebben a cikkben a kockarendszerhez tartozó kristálycsoportokhoz tartozó egyszeresen tranzitív gömkitöltéseket vizsgáljuk. Kidolgozunk egy algoritmust és egy programot, amely megkeresi minden egyes említett kristálycsoporthoz a megfelelő optimális gömbkitöltés sűrűségét az optimális magpontok egy reprezentáns elemét továbbá az optimális gömbsugarat. Bebizonyítjuk az algoritmus konvergenciáját és táblázatban összefoglaljuk az eredményeket. A többszörösen tranzitív esetekre is alkalmazható az algoritmus módosított változata, ennek eredményeit később közöljük.

1. Bevezetés

A huszadik század első harmadában a fizika számos új eredményt mutatott fel. Ezek egyike volt az anyagok, ezen belül a kristályok szerkezetének kutatása. Ezen vizsgálatok közben a röntgen diffrakciós technika segítségével feltérképezték a kristályos szerkezetű anyagokat. Már jóval korábban A. Schoenflies, E.S. Fedorov és később, főleg L. Bieberbach munkássága által előtérbe került a kristályok geometriájának vizsgálata is ([1], [4], [5], [10]). Érdekes lehet az egyes kristályok felépítésével kapcsolatban az adott tércsoporthoz tartozó optimális, azaz maximális sűrűségű gömbkitöltések meghatározása.

Ezt a problémát U. Sinogowitz az 1940-es években írt [11] cikkében kezdeményezte a legsűrűbb rácsszerű gömbkitöltés analógiájára és a kétdimenziós térben, vagyis az euklideszi síkban, meg is oldotta az analóg feladatot (lásd még [6]). Térbeli eredményeinek zömét folyóiratban, könyvben (ismereteink szerint) nem publikálta.

* Készült az OTKA T 20498 (1996–99) támogatásával.

Mint ismeretes, a legsűrűbb rácsszerű gömbkitöltés az $\mathbf{Fm\bar{3}m}$ jelű tércsoporthoz tartozik, amely a lapcentrált kockarácsnak a teljes szimmetriacsoportja. Ugyanilyen $\frac{\pi}{\sqrt{18}} \approx 0.74048$ sűrűséggel kaphatunk még szabályos és nem szabályos gömbrendszereket is.

G. Horváth Á. és Molnár E. [7] cikkükben igazolták: a tíz fixpontmentes tércsoport mindegyikére az optimális gömbkitöltés sűrűsége szintén $\frac{\pi}{\sqrt{18}} \approx 0.74048$ és leírták az adott tércsoporthoz tartozó optimális gömbkitöltéseket.

Ez a cikk a kockarendszerekhez tartozó kristálycsoportok vizsgálatával foglalkozik. Ezeken belül az adott tércsoporthoz tartozó egyszeresen tranzitív gömbkitöltéseket vizsgál, olyan gömbrendszereket, amelyeknek bármely két gömbjét az előbbi csoportnak pontosan egy eleme viszi egymásba. Későbbi dolgozatban tárgyaljuk majd azokat a gömbrendszereket, ahol a kitöltés többszörösen tranzitív.

Ennek a problémának a megoldása még viszonylag egyszerű tércsoportok esetén is, tisztán elméleti eszközökkel, igen bonyolult feladat ([7], [12], [13]). Ez indokolja a feladat algoritmikus megközelítését ([3], [8]).

A dolgozatban megadjuk a háromdimenziós kockarendszerhez tartozó adott tércsoportok esetén azt az algoritmust, amely megkeresi az összes említett optimális kitöltést és amelynek segítségével a tércsoport optimális gömbkitöltésének sűrűségére elvileg tetszőleges pontos eredményt kaphatunk. Továbbá megadjuk az algoritmus alapján elkészített program segítségével a háromdimenziós kockarendszerhez tartozó összes tércsoportra az optimális sűrűséget legalább kettő, a maximális gömbsugarat legalább három tizedesjegy pontosságig. A program megadja még az (egyik) optimális gömb középppontját, amely egyben az optimális kitöltéshez tartozó Dirichlet–Voronoi cella magpontja is. A megközelítés pontosságát csak a program futási ideje befolyásolhatja (erre az algoritmus tárgyalásánál még visszatérünk). A [12], [13] munkákból ismerjük az **F23**, **I432**, **F432**, $\mathbf{P\bar{4}3m}$, $\mathbf{Fd\bar{3}m}$, $\mathbf{Pn\bar{3}m}$, $\mathbf{I\bar{4}3m}$ tércsoportok optimális gömbkitöltésének a sűrűségét ill. az optimális Dirichlet–Voronoi cella magpontját (középpontját). Ezekre a tércsoportokra a kifejlesztett programot lefuttatva az egzakt eredménnyel megegyező eredményt kapunk az adott pontosságon belül [8]. Továbbá meghatároztuk még a síktükrözések által generált $\mathbf{Pm\bar{3}m}$, $\mathbf{F\bar{4}3m}$, $\mathbf{Fm\bar{3}m}$ csoportokhoz, valamint a síktükrözéseket tartalmazó $\mathbf{Im\bar{3}m}$ csoportokhoz az optimális gömbkitöltés sűrűségének egzakt értékét is (2. Táblázat). (A sűrűség „könnyen"meghatározható még más síktükrözéseket tartalmazó tércsoportokra is, de ezt egy következő dolgozatban tárgyaljuk más finomításokkal együtt).

2. Alapfogalmak

Röviden összefoglaljuk a kristálycsoportokra vonatkozó ismereteket, jelöléseinket. Az $\mathbf{E}^3$ euklideszi tér egybevágóságainak csoportját jelölje $\mathrm{Iso}\,\mathbf{E}^3$.

2.1. Definíció. A $\mathbf{G}$ transzformáció-csoportot az $\mathbf{E}^3$ tér diszkrét csoportjának nevezzük, ha teljesülnek az alábbi feltételek:

a. $\mathbf{G} \subset \mathrm{Iso}\,\mathbf{E}^3$
b. Tetszőleges $X \in \mathbf{E}^3$ esetén az X pont pályája (orbitja):

$$X^{\mathbf{G}} := \{X^{\alpha} \in \mathbf{E}^3 : \alpha \in \mathbf{G}\}$$

diszkrét ponthalmaz az $\mathbf{E}^3$ térben (nincs torlódási pontja).

(A transzformációk kitevőbe írása balról jobbra arra is utal, hogy ez lesz a végrehajtás sorrendje, ez a megállapodás az irodalomban nem egységes.)

2.2. Definíció. Az $F_{\mathbf{G}}$ zárt ponthalmazt a $\mathbf{G}$ diszkrét csoport alaptartományának (fundamentális tartományának) nevezzük ha teljesülnek a következő feltételek:
a. Minden $P \in \mathbf{E}^3$ esetén létezik olyan $A \in F_{\mathbf{G}}$ pont, hogy $P \in A^{\mathbf{G}}$.
b. Tetszőleges $A, B \in \mathrm{Int}\, F_{\mathbf{G}}$ belső pontokra igaz, hogy ha $B \in A^{\mathbf{G}}$, akkor $A = B$.
c. $\mathrm{Int}\, F_{\mathbf{G}}$ egyszeresen összefüggő $\mathbf{E}^3$-ban.

2.3. Definíció. A $\mathbf{G}$ diszkrét csoportot kristálycsoportnak nevezzük, ha létezik korlátos alaptartománya.

2.1. TÉTEL (Schoenflies–Bieberbach). *Ha a $\mathbf{G}$ diszkrét csoport kristálycsoport, akkor tartalmaz 3 lineárisan független párhuzamos eltolást.*

Megjegyzések.
1. A tétel általánosan n dimenzióban is igaz ($1 \leq n \in N$)
2. A tétel megfogalmazható úgy is, hogy a csoport egybevágóságainak lineáris része által alkotott csoport, az un. pontcsoport véges ortogonális csoport.

Ekkor a G tércsoport $\Gamma(\mathbf{G})$ eltoláspontja 3 független eltolás által generálható, és $\Gamma(\mathbf{G})$ egy O^{Γ}-val jelölt háromdimenziós pontráccsal szemléltethető (ahol O a tér tetszőleges pontja).

$\mathbf{S}(\Gamma)$-val jelöljük a $\Gamma(\mathbf{G})$ rács teljes szimmetriacsoportját és $\mathbf{S_0}(\Gamma)$-val az $\mathbf{S}(\Gamma)$ pontcsoportját. Nyilvánvaló, hogy $\mathbf{S_0}(\Gamma)$ az $\mathbf{S}(\Gamma)$ egy véges részcsoportja.

2.4. Definició. A $\mathbf{G}$ tércsoport affin ekvivalens a $\mathbf{G}'$ tércsoporttal ha létezik egy β affin transzformáció a térben, amelyre $\beta^{-1} \circ \mathbf{G} \circ \beta = \mathbf{G}'$.

Tehát bármely X pont $\mathbf{G}$-pályájára alkalmazva a β affinitást éppen az X^{β} pont $\mathbf{G}'$ pályáját kapjuk: $(X^{\mathbf{G}})^{\beta} = (X^{\beta})^{\mathbf{G}'}$.

2.5. Definíció. A Γ rácsot a Γ' ráccsal ekvivalensnek mondjuk ha létezik a térben egy β affin transzformáció amelyre $\beta(\Gamma) = \Gamma'$ és $\beta^{-1} \circ \mathbf{S}(\Gamma) \circ \beta = \mathbf{S}(\Gamma')$.

2.2. TÉTEL. *A $\mathbf{G_0}$ pontcsoportnak minden egybevágósága a $\Gamma(\mathbf{G})$ rácsnak egy szimmetrialeképezése és ezért $\mathbf{G_0}$ egy részcsoportja az $\mathbf{S_0}(\Gamma)$ pontcsoportnak.*

2.3. TÉTEL (Kristálytani korlátozás). *Legyen $\mathbf{G}$ kristálycsoport és $\mathbf{S_0}(\Gamma)$ a $\mathbf{G}$ rácsának pontcsoportja. Az $\mathbf{S_0}(\Gamma)$ pontcsoport forgatásai csak 2, 3, 4, vagy 6-od rendűek lehetnek.*

Az $\mathbf{S_0}(\Gamma)$ az előzőekből következően véges csoport, továbbá érvényes rá a kristálytani korlátozás. A $\mathbf{G_0}$ pontcsoportokra ezek után csak véges sok lehetőség adódik. Így kapjuk a következő tételt.

2.4. TÉTEL (Geometriai kristályosztályok). *A* $\mathbf{G}$ *tércsoportok* $\mathbf{G_0}$ *pontcsoportjai 32 (metrikus) ekvivalencia-osztályt képeznek.*

2.5. TÉTEL (Bravais rácsok). *A háromdimenziós rácsok 14 affin ekvivalencia osztályt képeznek.*

Megjegyzés. A Bravais rácsokból kiolvasható, hogy a tércsoportok rendelkezhetnek szabad (affin) rácsparaméterekkel, de a kockarács esetén csak egy hasonlósági paraméter létezik.

Mindezek után a rácsok és a pontcsoportok ekvivalencia-osztályai segítségével felépíthetjük a lehetséges kristálycsoportokat három dimenzióban. A továbbiakat már nem részletezve kapjuk a következő tételt:

2.6. TÉTEL (Schoenflies, Fedorov, Bieberbach). *A háromdimenziós euklideszi térben a kristálycsoportoknak pontosan 219 affin ekvivalencia osztálya létezik A tércsoportok esetében az izomorfiából következik az affin ekvivalencia, tehát az izomorfiaosztályok száma is 219.*

Megjegyzés. Ha az affin ekvivalenciánál csak az orientációtartó affin transzformációkat vesszük figyelembe, akkor 230 osztály létezik a 2.4 definíció szellemében.

2.6. Definíció. Egy $A \in \mathbf{E}^3$ pontnak a $\mathbf{G}$ diszkrét csoporthoz tartozó $\mathbf{G}_A$ stabilizátorcsoportja az A-t helybenhagyó $\mathbf{G}$-beli transzformációból áll

$$\mathbf{G}_A := \{\alpha \in \mathbf{G} : A^\alpha = A\}.$$

A továbbiakban, rögzített $\mathbf{G}$ diszkrét csoport esetén, kizárólag olyan $P \in \mathbf{E}^3$ jellegzetes (karakterisztikus) pontokkal foglalkozunk, amelyeknek a stabilizátorcsoportja az identitásból áll, azaz $\mathbf{G}_P = 1$. Szemléletesen fogalmazva a P pont „szabadsági foka" három, vagyis P egy (elég kicsiny) gömbkörnyezetében ilyen pontok vannak. (Ha $\mathbf{G}_P \neq 1$, akkor a P pont „szabadsági foka" értelemszerűen csökken.)

Legyen $\mathbf{G}$ kristálycsoport, $X, Y \in \mathbf{E}^3$ és $\varrho(X,Y)$ az $\mathbf{E}^3$-beli távolságfüggvény.

2.7. Definíció. Az $X^{\mathbf{G}}$ pályához tartozó, X magpontú zárt Dirichlet–Voronoi cella, röviden D-V cella:

$$D(X^{\mathbf{G}}) := \{Y \in \mathbf{E}^3 : \varrho(X,Y) \leq \varrho(X^g,Y),\ \forall g \in \mathbf{G}\}.$$

Ez a definíció alapvető fontosságú, ezért fejtsük ki részletesebben:

Egy kristálycsoporthoz általában többféle alaptartományt is megadhatunk. Egy ilyen nevezetes poliéder alaptartomány megadását teszi lehetővé a **Dirichlet–Voronoi** cella: Tekintsünk egy jellegzetes P pontot az $\mathbf{E}^n$ n-dimenziós euklideszi térből, tehát amelyre igaz az, hogy a $\mathbf{G}$ kristálycsoportnak a P pontot saját magára képező $\mathbf{G}_P$ részcsoportja csak az egységelemből áll. Képezzük a P pont pályáját alkotó pontok $P^{\mathbf{G}}$ halmazát. Vegyük azon Y pontok D_P halmazát az euklideszi térből, melyek P-hez közelebb (nem távolabb) vannak, mint a pálya többi pontjához. Az így értelmezett D_P halmazt a $P^{\mathbf{G}}$ pontrendszer P magponthoz tartozó Dirichlet–Voronoi cellájának nevezzük.

Az így definiált cellákra igaz a következő két állítás:

a. Tetszőleges $A, B \in \operatorname{Int} D(P^{\mathbf{G}})$ pontokra a $B \in A^{\mathbf{G}}$ feltételből következik, hogy $A = B$. Ami azt jelenti, hogy a cellák nem nyúlnak egymásba.

b. Az euklideszi tér tetszőleges Z pontjához létezik olyan X pont, amely eleme a zárt $D(P^{\mathbf{G}})$ cellának és $Z \in X^{\mathbf{G}}$. Tehát a $D(P^{\mathbf{G}})$ cella $\mathbf{G}$-nél származó képei lefedik a teret.

Ez a két állítás azt jelenti, hogy a $D(P^{\mathbf{G}})$ Dirichlet–Voronoi cellák alaptartományok, de csak akkor, ha a $\mathbf{G}_P$ stabilizátor triviális.

2.8. Definíció. $\mathbf{G}_X = 1$ esetén $D(X^{\mathbf{G}})$ cellába írható X középpontú maximális gömb sugara:

$$r(X) := \min_{g \in \mathbf{G}-1} \left\{ \frac{1}{2} \varrho(X, X^g) \right\}.$$

2.9. Definíció. Az $X^{\mathbf{G}}$ orbithoz tartozó gömbkitöltés sűrűsége:

$$\delta(X^{\mathbf{G}}) := \frac{4r^3(x)\pi}{3 \operatorname{Vol}\left(D(X^{\mathbf{G}})\right)}.$$

Legyen $Z, Y \in X^{\mathbf{G}}$ és $h \in \mathbf{G}$-re teljesüljön, hogy $Y^h = Z$. Ekkor $\left(D(Y^{\mathbf{G}})\right)^h = D(Z^{\mathbf{G}})$. Ez nyilván teljesül az egyes cellákhoz tartozó maximális sugarú gömbökre is. Az X pontrendszer és a kialakuló gömbrendszer $\operatorname{Sym}(X^{\mathbf{G}})$ szimmetriacsoportja mindenképpen tartalmazza a $\mathbf{G}$ szimmetriacsoportot, de lehet bővebb is nála:

$$\mathbf{G} \leq \operatorname{Sym}(X^{\mathbf{G}}).$$

2.10. Definíció. Ha $\operatorname{Sym}(X^{\mathbf{G}}) = \mathbf{G}$ akkor az $X^{\mathbf{G}}$ pályát karakterisztikusnak mondjuk. Egyébként a pálya (orbit) nem karakterisztikus.

3. A probléma megfogalmazása

Adott $\mathbf{G}$ csoport esetén keressük azt az $X^{\mathbf{G}}$ orbitot, melyre $\mathbf{G}_X = 1$, továbbá az orbithoz tartozó gömbkitöltés sűrűsége a maximális.

A $\mathbf{G}$ csoporthoz tartozó optimális sűrűség:

$$\delta(\mathbf{G}) := \max_{X, p(\mathbf{G})} \left\{ \delta(X^{\mathbf{G}}) \right\}.$$

(Ahol $p(\mathbf{G})$ a tércsoport esetleg fellépő szabad (affin) paramétereit jelöli.)

3.1. Definíció. Azok a $h \in \operatorname{Iso} \mathbf{E}^3$ egybevágóságok, amelyekre minden $X \in \mathbf{E}^3$ esetén $(X^{\mathbf{G}})^h = (X^h)^{\mathbf{G}}$, a $\mathbf{G}$ kristálycsoport metrikus normalizátor csoportját alkotják:

$$N(\mathbf{G}) := \{h \in \operatorname{Iso}.E^3 \ : \ h^{-1}\mathbf{G}\, h = \mathbf{G}, \}.$$

Legyen az $N(G)$ metrikus normalizátor csoport alaptartománya $F(N(\mathbf{G}))$.

Megjegyzések. Amikor az optimális sűrűségű gömbkitöltéshez tartozó orbitot keressük, az orbit egy elemét elég az $F(N(\mathbf{G})) \subseteq F_{\mathbf{G}}$ alaptartományból keresni:

$$\delta(\mathbf{G}) := \max_{\substack{X \in F(N(\mathbf{G})) \\ p(\mathbf{G})}} \{\delta(X^{\mathbf{G}})\}.$$

A metrikus normalizátor csoport és alaptartománya is függhet a $\mathbf{G}$ kristálycsoport szabad affin paramétereitől.

4. Az algoritmus

A dolgozatban a kockarendszerhez tartozó tércsoportokkal foglalkozunk. Azért választottuk éppen ezeket, mert, mint a rendszerhez tartozó Bravais rácsok mutatják, ez az osztály — a nyilvánvaló hasonlóságtól eltekintve — nem rendelkezik szabad paraméterekkel. A szabad (affin) paraméterek igen megnehezítik a feladat algoritmikus megközelítését.

A kockarendszerhez tarozó tércsoportokat a [14] atlasz alapján adjuk meg (lásd még [15]).

Tekintsük tehát egy kockarendszerhez tartozó tetszőleges $\mathbf{G}$ tércsoportot, ahol a kocka éleinek a hosszát válasszuk egységnyinek. Válasszunk ki egy kockát a rendszerből és helyezzük el azt egy koordinátarendszerben az 1. ábrán látható módon. A $\mathbf{G}$ tércsoportnak legalább egy alaptartományát (és így a metrikus normalizátor csoportjának megfelelő alaptartományát is) tartalmazza egy 0.5 egység oldalélű az 1. ábrán szemléltetett kocka [9].

Jelöljük ezt a kockát W-vel.

Tehát az alaptartomány definíciója miatt elég az optimális gömb középpontját W-ben keresni. A 3. fejezet 2. megjegyzéséből következően az adott $\mathbf{G}$ tércsoport metrikus normalizátor csoportjának alaptartományától függően W-t érdemes további kisebb tartományokra felosztani.

Tekintsük a W kockát, amelyet a tartalmazó egységkockával együtt elhelyeztünk egy koordinátarendszerben (lásd 1. ábra). A W kockára húzzunk egy pontrácsot, amelynek rácspontjai a következők:

$$P_t = (ei, ej, ek) \quad \text{ahol} \quad i, j, k \in \{1, 2, \ldots n\} \quad \text{és,} \quad 1 \ll n \in \mathbf{N}$$

$$\text{továbbá legyen} \quad e = \frac{1}{2n}.$$

(t az i, j, k értékeitől függő paraméterhármas.) Így kaptunk egy a W kockára húzott finom véges pontrácsot. Jelöljük ezen rács pontjainak a halmazát Q-val.

Válasszunk ki a Q pontrácsból egy tetszőleges P_t pontot, és a $\mathbf{G}$ tércsoportra vonatkozóan számítsuk ki a P_t ponttal ekvivalens pontpozíciók koordinátáit a [14] atlasz alapján.

Ha egy P_t-vel $\mathbf{G}$-ekvivalens pont nincs a W kockát tartalmazó egységkockában, akkor a $\mathbf{G}$ csoporthoz tartozó egységeltolásokkal — amelyek a kocka éleivel párhuzamosak — visszatoljuk őt az egységkockába.

Tehát a P_t pontnak a [14] tabella alapján számított, a $\mathbf{G}$ tércsoportra vonatkozó, ekvivalens pontjait visszajuttattuk az egységkockába, és kiszámítottuk ezek koordinátáit. Jelöljük ezt a ponthalmazt H_t-vel. Toljuk el a $\mathbf{G}$ tércsoporthoz tartozó, a koordináta tengelyekkel párhuzamos, egységeltolások segítségével a H_t ponthalmazt az egységkocka körül elhelyezkedő 26 kocka mindegyikébe. Az így előálló ponthalmazt a H_t halmazzal együtt nevezzük E_{P_t}-nek.

Az ilyen módon megkapott ponthalmaz vizsgálata már biztosan elegendő a $\mathbf{G}$ tércsoport optimális gömbkitöltésének meghatározásához, mert ha ezen 27 kockán kívüli P_t ponttal ekvivalens pontot választunk, annak távolsága P_t-től már biztosan nagyobb mint 1, azonban az optimális sugár biztosan kisebb mint 0.5. Így ezek a pontok már biztosan nem befolyásolják az optimális sugár nagyságát.

Számítsuk ki minden $X \in E_{P_t}$ és $X \neq P_t$ esetén az X és P_t pontok távolságát, majd ezen távolságok közül válasszuk ki a minimálisat. Így megkapjuk a 6. definíció alapján a P_t ponthoz hozzárendelt maximális gömbsugarat.

$$r(P_t) = \min_{X \in E_{P_t}, X \neq P_t} \left\{ \frac{1}{2} \varrho(P_t, X) \right\}. \tag{4.1}$$

Ezekután kiszámolható bármely rögzített $P_t \in Q$ esetén a minimális $r(P_t)$ gömbsugár. Válasszuk ki ezen minimális gömbsugarak közül a P_t-re vonatkozó maximálisat, és jelöljük ezt a $\mathbf{G}$ tércsoporttól és az $n \gg 1$ természetes számtól függő értéket $R_n^{\mathbf{G}}$-vel.

$$R_n^{\mathbf{G}} := \max_{P_t \in Q} \{r(P_t)\} = \max_{P_t \in Q} \left\{ \min_{X \in E_{P_t}, X \neq P_t} \left\{ \frac{1}{2} \varrho(P_t, X) \right\} \right\}. \tag{4.2}$$

Így ha a $\mathbf{G}$ tércsoporthoz tartozó optimális gömb sugarát $R_{\text{opt}}^{\mathbf{G}}$-vel jelöljük, akkor nyilvánvalóan teljesül a következő egyenlőtlenség:

$$R_n^{\mathbf{G}} \leq R_{\text{opt}}^{\mathbf{G}}.$$

Ezzel előállítottuk a $\mathbf{G}$ tércsoport optimális gömbkitöltéséhez tartozó gömbsugárnak, és így egyben a $\mathbf{G}$ tércsoport optimális gömbkitöltéséhez tartozó sűrűségnek is egy alsó korlátját.

Továbbiakban elkészítünk egy felső korlátot is, és megmutatjuk, hogy n növekedésével az alsó és felső korlát különbsége csökken, tehát $R_n^{\mathbf{G}} \to R_{\text{opt}}^{\mathbf{G}}$.

Megjegyzések. Az algoritmus műveletigénye önmagában elég nagy, azonban a következő néhány megjegyzés segítségével az E_{P_t} ponthalmaz elemszáma lecsökkenthető. A megjegyzések jelentősége a konkrét program futtatásánál látható, mert segítségükkel a futási idő lecsökken.

1. Elegendő az E_{P_t} ponthalmazból azokat a pontokat vizsgálni amelyeknek a P_t ponttól való távolsága kisebb mint egy, mert az egységnyi nagyságú a koordinátatengelyekkel párhuzamos eltolások biztos hozzátartoznak a $\mathbf{G}$ csoporthoz.

2. Abban az esetben, ha létezik egy $\gamma \in \mathbf{G}$, $\gamma \neq \mathrm{id}.$, hogy $\varrho(P_t, P_t^\gamma) < 0.5$ akkor nem kell vizsgálnunk a további egységeltolásokkal kapható ${}^1P_t^\gamma$ pontokat, mert a háromszögegyenlőtlenség miatt ezek távolsága P_t-től már 0.5-nél biztosan nagyobb (2. ábra).

3. Ha létezik egy $\gamma \in \mathbf{G}$ a P_q és P_p alappontok esetén, úgy hogy

$$\varrho(P_p, P_p^\gamma) < r(P_q), \quad \text{ahol} \quad p \neq q \quad \text{és} \quad P_p, P_q \in Q, \quad \gamma \in \mathbf{G}, \quad \gamma \neq \mathrm{id}.,$$

ahol a P_p pontot és a vele $\mathbf{G}$-ekvivalens pontokat nem kell figyelembe venni a további számításoknál.

5. A konvergencia bizonyítása, becslések

Jelöljük a W kockában elhelyezkedő (egyik) optimális gömb középpontját és annak koordinátáit a következő módon:

$$K^{\mathbf{G}}_{\mathrm{opt}}(x_{\mathrm{opt}}, y_{\mathrm{opt}}, z_{\mathrm{opt}}).$$

Válasszunk először egy $P(x, y_{\mathrm{opt}}, z_{\mathrm{opt}})$ pontot a W kockából és jelöljük röviden $P(x)$-szel. Vezessünk be a 2.8. definíció felhasználásával egy $M(x)$ segédfüggvényt az $x \in \left[0, \frac{1}{2}\right]$ zárt szakaszon.

Legyen $P^{\min}(x)(A, B, C)$ a $P(x)$-hez legközelebbi pontja az $E_{P(x)} \setminus \{P(x)\}$ ponthalmaznak. Továbbá legyen a $r\big(P(x)\big) = \frac{1}{2}\big(P(x), P^{\min}(x)\big)$.

Ezután legyen

$$M(x) = 4\Big(r\big(P(x)\big)\Big)^2 = (A - x)^2 + (B - y_{\mathrm{opt}})^2 + (C - z_{\mathrm{opt}})^2. \tag{5.1}$$

Az $M(x)$ függvény $x \in \left[0, \frac{1}{2}\right]$ esetén folytonos. A grafikonja pedig parabolaívekből esetleg vízszintes szakaszokból áll (3. ábra). Ez abból adódik, hogy a $P^{\min}(x)$ pont A, B, C koordinátái a [14], [15] táblázatok alapján az x, y_{opt}, z_{opt} és (véges sok) konstans összegéből illetve különbségéből állíthatók elő, ráadásul véges sokféle módon. Ezt szemlélteti az 1. táblázat, amely a [14] és [15] alapján készült és a 227-es számú, a kockarendszerhez tartozó $\mathbf{Fd\bar{3}m}$ tércsoport elemeit tartalmazza. Látható, hogy a tércsoport elemei véges sok típusba tartoznak (4. ábra 1. táblázat).

1. Táblázat

1	$(x,y,z) \to$	$(x,y,z)+\Gamma$
m	$(x,y,z) \to$	$(-x,y,z)+\left(\frac{1}{4},\frac{1}{4},\frac{1}{4}\right)+\Gamma$
.m	$(x,y,z) \to$	$(y,x,z)+\Gamma$
$\overline{\mathbf{1}}$	$x,y,z) \to$	$(-x,-y,-z)+\left(\frac{1}{4},\frac{1}{4},\frac{1}{4}\right)+\Gamma$
2	$(x,y,z) \to$	$(-x,-y,z)+\Gamma$
.2	$(x,y,z) \to$	$(y,x,-z)+\left(\frac{1}{4},\frac{1}{4},\frac{1}{4}\right)+\Gamma$
3	$(x,y,z) \to$	$(z,x,y)+\Gamma$
$\overline{\mathbf{3}}$	$(x,y,z) \to$	$(-z,-x,-y)+\left(\frac{1}{4},\frac{1}{4},\frac{1}{4}\right)+\Gamma$
4	$(x,y,z) \to$	$(-y,x,z)+\left(\frac{1}{4},\frac{1}{4},\frac{1}{4}\right)+\Gamma$
$\overline{\mathbf{4}}$	$(x,y,z) \to$	$(y,-x,-z)+\Gamma$

Az egységeltoláson kívül az **F** jelű csoportoknál 4 lapcentráló eltolás $\left((0,0,0),\left(\frac{1}{2},\frac{1}{2},0\right),\left(\frac{1}{2},0,\frac{1}{2}\right),\left(0,\frac{1}{2},\frac{1}{2}\right)\right)$, az **I** jelű csoportoknál 2 $\left((0,0,0),\left(\frac{1}{2},\frac{1}{2},\frac{1}{2}\right)\right)$, a **P** jelű primitív rácsoknál — formálisan — 1 eltolás $\left((0,0,0)\right)$ adódik még hozzá.

Vizsgájuk meg az $M(x)$ grafikonjában előforduló lehetséges parabolákat.

A táblázatokból látható, hogy az x változó A, B, C közül pontosan egyben fordulhat elő. Ezzel a vizsgálat két részre bontható:

1. Ha A tartalmazza az x változót.

2. Ha B vagy C tartalmazza az x változót (B és C szerepe ebben az esetben felcserélhető).

1. Ekkor a következő lehetőségek vannak:

a. Az A koordinátájában az x együtthatója 1. Ekkor nyilván $M(x)$ =konstans.

b. Az x együtthatója -1. Ebben az esetben

$$M(x) = (k-2x)^2 + \text{nem negatív konstans.}$$

A k lehetséges értékei:

$$k \in \left\{ -\frac{1}{4} \quad \text{ha} \quad x \in \left[0,\frac{1}{4}\right], 0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1 \quad \text{ill.} \quad \frac{5}{4} \quad \text{ha} \quad x \in \left[\frac{1}{4},\frac{1}{2}\right] \right\}.$$

Ez abból adódik, hogy a kockarendszerekhez tartozó tércsoportok esetén a [14] táblázatban $\left\{0,\frac{1}{4},\frac{1}{2},\frac{3}{4},1,\frac{5}{4}\right\}$ eltolási koordináták fordulhatnak elő, és emellett még figyelembe vettük a 4. fejezet megjegyzéseit. Így látható, hogy az 1/b esetben az $M(x)$ milyen típusú parabolákat tartalmazhat:

$$M(x) = 4x^2 - 4kx + \text{konstans},$$

ahol k az előbb felsoroltak közül való. Az 5. ábrán vázoltuk ezen lehetséges parabolaíveket.

2. Ebben az esetben:

$$M(x) = (A - x)^2 + (B - y_{\text{opt}})^2 + (C - z_{\text{opt}})^2 =$$

$$= (k_1 \pm z_{\text{opt}} - x)^2 + (k_2 \pm x - y_{\text{opt}})^2 + \text{konst.} =$$

$$= 2x^2 - 2x(k_1 \pm z \mp k_2 \pm y) + \text{konstans} \quad \text{(ahol ez a konstans nemnegatív)}.$$

A $P^{\min}(x)(A, B, C)$ pont meghatározásából következően ebben az esetben is $M(x) < 1$. Ebből következően:

$$M(x) = 2x^2 - 2x(k_1 \pm z_{\text{opt}} \mp k_2 \pm y_{\text{opt}}) + \text{konstans} < 1, \quad \text{ezért}$$

$$2x^2 - 2x(k_1 \pm z_{\text{opt}} \mp k_2 \pm y_{\text{opt}}) < 1 \quad \text{is teljesül minden} \quad x \in \left[0, \frac{1}{2}\right] \quad \text{esetén.}$$

Jelöljük a $2x$ tag együtthatóját E-vel.

Az előző feltételből következően E-re a következő egyenlőtlenség teljesül:

$$-\frac{1}{2} < E < 1.$$

Tehát a 2. esetben:

$$M(x) = 2x^2 - 2xE + \text{konstans}, \quad \text{ahol} \quad -\frac{1}{2} < E < 1.$$

A 2. esetben előforduló parabolaívek két szélső helyzetét ábrázolja a 6. ábra.

Könnyen látható ezek után, hogy a legnagyobb meredekséget, amit a lehetséges parabolaívek elérhetnek több parabola is létrehozza, például az

$$M(x) = 4x^2 - x + \text{konstans}$$

parabola az $x \in \left[0, \frac{1}{2}\right]$ intervallumon. Ennek a maximális meredekségnek a nagysága a derivált becslésével: $|m| = 3$.

Továbbra is vizsgáljuk a $P(x, y_{\text{opt}}, z_{\text{opt}})$ pontokat, amelyeknél $x \in \left[0, \frac{1}{2}\right]$. Tegyük fel, hogy ki tudjuk számolni az $M(x)$ értékeit minden $(ei, y_{\text{opt}}, z_{\text{opt}})$, $e = \frac{1}{2n}$, $i \in \{1, 2, \ldots, n\}$ pontban, ahol $n \gg 1$ egész szám. A 7. ábrán kis téglalapokkal érzékeltettük az $M(x)$ számolt értékeit. Ismertek tehát $M(x)$ értékei diszkrét pontokban, továbbá tudjuk, hogy $M(x)$ grafikonja parabolaívekből és esetleg vízszintes szakaszokból áll. Ismert az is, hogy mekkora az a maximális emelkedés, amit az $M(x)$ grafikonját alkotó parabolák létrehoznak. A maximális emelkedést a 8. ábrán látható módon érhetjük el. Látható, hogy a maximális emelkedés nagysága,

egy adott $\mathbf{G}$ tércsoport esetén csak a maximális meredekségtől és az n-től függ. Jelöljük a maximális emelkedést $s(n)$-nel. (Ez szélsőséges helyzet a valóságban nem áll elő, de most célunk egy felső becslés megadása.)

Az $s(n)$ értékére felső becslés a következő:

$$s(n) < \frac{3}{4n}. \tag{5.2}$$

Megjegyzés. Az $s(n)$ (5.2) felső becslése az adott G tércsoport normalizátorcsoportjának figyelembevételével lényegesen javítható, hiszen lecsökken a kiindulási kocka éle, továbbá a létrejövő lehetséges parabolaívek is kisebb intervallumban értelmezettek, így kisebb maximális meredekség és emelkedési érték adódik.

Ha kiválasztjuk az

$$(ei, y_{\text{opt}} z_{\text{opt}}), \quad e = \frac{1}{2n}, \quad i \in \{1, 2, \ldots, n\}$$

pontok közül azt, amelynél az $M(x)$ maximális, akkor a 8. ábrán látható módszerrel az $M^{\mathbf{G}}_{\text{opt}}(x_{\text{opt}})$ értéknek egy felső korlátját adhatjuk meg (9. ábra):

$$M^{\mathbf{G}}_{\text{opt}}(x_{\text{opt}}) \le M_{\max}(ke) + s(n) \quad \text{ahol} \quad k \in \{1, 2, \ldots, n\} \quad \text{és} \tag{5.3}$$

$$M_{\max}(ke) \ge M(ie) \quad \text{minden} \quad i \in \{1, 2, \ldots, n\}.$$

Megjegyzés. 1. Ha az előző gondolatmenetben y és z koordinátákról csak azt tesszük fel, hogy y és z állandók, akkor az M függvény ilyen koordinátájú pontokon felvett értékére szintén tudunk egy felső korlátot biztosítani. Az ezen koordinátájú pontoknál mért maximumhoz adjuk hozzá az $s(n)$ értéket.

2. Az előzőekben feltettük, hogy speciálisan a $P(x, y_{\text{opt}}, z_{\text{opt}})$ vagy általánosabban $P(x, y_{\text{konst}}, z_{\text{konst}})$ pontokat vizsgáljuk, ahol $x \in \left[0, \frac{1}{2}\right]$. Ebben az esetben meg is adtuk egy lehetséges felső korlátját az $M^{\mathbf{G}}_{\text{opt}}(x_{\text{opt}})$ értékének. (És nyilván ezen keresztül az optimális sugárnak és sűrűségnek is. (A konkrét tércsoport esetén azonban ezzel még nem adtunk felső korlátot az optimális sugár és sűrűség számára, hiszen a W kockára húzott Q rács nem biztos, hogy tartalmazza a $(ei, y_{\text{opt}}, z_{\text{opt}})$, $e = \frac{1}{2n}$, $i \in \{1, 2, \ldots, n\}$ pontokat.

Továbbra is jelölje $K^{\mathbf{G}}_{\text{opt}}(x_{\text{opt}}, y_{\text{opt}}, z_{\text{opt}})$ a W-ben levő egyik optimális gömb középpontját.

Definiáljuk általánosabb esetben is a (5.1) egyenlettel meghatározott M függvényt:

$$M(x, y, z) := 4\Big(r\big(P(x, y, z)\big)\Big)^2 \quad \text{ahol} \quad x, y, z \in \left[0, \frac{1}{2}\right]. \tag{5.4}$$

Jelölje továbbá $K^{\mathbf{G}}_{\max}(x_{\max}, y_{\max}, z_{\max})$ a Q rács azon pontrácsait, amelyhez egy adott G és adott n esetén a maximális gömbsugár tartozik. Ezt a sugarat a (4.2) egyenlet segítségével definiáltuk.

A 10. ábrán szimbolikusan érzékeltettünk egy mozgatást, amely a $K^{\mathbf{G}}_{\max}(x_{\max}, y_{\max}, z_{\max})$ pontot $K^{\mathbf{G}}_{\text{opt}}(x_{\text{opt}}, y_{\text{opt}}, z_{\text{opt}})$ pontba viszi. Ez a mozgatás a W kocka éleivel párhuzamosan történik, tehát úgy, hogy az x, y, z koordináták közül pontosan kettő mindig állandó (lásd 5. fejezet 1. és 2. megjegyzése).

A 10. ábrán érzékeltetett mozgatás legfeljebb három szakaszra bontható. Minden egyes ilyen szakaszra megadhatjuk a megfelelő $M(x, y, z)$ függvény egy felső korlátját. Így végül az $M^{\mathbf{G}}_{\text{opt}}(x_{\text{opt}}, y_{\text{opt}}, z_{\text{opt}})$ egy felső korlátjához jutunk. Jelöljük a 10. ábra mozgatásainak fordulópontjait K_1 illetve K_2-vel az ábrán látható módon. Az M függvény ezen pontokhoz tartozó értékeit jelöljük rendre M_1 illetve M_2-vel. Az (5.3) és a fentiek alapján:

$$\text{A } K^{\mathbf{G}}_{\max}K_1 \text{ szakaszon } : M^{\mathbf{G}}{}_{\max} + s(n) \geq M_1$$

$$\text{A } K_1K_2 \text{ szakaszon } : M^{\mathbf{G}}{}_{\max} + 2s(n) \geq M_2$$

$$\text{A } K_2K^{\mathbf{G}}_{\text{opt}} \text{ szakaszon } : M^{\mathbf{G}}{}_{\max} + 3s(n) \geq M^{\mathbf{G}}_{\text{opt}}.$$

Amiből következik, hogy:

$$M^{\mathbf{G}}_{\max} + 3s(n) \geq M^{\mathbf{G}}_{\text{opt}}.$$

Így becsléseink alapján:

$$M_{\max} + 3s(n) \geq M^{\mathbf{G}}_{\text{opt}} \geq M^{\mathbf{G}}_{\max}, \tag{5.5}$$

amiből a $\mathbf{G}$ tércsoporthoz tartozó optimális gömbsugárnak és optimális sűrűségnek is egy n-től függő alsó és felső korlátja is megadható.

(5.2) és (5.3) alapján ha $n \to \infty$ akkor a $s(n) \to 0$. Ezért ezzel az algoritmussal elvileg tetszőleges pontossággal meg lehet határozni a $\mathbf{G}$ tércsoport optimális sűrűségét. A felosztás finomításával a becsléseink szerint is jobbak, továbbá a lokalizáció lehetősége jelentősen javítja a konkrét számolások pontosságát. Az algoritmus polinomiális lépésszámú.

6. Az algoritmusra épülő program

a. Bemeneti adatok és eredmények

A 2. táblázatban ismertetjük a kockarácshoz tartozó tércsoportokra kapott optimális gömbkitöltések sűrűségét. Ezen eredmények közléséhez rendelkezésünkre áll az algoritmus alapján kifejlesztett program. Ennek felhasználói és fejlesztői dokumentációja a [8] munkában megtalálható.

A program megírása a Turbo C++ 3.1. fejlesztőrendszerének segítségével történt, a program egy objektum alapú Windows alkalmazás, mellyel a tércsoportok az eddig megszokott módon könnyen kezelhetők.

A program inputja a korábbiakban már említett kristálycsoport definíció. Ilyen csoport leírás található pl. a [14], [15] táblázatokban.

Az eredmények szöveg file-ba menthetők. Egy output file a következő adatokat tartalmazza:

1. Tércsoport neve
2. Az optimális gömbközéppont koordinátái
3. A cellába írt maximális gömb sugara
4. A cella térfogata
5. Az optimális gömbkitöltés sűrűsége.

b. Az algoritmus paraméterei

A felosztás finomsága $\frac{1}{2n}$

Ezzel a kockára húzott Q térrács finomságát adjuk meg.

A program által megengedett értékek: [2,2000] intervallumbeli egész számok.

Alapértelmezés szerinti értékek: $n := 100$.

Kezdő kocka éle:

Az első lépésben vett kocka élhosszúsága.

Megengedett értékek: $(0, 1]$ itervallumbeli valós számok.

Alapértemezés szerinti érték: 0.5.

Lehetőségünk van egy tetszőleges Q-beli pont tetszőleges környezetére futtatni az algoritmust.

Alappont koordinátái:

Ez lesz a kezdő kocka középpontja. Ezzel az alapponttal indul az algoritmus.

Megengedett értékek: $(0, 0.5)$ intervallumbeli valós számok.

Alapértelmezés szerinti kiinduló pont: $x : 0.25$ $y : 0.25$ $z : 0.25$

Lehetőségünk van az általunk kiválasztott pont egy tetszőleges környezetét vizsgálni a gömbkitöltés szempontjából.

c. Az adatszerkezet felépítése

Lehetőség van összetett adatszerkezetek megvalósítására, amelyekkel az algoritmus kódolása egyszerűbbé, áttekinthetőbbé válik. Az alábbiakban említett osztályok megvalósításával a tércsoportok könnyen kezelhetők.

A program input adata a **G** kristálycsoport definíciója. A csoportot ekvivalens pontpozíciók és rács-eltoltjaik felsorolásával lehet megadni. Az eltolás helyvektorait egy koordinátahármassal írjuk le, erre, valamint az $\mathbf{E}^3$ euklideszi tér pontjainak megadására megfelel egy Point elnevezésű osztály. A leírásban az alappont x, y, z koordinátáival kifejezve fel kell sorolni a vele ekvivalens pontok koordinátáit. A kifejezésben szerepelhet egy x, $-x$, y, $-y$, z, vagy $-z$ hivatkozás az alappont egy koordinátájára, illetve a koordináta ellentettjére, és ezt megelőzheti egy valós szám. Ezek a kifejezések a Polynom osztály objektumai. Az ekvivalens pontkoordinátákat az Expression osztállyal adjuk meg. Egy tércsoport definíció minimum 1, (a $(0,0,0)$ vektort soroljuk ide) maximum 4 centráló eltolásvektort tartalmaz, valamint 12, 24, vagy 48 ekvivalens pontpozíciót (lásd még az 1. Táblázatot). Megjegyezzük, hogy a cella térfogata ezen adatokból könnyen számolható:

D-V cella térfogat=1/(centráló eltolások száma $*$ pontpozíciók száma).

Mindezen adatszerkezetek tárolásához megvalósítunk egy paraméterezett tömbosztályt (MyArray), amellyel létrehozzuk az eltolások és ekvivalens pontpozíciók tömbjét. A program a csoport adatainak tárolásához definiál egy 4 elemű Point típusú elemeket tartalmazó-, és egy 48 elemű Expression objektumokat tartalmazó tömböt.

2. Táblázat

A tércsoport neve és száma	Az optimális gömb középpontja	Az optimális gömbsugár	Az optimális sűrűség
P23 No.195	$x = 0$ $y = 0.293$ $z = 0.293$	0.207	0.44
F23 No.196	$x = 0$ $y = 0.129$ $z = 0.208$	• 0.128532 0.129	• 0.426946 0.42
I23 No.197	$x = 0$ $y = 0.186$ $z = 0.301$	0.186	0.64
$P2_13$ No.198	$x = 0$ $y = 0.250$ $z = 0.375$	0.233	0.64
$I2_13$ No.199	$x = 0.125$ $y = 0.125$ $z = 0.375$	0.176	0.55

A tércsoport neve és száma	Az optimális gömb középpontja	Az optimális gömbsugár	Az optimális sűrűség
Pm$\bar{3}$ No.200	$x = 0.146$ $y = 0.146$ $z = 0.354$	0.146	0.32
Pn$\bar{3}$ No.201 (origó a **23**-ban)	$x = 0$ $y = 0.186$ $z = 0.301$	0.186	0.64
Fm$\bar{3}$ No.202	$x = 0.104$ $y = 0.250$ $z = 0.396$	0.104	0.45
Fd$\bar{3}$ No.203 (origó a **23**-ban)	$x = 0.032$ $y = 0.200$ $z = 0.402$	0.103	0.44
Im$\bar{3}$ No.204	$x = 0.129$ $y = 0.129$ $z = 0.311$	0.129	0.43
Pa$\bar{3}$ No.205	$x = 0.079$ $y = 0.366$ $z = 0.196$	0.176	0.55
Ia$\bar{3}$ No.206	$x = 0.183$ $y = 0.400$ $z = 0.372$	0.144	0.6
P432 No.207	$x = 0.143$ $y = 0.394$ $z = 0.306$	•0.156183 0.156	•0.383009 0.38
P4$_2$32 No.208	$x = 0$ $y = 0.250$ $z = 0.250$	0.176	0.55
F432 No.209	$x = 0.074$ $y = 0.136$ $z = 0.250$	•0.109398 0.109	•0.526492 0.53
F4$_1$32 No.210	$x = 0.069$ $y = 0.206$ $z = 0.431$	0.097	0.37

A tércsoport neve és száma	Az optimális gömb középpontja	Az optimális gömbsugár	Az optimális sűrűség
I432 No.211	$x = 0.129$ $y = 0.129$ $z = 0.311$	0.128	0.42
P4$_3$32 No.212	$x = 0.125$ $y = 0.125$ $z = 0.375$	0.176	0.55
P4$_1$32 No.213	$x = 0.125$ $y = 0.375$ $z = 0.375$	0.176	0.55
I4$_1$32 No.214	$x = 0.246$ $y = 0.173$ $z = 0.375$	0.125	0.39
P$\bar{4}$3m No.215	$x = 0$ $y = 0.185$ $z = 0.369$	•0.130601 0.131	•0.223948 0.22
F$\bar{4}$3m No.216	$x = 0$ $y = 0.125$ $z = 0.250$	•0.088388 0.088	•0.277680 0.28
I$\bar{4}$3m No.217	$x = 0$ $y = 0.178$ $z = 0.355$	•0.125529 0.126	•0.397710 0.40
P$\bar{4}$3n No.218	$x = 0.125$ $y = 0.125$ $z = 0.375$	0.176	0.55
F$\bar{4}$3c No.219	$x = 0.074$ $y = 0.250$ $z = 0.364$	0.109	0.52
I$\bar{4}$3d No.220	$x = 0.132$ $y = 0.390$ $z = 0.420$	0.143	0.58
Pm$\bar{3}$m No.221	$x = 0.104$ $y = 0.250$ $z = 0.396$	•0.103553 0.104	•0.223266 0.22

A tércsoport neve és száma	Az optimális gömb középpontja	Az optimális gömbsugár	Az optimális sűrűség
Pn$\bar{3}$n No.222 (origó a **43**-ban)	$x = 0.151$ $y = 0.435$ $z = 0.320$	0.135	0.49
Pm$\bar{3}$n No.223	$x = 0.129$ $y = 0.129$ $z = 0.311$	0.129	0.43
Pn$\bar{3}$m No.224 (origó a $\bar{4}$**3m**-ben)	$x = 0$ $y = 0.178$ $z = 0.355$	•0.125529 0.126	•0.397710 0.40
Fm$\bar{3}$m No.225	$x = 0.158$ $y = 0.250$ $z = 0.435$	•0.06531 0.065	•0.223949 0.22
Fm$\bar{3}$c No.226	$x = 0.076$ $y = 0.155$ $z = 0.301$	0.076	0.35
Fd$\bar{3}$m No.227 (origó a $\bar{4}$**3m**-ben)	$x = 0$ $y = 0.089$ $z = 0.178$	•0.012552 0.013	•0.198855 0.20
Fd$\bar{3}$c No.228 (origó a **23**-ban)	$x = 0.125$ $y = 0.125$ $z = 0.500$	0.088	0.55
Im$\bar{3}$m No.229	$x = 0.084$ $y = 0.202$ $z = 0.320$	•0.083568 0.084	•0.234685 0.23
Ia$\bar{3}$d No.230	$x = 0.053$ $y = 0.352$ $z = 0.461$	0.114	0.60

(A táblázatban • jelöli az optimális sugár illetve sűrűség pontos értékeit [12], [13].

Ezúton is szeretnénk köszönetet mondani Molnár Emil kollégának a dolgozat elkészítéséhez nyújtott sok értékes tanácsáért és segítségéért.

IRODALOM

[1] Bieberbach, L., Über die Bewegungsgruppen der euklidischen Räume I–II, *Math. Annalen* **70** (1910), 297–336; **72** (1912) 400–412.

[2] Brown, H., Bülow, R. Neubüser, J. Wondratschek, H., Zassenhaus, H., *Crystallographic Groups of Four-Dimensional Space* (New York, 1978).

[3] Diószegi, F., „Tércsoportok Dirichlet–Voronoj celláinak és optimális gömbkitöltésének meghatározása számítógéppel", szakdolgozat, ELTE TTK, 1989.

[4] Fedorov, E. S., Reguläre Plan und Raumtheilung, *Abh. K. Bayer. Akad. d. Wiss.* **11** (1889) Cl. XX. Abth. 11, 465–588.

[5] Fedorov, E. S., Zusammenfassung der kristallographischen Resultate des Herrn Schoenflies und der meinigen. *Z. Kristallogr.* **20** (1982) S. 25–75.

[6] Fejes-Tóth, G., „New results in the Theory of Packing and Covering" in: *Convexity and its Applications* Ed. P. M. Gruber, J. M. Wills, (1983).

[7] G. Horváth, Á., Molnár, E., Densest ball packing by orbits of the 10 fixed point free euklidean space groups, *Studia Sci. Math. Hung.* **29** (1994) 9–23.

[8] Máté, Cs., „Tércsoportok optimális gömbkitöltésének meghatározása számítógéppel", szakdolgozat, ELTE TTK. (Prog. Mat.), 1994.

[9] Molnár, E., Konvexe Fundamentalpolyeder und D-V Zellen für 29 Raumgruppen, die Coxetersche Spiegelungsuntergruppen enthalten, *Beiträge zur Algebra und Geometrie* **14** (1983) 33–75.

[10] Schoenflies, A., *Kristallsysteme und Kristallstruktur* (Leipzig, 1881).

[11] Sinogowitz, U., Herteilung aller homogenen, nicht kubischen Kugelpackungen, *Z. Kristallographie* **105** (1943) 23–52.

[12] Szirmai, J., Optimale Kugelpackungen für die Raumgruppen F23, P432 und F432, *Periodica Polytechnica Ser. Mech. Eng.* **36** No. 3–4 (1992) 317–331.

[13] Szirmai, J., Néhány tércsoport optimális gömkitöltése. *Alkalmazott Matematikai Lapok* **17** (1993) 87–99.

[14] *Atlas postranstvennyh grupp kubicheskoi sistemy* (Nauka, Moskva, 1980).

[15] International Tables for X-Ray Crystallography, Vol 1., Henry, N. F. M and Lonsdale, K. Symmetry Groups, *Kynoch Press, Birmingham* 1969. New edition by Theo Hahn, Vol. A, *Reidel Co, Dordrecht* 1983.

(Beérkezett 1997. január 30.)

MÁTÉ CSILLA ÉS SZIRMAI JENŐ
BUDAPESTI MŰSZAKI EGYETEM,
MATEMATIKAI INTÉZET, GEOMETRIAI TANSZÉK
H-1521, BUDAPEST, EGRY J. UTCA 1, II. EMELET 22.
e-mail: geometry@ccmail.bme.hu
szirmai@math.bme.hu

DETERMINATION OF DENSEST BALL PACKINGS UNDER CUBIC CRYSTALLOGRAPHIC GROUPS BY COMPUTER

CSILLA MÁTÉ AND JENŐ SZIRMAI

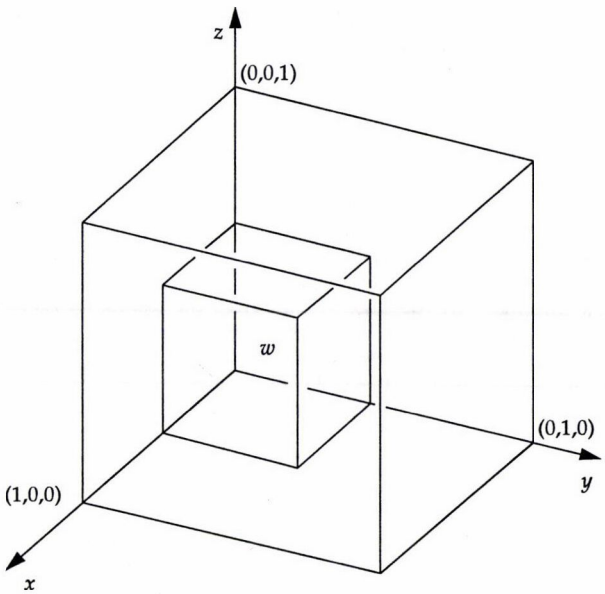
The famous unsolved KEPLER conjecture [6] about the densest ball packing of the whole Euclidean space $\mathbf{E}^3$ with equal balls motivated the initiative of U. Sinogowitz who posed the problem to find the densest homogeneous ball packing under a given space group [11]. The maximal density $\pi/\sqrt{18} \approx 0.74048$ of the lattice-like ball packing occurs at other space groups as well [7].

The author reports a computer algorithm, which determines the densest simple transitive ball packing for each cubic crystallographic space group.

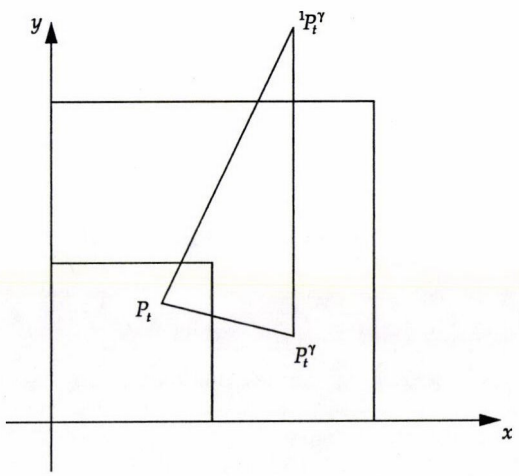
The author proves here the convergence and gives the results in Table 2 where the known exact data [12], [13] are indicated too. A complete algorithm for every orbit type is in progress on the base of [9].

MSC(1991) Codes: 51M20, 52C17, 65Y25

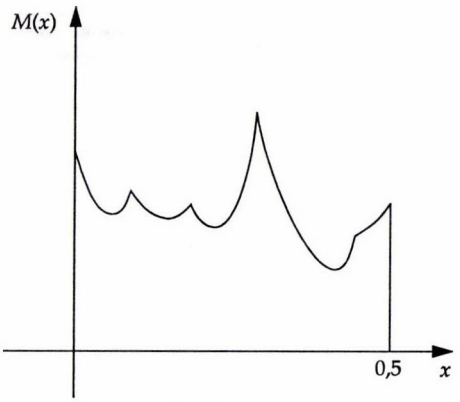
Key Words and Phrases: Densest ball packings under cubic crystallographic groups by computer.



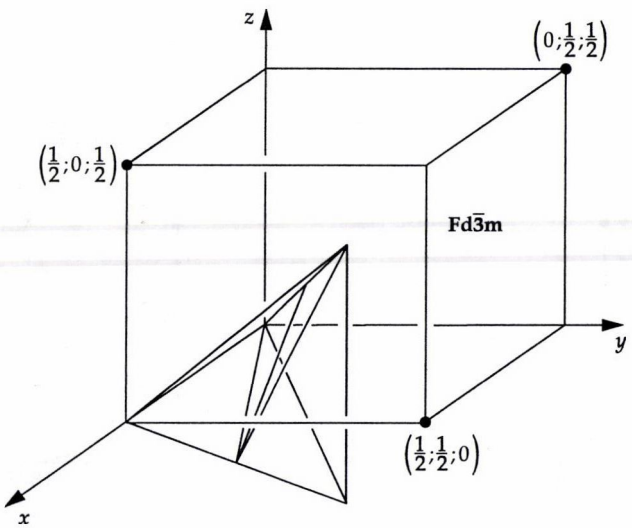
1. ábra



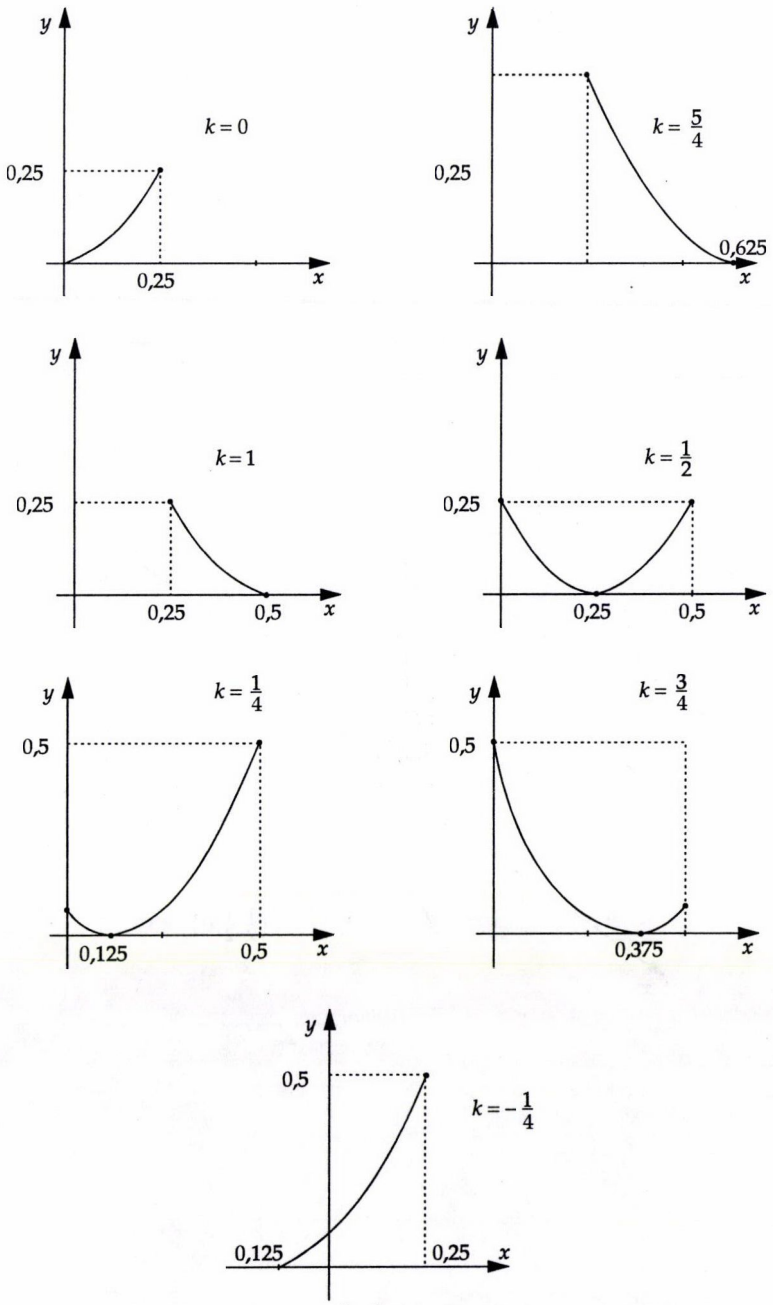
2. ábra


M(x)
0,5
x

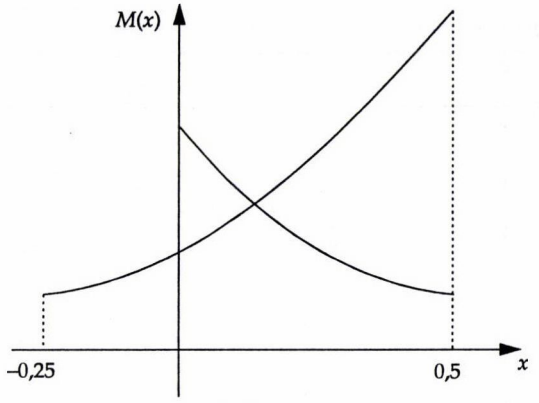
3. ábra


z
$\left(0;\frac{1}{2};\frac{1}{2}\right)$
$\left(\frac{1}{2};0;\frac{1}{2}\right)$
Fd$\bar{3}$m
y
$\left(\frac{1}{2};\frac{1}{2};0\right)$
x

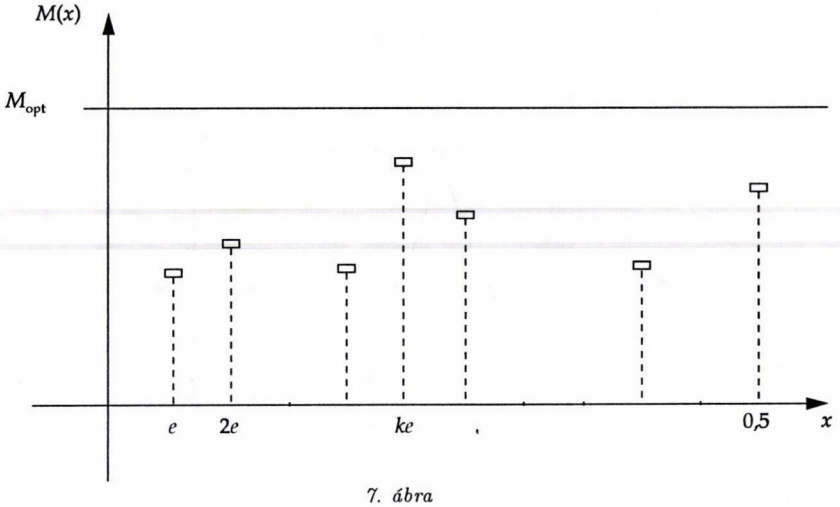
4. ábra


y
k = 0
0,25
0,25
x
y
k = 5/4
0,25
0,625
x
y
k = 1
0,25
0,25
0,5
x
y
k = 1/2
0,25
0,25
0,5
x
y
k = 1/4
0,5
0,125
0,5
x
y
k = 3/4
0,5
0,375
x
y
0,5
k = −1/4
0,125
0,25
x

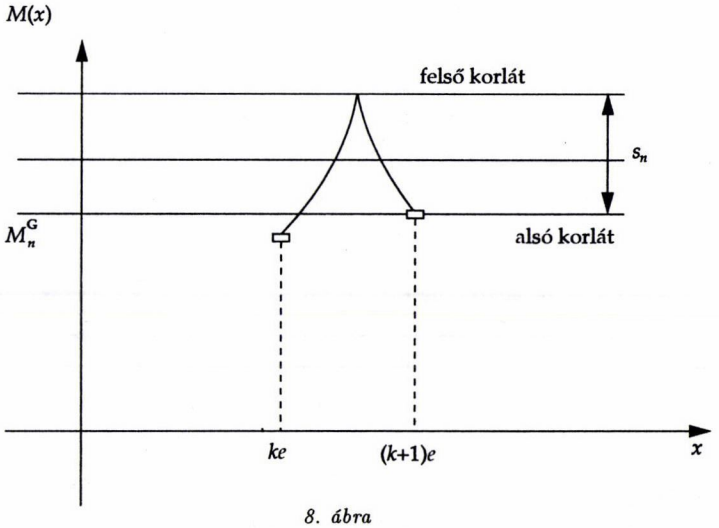
5. ábra



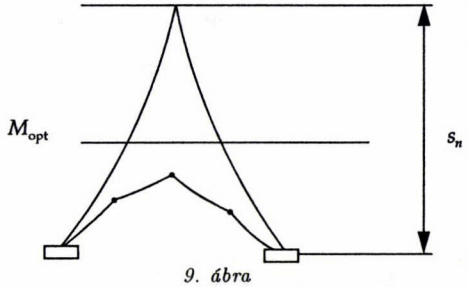
6. ábra



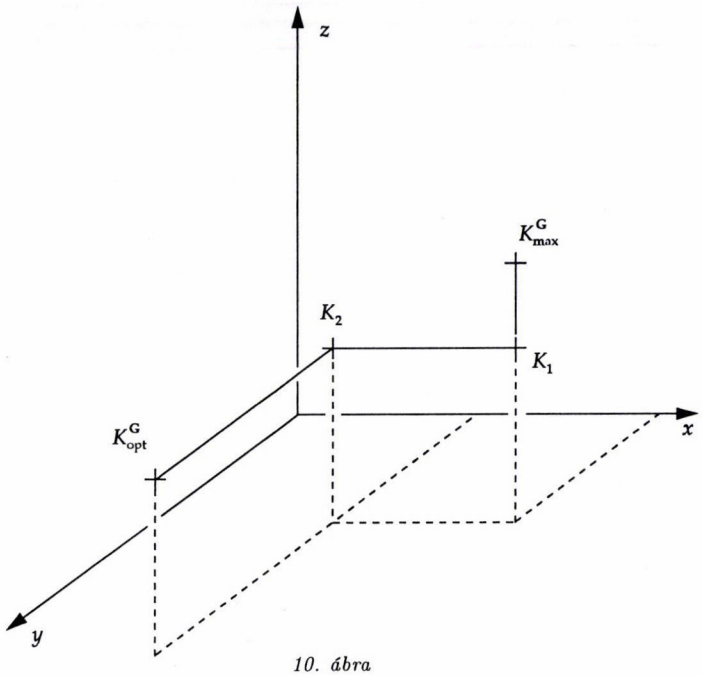
7. ábra



8. ábra



9. ábra



10. ábra

Alkalmazott Matematikai Lapok **19** (1999) 113–114

Könyvismertetés

SZILI LÁSZLÓ ÉS TÓTH JÁNOS: Matematika és *Mathematica,* ELTE Eötvös Kiadó, Budapest, 1996.

Nagyon nehezen, valószínűtlenül hosszú idő után sikerült formába öntenem az alábbiakat. (Közben a 2.2. verzió után megjelent a 3.0. változat is — lásd Természet Világa 1998. november.) Igazából máig sem sikerült tisztáznom magamban, mi a könyv pontos célja; kiknek íródott; kinek ajánljam, mint felsőoktatásban tanító pedagógus. Ezzel kapcsolatos kételyeim, morgolódásaim helyett inkább pozitív élményeket szeretnék rögzíteni. Ugyanakkor el kell ismernem, hogy ennél jobb, komplexebb könyvet nem tudok elképzelni — csakhogy miről is, milyen célra is?

Az már történelem, hogy a hatvanas évektől a számítógépek terjedése jelentősen ösztönözte a matematika fejlődését. Napjainkban a program(csomagok) fejlődése forradalmasítja a matematikával, mint tudománnyal való tevékenységet (kutatást). Minderről a könyv bevezetőjében bőségesen olvashatunk, akárcsak az egyes programok összehasonlításáról. Egy átlagos kutató, de egy picit is igényes egyetemi oktató sem engedheti meg magának azt a luxust, hogy ne használjon rendszeresen valami olyan segédanyagot (szoftvert), amely jelenségeket, összefüggéseket demonstrál, sejtéseket kipróbál, eredményeket diszkutál, stb. Az ilyen lehetőségek fantasztikusan gazdag gyűjteménye a *Mathematica* programcsomag, ami „azt is tudja, amit el se tudunk képzelni". (Különben ez a fő „hibája" is: valószínűleg egy teljes élet is kevés az alapos megismeréséhez, és ennek megfelelően a kézikönyvek, a súgó egyaránt nagyon nehézkesek — bár a HELPje sokkal kellemesebb, mint amit a WINDOWS-ban megszokhattunk. Inkább a programozási nyelvek súgóinak PÉLDÁI köszönnek vissza!)

Maga a könyv szerintem egy csokrot próbál átnyújtani, hogy a matematika egyes területein hol, hogyan érdemes a *Mathematica*-hoz nyúlni. A nagyobb témakörökön belüli igen részletes felosztásnak köszönhetően a triviális példákon kívül olyan speciális problémakörökből olvashatunk feladatokat, mint a

— stabilitáselmélet és variációszámítás a differenciálegyenleteken belül;

— lánctörtek, Moebius-függvény, tizedes törtek közönségessé átírása számelméletből;
— Markov láncok, korreláció számítása, maximum likelihood becslés valószínűségszámításból;
— feszítő fák a gráfelméletből;
— polinomrendszerek megoldása Gröbner-bázisokkal;
— lineáris programozás;
— animáció (függvényrendszerek ábrázolása); stb.

A szerzők teljességre törekvése miatt gyakran volt olyan érzésem, hogy „ezt akkor sem értettem (volna), amikor tanultam az egyetemen", de aztán rájöttem, hogy szerencsére nem vizsgáznom kell a könyvből, így ezt a részt nyugodtam átugorhatom! Egyre több részről derült ki, hogy nekem (is) szól, sőt egyre több dolgot sikerült is reprodukálni, bizonyosakat tovább is fejlesztettem. Ezután már saját problémáimat kezdtem megoldani, és pillanatnyilag sokkal több időt töltök *Mathematica*-val, mint DERIVE-val és CABRI-val együttvéve — persze ebben az újdonságnak is szerepe van, és nem is lesz így mindig.

Természetesen nagyon sokféle könyvet lehet írni a *Mathematica*-ról. (Kétszáznál több kötetre rúg az irodalma!) Még több lehetőség adódik, ha az adott címre szorítkozunk. A jelen könyv szerzői nem sokat meditáltak, optimalizáltak: megírták, amit írtak. Kifejezett céljuk volt a program népszerűsítése, reklámozása. Rengeteg frappáns példájuk; az elviselhető mennyiségű eszközismertetés (a kulcsszavak tizedét se tukmálják az olvasóra); kísérletezésre való buzdításuk az, ami rávett arra, hogy egy (nem tőlük kapott) kalózmásolattal leüljek, s kipróbáljam. Már keresek elfogadható árú verziót, illetve szeretném az oktatásban is használni. Pillanatnyilag ketten írnak szakdolgozatot nálam a *Mathematica* iskolai (14–16 év közötti tanulók számára) oktathatóságából.

Az árral kapcsolatban: az általam ismert legolcsóbb lehetőség a következő: 10 munkaállomásra szóló licence ára $3 \times 1140 = 3420$ angol fontnak megfelelő forint, amit 3 évi részletben kell fizetni, és ezalatt automatikusan megkapjuk az esetleges újabb verziókat.

TÖRÖK TURUL

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 után, a kézirat befejezéseképpen fel kell tüntetni a szerző teljes nevét és a munkahelye (illetve lakása) pontos 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 kel 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 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-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 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* (Elsevier Publishing Company, Amsterdam and New York, 1960).

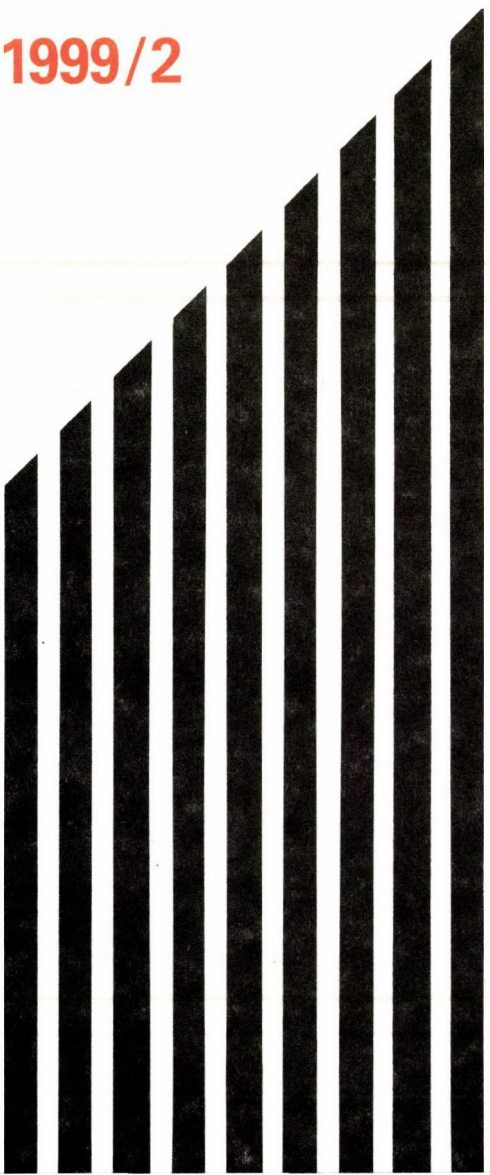
A dolgozatok szövegében az irodalmi hivatkozás számait szögletes zárójelben kell megadni, mint például [5] vagy [4, 76–78]. A szerzők a dolgozatukról 50 darab ingyenes különlenyomatot kapnak. A dolgozatok után szerzői díjat az Alkalmazott Matematikai Lapok nem fizet.

TARTALOMJEGYZÉK

INDEX

Alkalmazott matematikai lapok

1999/2



A MAGYAR TUDOMÁNYOS AKADÉMIA
MATEMATIKAI TUDOMÁNYOK
OSZTÁLYÁNAK KÖZLEMÉNYEI

19.
KÖTET

ALKALMAZOTT MATEMATIKAI LAPOK

A MAGYAR TUDOMÁNYOS AKADÉMIA MATEMATIKAI 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

XIX. kötet 2. szám

Szerkesztőség és kiadóhivatal: 1027 Budapest, Fő u. 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) 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 u. 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) Osztálya a következő idegen nyelvű folyóiratokat adja ki:

1. Acta Mathematica Hungarica,
2. Acta Physica Hungarica.

Alkalmazott Matematikai Lapok **19** (1999) 115–132

EGY CÍMKÉZŐ ELJÁRÁS A LEGRÖVIDEBB UTAK FÁJÁNAK MEGHATÁROZÁSÁRA RITKA HÁLÓZATOKBAN

MARTON LÁSZLÓ

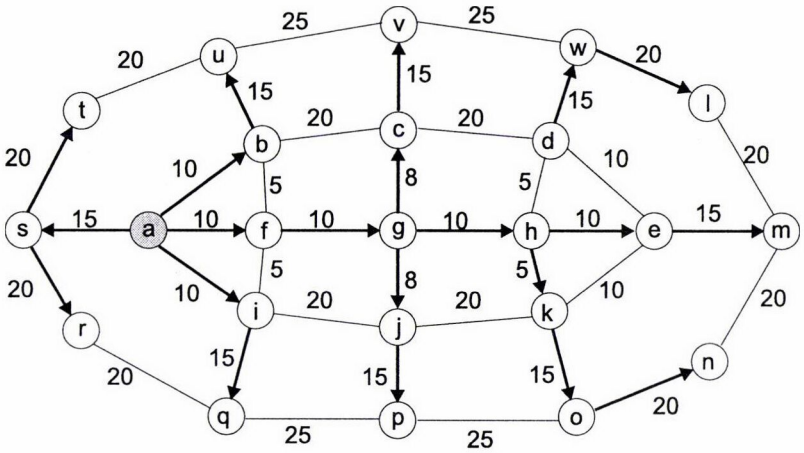
Győr

Számos hálózati alkalmazás számítógépes modelljének központi, a számitástechnikai hatékonyság szempontjából meghatározó része az egy pontból mint kezdőpontból kiinduló minimális hosszú utak fájának meghatározása. Gyakran — így például a közlekedéstervezési feladatoknál is — előforduló speciális eset, hogy a hálózat ritka, az egy pontból kiinduló élek száma korlátozott és minden pontból, vagy viszonylag nagyszámú pontból kiindulva meg kell határoznunk a minimális utak fáját. A cikkben ilyen jellegű feladatokra javaslunk egy faépítő címkéző algoritmust, bizonyítjuk helyességét és hatékonyságát, ez utóbbit számítógépes demonstrációval is kiegészítve. A bemutatott algoritmus a szakirodalomban jól ismert általános címkéző eljárás (general label-setting method) egy, az egy pontból kiinduló élek hossz szerinti előrendezését kihasználó változata.

1. Bevezetés

Hálózatunk egy *egyszerű, irányított, összefüggő* gráf, az élekhez rendelt nemnegatív értékek az élhosszak. A hálózat pontjait 1-gyel kezdődő sorszámozással azonosítjuk, a legnagyobb sorszámú pontot jelölje N, az (i,j) jelöli az i pontból a j pontba mutató élt, $h(i,j)$ ennek a hosszát. Az i-ből a j-be vezető útnak nevezünk egy $P(i,j) = (i_1,j_1),\ldots,(i_k,j_k)$ élsorozatot, amelyre igaz, hogy $i_1 = i$, $j_k = j$ és $j_l = i_{l+1}$ ha $1 \leq l < k$. Az út hosszán az élek összhosszát értjük. *Körútnak* nevezünk egy utat, ha kezdő és végpontja azonos. *Szimmetrikus* a hálózat, akkor, ha az $)i,j)$ él létezéséből következik a (j,i) él létezése, és $h(i,j) = h(j,i)$. Példahálózatunk (ld. 1.–3. ábrák) ilyen, az ábrákon az élek irányítását nem jelöltük, illetve minden szakasz két élt jelent. A szimmetricitás fejtegetéseink szempontjából lényegtelen tulajdonság. Megjegyezzük, hogy az ábrákon — a könnyebb áttekinthetőség kedvéért — a pontokat betűkkel azonosítjuk.

Irányított fának nevezünk egy olyan részhálózatot, amely nem tartalmaz körutat és egy kijelölt pontból a *gyökér* vagy *kezdőpontból* minden más pontba pontosan egy út vezet. *Minimális fának* nevezünk egy olyan irányított fát, amelyben min-



1. ábra

den út — az eredeti hálózatban a kezdőpontból a végpontba vezető utak között — minimális hosszúságú (1. ábra, az a kezdőponttal).

Az irányított fa egyszerűen leírható az ún. *címkékkel,* amelyek minden ponthoz (kivéve a kezdőpontot) megadják a kezdőpontból hozzá vezető úton a megelőző pontot. A fában mint *távolságot* rendeljük hozzá minden ponthoz kezdőpontból hozzá vezető út hosszát (2.–3. ábra). Az i pont címkéjét jelölje $c(i)$, távolságát pedig $t(i)$. A kezdőpont távolsága a definíció szerint 0, ha technikailag szükséges, akkor címkét is rendelünk hozzá, ez lehet bármilyen megkülönböztető érték (önmaga vagy 0).

A minimális fákat az ún. faépítő (*tree building*) technikával meghatározó eljárások mind elméleti, mind gyakorlati szempontból jelentősek. Ennek oka az, hogy egyrészt nem ismert olyan, a faépítésnél hatékonyabb eljárás, amely csak egy adott viszonylatban, tehát két pont között keresné meg a legrövidebb utat, másrészt a jelenleg ismert, a legrövidebb utat minden viszonylatban meghatározó algoritmusok nagy hálózatoknál a gyakorlatban már nehezen alkalmazhatók. A témakör összefoglalását és bő irodalomjegyzékét találhatjuk a [3, 4] cikkekben.

2. Alapeljárás

A faépítés egyik (sok és sokféle változattal bővelkedő) alap-algoritmusa az általános címkéző eljárás (*general label-setting method*) vagy más néven Dijkstra — féle eljárás [1], amelynek egy, a számítástechnikai megvalósításhoz közelítő megfogalmazása a következő:

Definiáljuk a hálózat pontjainak két részhalmazát:

— a K a *kész pontok* halmaza, elemei már végleges (nem rövidíthető) távolsággal és végleges (nem áthelyezhető) címkével rendelkeznek

— az A az *aktív pontok* halmaza (aktivitás halmaz), elemeinek már van (ideiglenes) címkéje és véges a távolsága de ezek még változhatnak.

Lépések:

a, Rendeljük a kezdőponthoz a 0, a többihez a végtelen távolságértéket. A kezdőpont címkéje legyen önmaga. A K legyen üres, az A tartalmazza csak a kezdőpontot.

b, Válasszuk ki az A minimális távolságú elemét, jelölje ezt i. (Ha több pont is minimális távolságú, a legkisebb sorszámút választjuk, az egyértelműség kedvéért.) Az i-t töröljük az A-ból és vegyük hozzá a K-hoz.

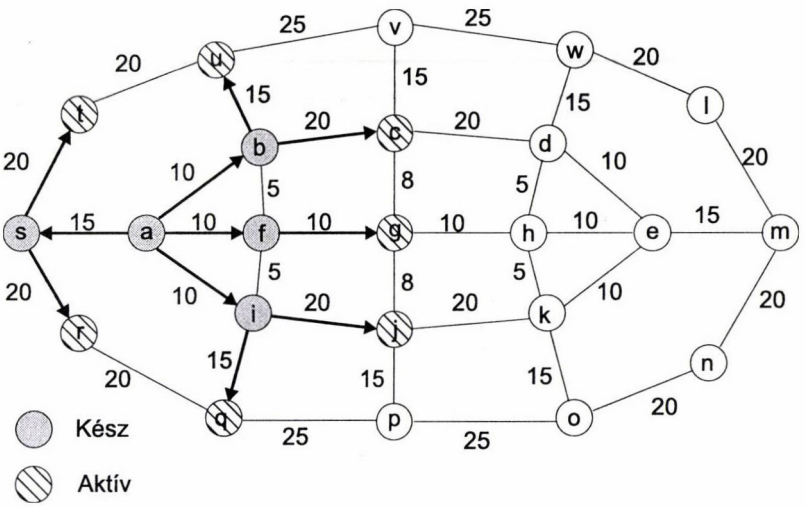
c, Minden olyan (i, j) élre, amelyre a $t(i) + h(i, j) < t(j)$ rövidítési lehetőség fennáll:

c1, Ha a j pont még nincs az A-ban, akkor hozzávesszük, címkéjét beállítjuk és távolságát módosítjuk: $c(j) = i$ $t(j) = t(i) + h(i, j)$

c2, Ha a j pont már az A-ban van, akkor címkéjét és távolságát módosítjuk: $c(j) = i$ $t(j) = t(i) + h(i, j)$

d, Ha az A halmazban van elem, folytatjuk a **b,** lépéstől, ha az A üres, az eljárás véget ért, a címkék meghatározzák a minimális fát.

Az algoritmus első néhány lépését a példahálózaton a 2. ábra és a hozzá tartozó 2.1. táblázat mutatja be.



2. ábra

Az algoritmus helyességének bizonyítása jól ismert (ld. pl. [1]), ezzel kapcsolatban csak azt jegyezzük meg, hogy a hálózatban egy kezdőponthoz több minimális fa is tartozhat, ezekben a pontok távolságadatai szükségképen azonosak, de a címkék eltérhetnek.

2.1. táblázat

K: Kész A: Aktív C: Cimke T: Távolság																							
1	**K**	*a*																					
	A	*b*	*f*	*i*	*s*																		
		a	**b**	**c**	**d**	**e**	**f**	**g**	**h**	**i**	**j**	**k**	**l**	**m**	**n**	**o**	**p**	**q**	**r**	**s**	**t**	**u**	**v**	**w**
	C	*a*	*a*				*a*			*a*										*a*				
	T	0	10	~	~	~	10	~	~	10	~	~	~	~	~	~	~	~	~	15	~	~	~	~
2	**K**	*a*	*b*																				
	A	*f*	*i*	*s*	*u*	*c*																	
		a	**b**	**c**	**d**	**e**	**f**	**g**	**h**	**i**	**j**	**k**	**l**	**m**	**n**	**o**	**p**	**q**	**r**	**s**	**t**	**u**	**v**	**w**
	C	*a*	*a*	*b*			*a*			*a*										*a*		*b*		
	T	0	10	30	~	~	10	~	~	10	~	~	~	~	~	~	~	~	~	15	~	25	~	~
3	**K**	*a*	*b*	*f*																			
	A	*i*	*s*	*g*	*u*	*c*																	
		a	**b**	**c**	**d**	**e**	**f**	**g**	**h**	**i**	**j**	**k**	**l**	**m**	**n**	**o**	**p**	**q**	**r**	**s**	**t**	**u**	**v**	**w**
	C	*a*	*a*	*b*			*a*	*f*		*a*										*a*		*b*		
	T	0	10	30	~	~	10	20	~	10	~	~	~	~	~	~	~	~	~	15	~	25	~	~
4	**K**	*a*	*b*	*f*	*i*																		
	A	*s*	*g*	*q*	*u*	*j*	*c*																
		a	**b**	**c**	**d**	**e**	**f**	**g**	**h**	**i**	**j**	**k**	**l**	**m**	**n**	**o**	**p**	**q**	**r**	**s**	**t**	**u**	**v**	**w**
	C	*a*	*a*	*b*			*a*	*f*		*a*	*i*							*i*		*a*		*b*		
	T	0	10	30	~	~	10	20	~	10	30	~	~	~	~	~	~	25	~	15	~	25	~	~
5	**K**	*a*	*b*	*f*	*i*	*s*																	
	A	*g*	*q*	*u*	*j*	*c*	*r*	*t*															
		a	**b**	**c**	**d**	**e**	**f**	**g**	**h**	**i**	**j**	**k**	**l**	**m**	**n**	**o**	**p**	**q**	**r**	**s**	**t**	**u**	**v**	**w**
	C	*a*	*a*	*b*			*a*	*f*		*a*	*i*							*i*	*s*	*a*	*s*	*b*		
	T	0	10	30	~	~	10	20	~	10	30	~	~	~	~	~	~	25	35	15	35	25	~	~

Az algoritmus hatékonyságáról első közelítésben azt mondhatjuk, hogy a hálózat pontjai számának négyzetével arányos a szükséges műveletek száma, hiszen:

— Az algoritmus minden **b**–**c** ciklusmenetében egy pont bekerül a K halmazba, tehát pontosan N menet szükséges.

— Minden menetben lezajlik egy, maximum $N-1$ elem közti minimumkeresés (A halmaz) és maximum $N-1$ élre vonatkozó rövidítési vizsgálat.

Az a tény, hogy a K halmazba egyszer már bekerült pont nem „reaktiválódhat", nem léphet vissza az A-ba, az algoritmus helyességbizonyításának része.

Nyilvánvaló, hogy a K halmaz csak a leírást és megértést (valamint a helyességbizonyítást) könnyíti, a számítástechnikai realizációkból el is hagyható.

Könnyen belátható az is, hogy bármilyen faépítő algoritmust is konstruálunk, a hálózat minden élét legalább egyszer meg kell vizsgálni, és ezt a Dijkstra algoritmus élenként pontosan egyszer teszi meg.

Ebből következően, az ebből az alapalgoritmusból származtatott, ennek alapelveit (egyszeri élvizsgálat, *kész* halmaz, *aktív* halmaz, *minimumkiválasztás*) követő algoritmusok hatékonysága az alábbi tényezőkön múlik:

— Az aktivitás halmaz elemszáma.

— Az aktivitás halmaz tárolási, kezelési módja, az e célra választott adatstruktúra.

A szakirodalom számos, a második tényezőre vonatkozó változatot, reprezentációt ismer, a következőkben ezeket foglaljuk össze.

Az aktivitás halmaz számítástechnikai reprezentációi az alábbi három típusba oszthatók:

2.1. *Rendezetlen* (halmaz, rendezetlen tömb, lista)

2.2. *Lineárisan rendezett* (tömb,lista)

2.3. *Többszintű ill. részben rendezett* (összetett listák, fák, kupacok stb.)

A rendezettség természetesen mindig a pont távolsága szerinti növekvő (nemcsökkenő) rendezettséget jelent.

Az algoritmusnak az aktivitás halmazt kezelő lépéseit elemezve látható:

b lépés: a *minimumkiválasztás* a 2.1. típusban egy tényleges, az elemszámmal egyenesen arányos lépésszámú minimumkeresést jelent, a másik két típusban a minimum közvetlenül adódik (első elem, csúcs elem). A *törlés* az aktivitáshalmazból az első két típusban nem műveletigényes, a 2.3. típusban viszont a legtöbb esetben a struktúra törlés utáni helyreállítása ugyanolyan műveletigényes feladat mint a besorolás.

c1 lépés: ez a 2.1. típusban nem műveletigényes, a másik kettőben viszont egy új elem *besorolását*, vagyis strukturábani helyének megkeresését és beillesztését jelenti, ami az elemszám valamilyen (a 2.2. típusban lineáris, a 2.3.-ban általában logaritmikus) függvénye.

c2 lépés: a műveletigény azonos jellegű mint a **c1** lépésnél, azzal az eltéréssel, hogy a 2.2. és 2.3. típusban *átsorolásról*, vagyis egy már a struktúrában lévő pont új helyének megkereséséről és áthelyezéséről van szó.

3. Módosított eljárás

3.1. Elvi megfontolások

Az előző elemzésből levonhatjuk a következtetést, hogy az aktív elemek számának csökkentése általános jelleggel javulást eredményez az algoritmus hatékonyságában.

A javítás alapötletét az adja, hogy konkrét példákat vizsgálva (2. ábra) úgy tűnik, hogy az aktivitás halmaz túl bő, sok olyan pont van, amely a halmazba való bekerüléséhez képest csak több lépésnyi várakozás után kerül a minimumpozici-

óba. Célszerű lenne tehát egy-egy lépésben kevesebb új pontot bevonni, elsősorban olyanokat, amelyek nagyobb eséllyel juthatnak a minimumpozicióba.

Erre egy lehetőséget ad, ha a hálózat éleit, pontosabban pontonként a kiinduló éleket, növekvő (nemcsökkenő) hossz szerint átrendezzük. Ez az *előrendezés* bizonyos típusú hálózatoknál és feladatoknál nem okoz lényeges többlet-műveletigényt. Ha például

— *a hálózatban a az egy pontból kiinduló élek száma egy* $k \ll N$ *konstanssal korlátozott, és*

— *a feladat minden (vagy viszonylag nagy számú) minimális fa meghatározása*

akkor az előrendezés műveletigényének egy fára jutó hányada $\approx k^2$, vagyis az N^2-hez képest egy elhanyagolhatóan kicsi konstans.

A rendezettség nyújt lehetőséget arra, hogy egy pont kész állapotba kerülésekor kevesebb új pont lépjen be az aktivitás halmazba, de természetesen gondoskodnunk kell arról, hogy sohasem hiányozhasson a következő minimum pontja.

Az előrendezés mint javító tényező egy konkrét algoritmussal kapcsolatban felvetődik a [5] cikkben. Egy hasonlóan az előrendezést feltételező faépítő algoritmust találhatunk a [2] dolgozatban is.

Jelen cikkben, megtartva a Dijkstra eljárás meghatározó jellemzőit, egy más címkézési, faépítési technikát írunk le, bizonyítva helyességét és elvi hatékonyságát.

3.2. Algoritmus

Az egyértelműség kedvéért úgy rendezünk, hogy ha két él azonos hosszú, akkor az kerüljön előre,amelyiknek kisebb a sorszáma. Az egy pontból kimutató élek sorrendjének nyilvántartására és kezelésére egy m mutatót vezetünk be. Az algoritmus végrehajtása folyamán az $m(i)$ jelenti az i pontból kimutató, soronkövetkező, még nem vizsgált él sorszámát. Kezdetben minden i-re, amelyből indul ki él, az $m(i) = 1$.

Az algoritmus leírásánál minden olyan elemet (halmaz, függvény, lépés stb.), amely az alapeljárásban is szerepel, és ugyanolyan jelentésű, szerepű de nem feltétlenül azonos tartalmú, értékű a két algoritmusban, az eredeti jelölés vesszős változatával jelölünk.

Lépések:

a′, Rendeljük a kezdőponthoz a 0, a többihez a végtelen távolságértéket. A kezdőpont címkéje legyen önmaga. A K' legyen üres, az A' tartalmazza csak a kezdőpontot. Minden i-pontra, amelyből indul ki él, legyen az $m(i) = 1$.

b′, Válasszuk ki az A' minimális távolságú elemét, jelölje ezt i'. (Ha több pont is minimális távolságú, a legkisebb sorszámút választjuk, az egyértelműség kedvéért.) Az i'-t töröljük az A'-ből és vegyük hozzá a K'-höz. Jelölje az i' címkéjét e, tehát $e = c'(i')$.

c′, Mind az e, mind az i' pontra hajtsuk végre az alábbi (**c1′–c4′**) belső eljárást.

c1′, Jelölje a pontot x.

c2′, Az $m(x)$ által mutatott éllel indulva, az x-ből kimutató éleket egyenként sorra véve (és természetesen mutatót léptetve) eljutunk az első olyan (x, y) élhez, amelyre a $t'(x) + h(x, y0) < t'(y)$ rövidítési lehetőség fennáll, vagy már nincs több, még nem vizsgált x kezdőpontú él. Az utóbbi esetben a belső eljárásnak vége, az előbbi esetben folytatjuk:

c3′, Ha az y még nem aktív, hozzávesszük az A'-höz, címkéjét, távolságát, és az x mutatóját módosítjuk:

$$c'(y) = x \qquad t'(y) = t'(x) + h(x, y) \qquad m(x) = m(x) + 1$$

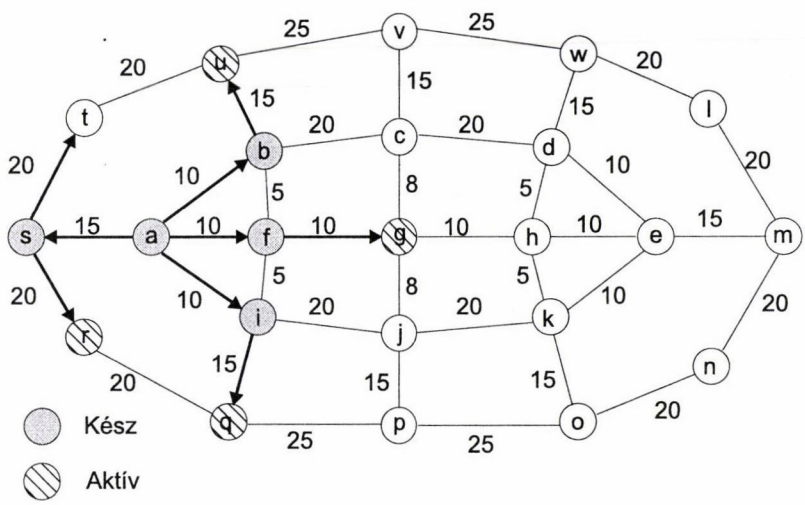
és a belső eljárásnak ezen a ponton is vége van. Ha az y már aktív volt, folytatjuk:

c4′, Mivel az y aktív, van címkéje, jelölje ezt z, tehát $z = c'(y)$. Mint az előző pontban:

$$c'(y) = x \qquad t'(y) = t'(x) + h(x, y) \qquad m(x) = m(x) + 1$$

A z ponttal ismételjük meg a vizsgálatot, tehát legyen $x = z$ és menjünk vissza a **c2′** pontra.

d′, Ha az A' halmazban van elem, folytatjuk a **b′,** lépéstől, ha az A' üres, az eljárás véget ért, a címkék meghatározzák a minimális fát.



3. ábra

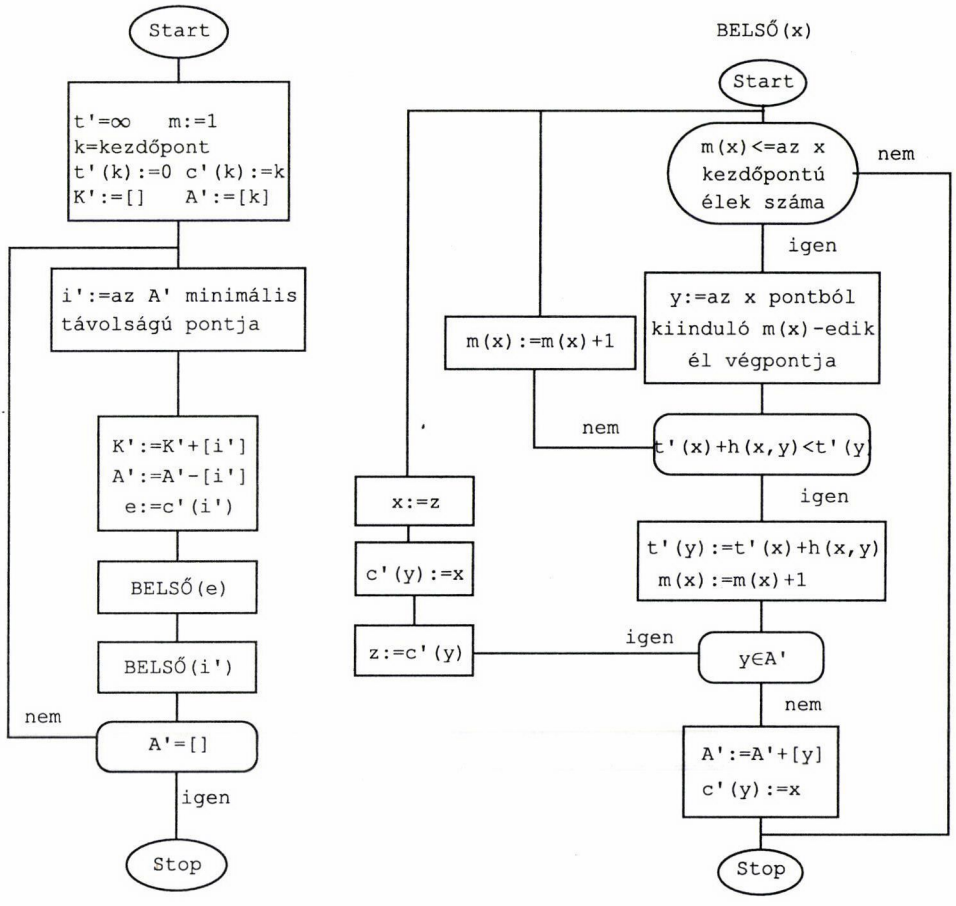
Az algoritmus első néhány lépését a példahálózaton a 3. ábra és a hozzá tartozó 3.1. táblázat mutatja be. Az algoritmust vázlatos diagramban is megadjuk (4. ábra) és a kulcsfontosságú belső eljárás működését külön is szemléltetjük (5. ábra).

3.1. táblázat

K: Kész A: Aktív C: Cimke T: Távolság																							
1	K	*a*																					
	A	*b*	*f*																				
		a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w
	C	*a*	*a*				*a*																
	T	0	10	~	~	~	10	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~	~
	M	3	1				1																
2	K	*a*	*b*																				
	A	*f*	*i*	*u*																			
		a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w
	C	*a*	*a*				*a*			*a*												*b*		
	T	0	10	~	~	~	10	~	~	10	~	~	~	~	~	~	~	~	~	~	~	25	~	~
	M	4	4				1			1												1		
3	K	*a*	*b*	*f*																			
	A	*i*	*s*	*g*	*u*																		
		a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w
	C	*a*	*a*				*a*	*f*		*a*										*a*		*b*		
	T	0	10	~	~	~	10	20	~	10	~	~	~	~	~	~	~	~	~	15	~	25	~	~
	M	–	4				–	1		1										1		1		
4	K	*a*	*b*	*f*	*i*																		
	A	*s*	*g*	*u*	*q*																		
		a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w
	C	*a*	*a*				*a*	*f*		*a*								*i*		*a*		*b*		
	T	0	10	~	~	~	10	20	~	10	~	~	~	~	~	~	~	25	~	15	~	25	~	~
	M	–	4				–	1		4								1		1		1		
5	K	*a*	*b*	*f*	*i*	*s*																	
	A	*g*	*u*	*q*	*r*																		
		a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w
	C	*a*	*a*				*a*	*f*		*a*								*i*	*s*	*a*		*b*		
	T	0	10	~	~	~	10	20	~	10	~	~	~	~	~	~	~	25	35	15	~	25	~	~
	M	–	4				–	1		4								1	1	3		1		

3.3. Helyesség

A módosított algoritmus *helyességét* illetően bebizonyítjuk, hogy *minimális fát* állít elő. A bizonyítást több lépésben végezzük. Először, az 5. ábra jelöléseit hasz-



4. ábra

nálva megvizsgáljuk az aktivitás halmazt új elemmel (y_u) bővítő valamint már aktív pontok (y_l, $l = 1, \ldots, u-1$) távolságait csökkentő belső eljárást.

Az egy pontból kiinduló élek rendezettségéből és az eljárás működési logikájából következően, alulvonással az eljárás végrehajtása előtti, felülvonással az utána kialakult értékeket jelölve, igazak az alábbi összefüggések:

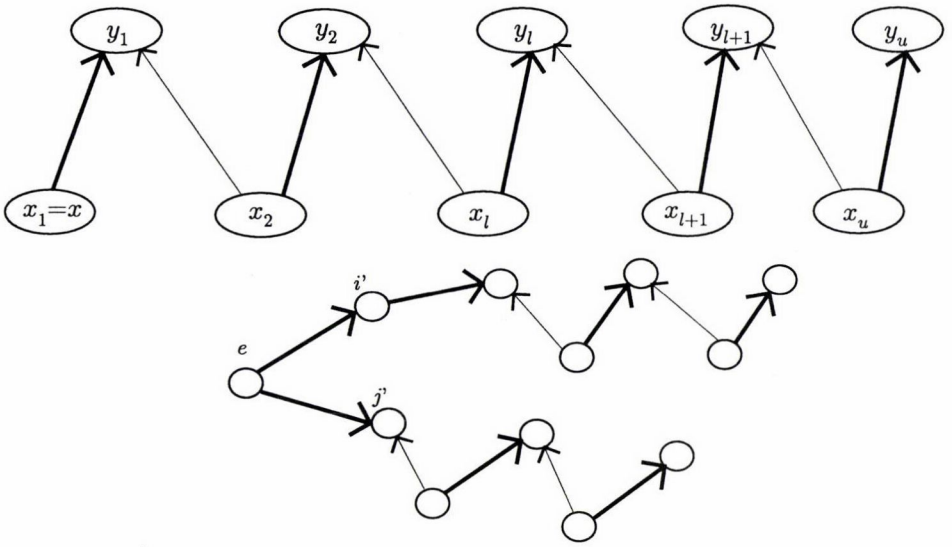
$$\underline{c'(y_1)} = x_2, \ldots, \underline{c'(y_1)} = x_{l+1}, \ldots \tag{3.1}$$

$$\underline{t'(y_1)} = \underline{t'(x_2)} + h(x_2, y_1), \ldots, \underline{t'(y_l)} = \underline{t'(x_{l+1})} + h(x_{l+1}, y_l), \ldots$$

$$h(x_2, y_1) \leq h(x_2, y_2) \ldots, h(x_{l+1}, y_l) \leq h(x_{l+1}, y_{l+1}), \ldots$$

$$\overline{c'(y_1)} = x_1, \ldots, \overline{c'(y_l)} = x_l, \ldots, \overline{c'(y_{l+1})} = x_{l+1}, \ldots$$

$$\overline{t'(y_1)} = \overline{t'(x_1)} + h(x_1, y_1), \ldots, \overline{t'(y_l)} = \overline{t'(x_l)} + h(x_l, y_l), \ldots$$



5. ábra

Vizsgáljuk meg az y_l és y_{l+1}, $l = 1, \ldots, u-1$ pontok új távolságainak viszonyát. Az y_l átcímkézése a

$$\overline{t'(y_1)} = \overline{t'(x_1)} + h(x_l, y_l) < \underline{t'(x_{l+1})} + h(x_{l+1}, y_l) = \underline{t'(y_l)} \tag{3.2}$$

reláció miatt történhetett csak. Az y_{l+1} új távolsága:

$$\overline{t'(y_{l+1})} = \overline{t'(x_{l+1})} + h(x_{l+1}, y_{l+1}), \ldots .$$

Mivel az y_{l+1} átcímkézése csak x_{l+1} mutatójának növelése után hajtódhatott végre, az y_l, y_{l+1}, x_{l+1} pontok mind különbözők, vagyis: $\underline{t'(x_{l+1})} = \overline{t'(x_{l+1})}$. Ebből, valamint az (3.1) és (3.2) egyenlőtlenségből adódik:

$$\overline{t'(y_l)} < \overline{t'(y_{l+1})} \qquad l = 1, \ldots, u-1) \tag{3.3}$$

vagyis:

3.1. ÁLLÍTÁS. *A belső eljárás folyamán módosított távolságok a módosítási sorrendben monoton nőnek.*

Mivel az algoritmus egy **b′–c′** lépésében a belső eljárás a minimumpozicióban lévő i', valamint az ennek címkéjeként szereplő e pontra hívódik meg, és az élek rendezettsége miatt $h(e, i') \leq h(e, j')$ (5. ábra), az 3.1. Állításból következik:

3.2. ÁLLÍTÁS. *Az algoritmus egy lépésének* (**b′–c′**) *végrehajtása után az aktivitás halmazban lévő pontok távolsága nagyobb vagy egyenlő mint a lépésben a kész halmazba bevont pont távolsága.*

Az előzőek segítségével bizonyítani tudjuk:

3.3. ÁLLÍTÁS. *A kész halmazba bekerülő pontok távolsága a bekerülés sorrendjében nem csökkenhet. Azonos távolság esetén a kisebb sorszámú pont kerül be előbb.*

Az állítás első része közvetlenül következik az 3.2. Állításból. A második rész igazolásához tegyük fel az ellenkezőjét és legyen a $p \in K'$, $q \in K'$, $t'(p) = t'(q)$, $p < q$ de a q a p-t megelőzően került K'-be. Feltehető, hogy ez közvetlenül a p bekerülése előtt történt. Az algoritmus $\mathbf{b}'$ lépésének definíciója miatt ez csak úgy lehet, hogy a q bekerülése előtt $t'(q) < t'(p)$ volt, és a q bekerülésekor rövidült le a p távolsága. Az (3.3) képletekből és az 3.2. Állításból következően a $t'(p)$ legfeljebb úgy rövidülhet le $t'(q)$-ra, ha (5. ábra) $q = i'$, $p = j'$ és $h(e, i') = h(e, j')$. Ez viszont ellentmond az egy pontból kiinduló élek definiált rendezettségének.

A 3.3. Állításból következik:

3.4. ÁLLÍTÁS. *A kész halmazba bekerült pontok már végleges távolsággal és címkével rendelkeznek, nem kerülhetnek vissza az aktivitás halmazba.*

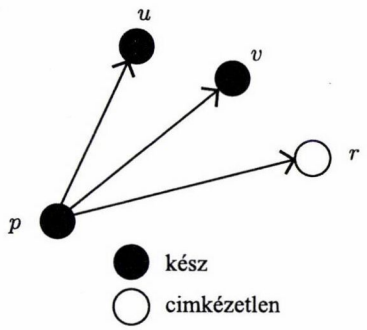
Be kell még bizonyítanunk azt is, hogy elérünk minden pontot, vagyis:

3.5. ÁLLÍTÁS. *Ha az aktivitás halmaz üressé vált, akkor minden, a kezdőpontból elérhető pont rendelkezik véges távolsággal és címkével.*

Tegyük fel az ellenkezőjét. Ha van ilyen pont, akkor van ezek között olyan, amelyik egy már kész pontból kiinduló él végpontja és ezek között az élek rendezettségében az első ilyen. A 6. ábra jelöléseivel: r az első címkézetlen pont, a megfelelő kész pont a p. Az r nem lehet a p első éle, mert akkor a p kész állapotba kerülésekor vizsgált és címkézett lett volna. Az r pontot a p-nél megelőző $u, \ldots, v$ pontok az r definíciója miatt már mind kész pontok. A p kész állapotba lépésekor ezek közül legalább egy p címkét kapott (hiszen, ha már mind rövidebb lett volna, akkor a belső eljárásban a rövidítésekkel eljutottunk volna az r-hez). Jelölje az élek rendezettségében az utolsó ilyet u. Tegyük fel, hogy az u jelenlegi címkéje is p. Ekkor viszont az u kész állapotba lépésekor, a követő pontok vizsgálatánál vagy eljutottunk volna az r-hez, vagy egy u-t követő és r-t megelőző pont vette volna fel a p címkét, ami ellentmond az eddigi feltételezéseknek illetve az u definíciójának. Tegyük fel, hogy az u jelenlegi címkéje már nem p. Ekkor viszont a címke áthelyezésekor lépett volna fel ugyanez a követő pont vizsgálat, ami ugyanezt az ellentmondást okozza. Minden esetben ellentmondásra jutunk, következésképpen az r pont nem létezik, így az 3.5. Állítás igaz.

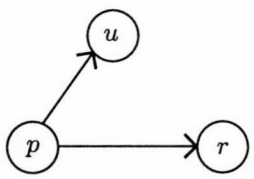
3.6. ÁLLÍTÁS. *A módosított algoritmus végrehajtása folyamán a pontok címkéi csak kész pontok lehetnek.*

Az állítás a kezdőállapotra vonatkozóan nyilván igaz. Tegyük fel, hogy van olyan pont, amelyre az állítás hamis lesz. Legyen a végrehajtás folyamatában az első ilyen az r pont (7. ábra), címkéje a p aktív pont. Az r nem kaphatta a p címkét úgy, hogy p került volna minimumhelyzetbe, mert akkor p már kész lenne és nem



6. ábra

lehet újra aktív (3.4. Állítás). A másik lehetséges mód az lenne, hogy egy az r-t a p-nél megelőző u pont címkéje volt a p és ennek megváltozásánál jutottunk el a p mutatójának növelésével az r-ig. Ez ismét ellentmond az r definíciójának, hiszen akkor az u lenne az első ilyen pont, lévén hogy a p nem lehetett még kész pont (3.4. Állítás).



7. ábra

3.7. ÁLLÍTÁS. *Ha a $p \in K'$ és az u egy p kezdőpontú él végpontja és a p mutatója az u-t már túllépte, akkor az aktuális távolságokra nézve igaz, hogy $t'(u) \leq t'(p) + h(p,u)$ és $u \in K'$ vagy $u \in A'$.*

Nyilvánvaló, hogy az azonos pontokhoz tartozó t' értékek az algoritmus előrehaladtával csak csökkenhetnek. Mivel a mutató már túllépte az u-t, szükségképpen megtörtént a $t'(u)$ és a $t'(p) + h(p,u)$ összehasonlítása és ha ennél $t'(u) > t'(p) + h(p,u)$ volt, a $t'(u) = t'(p) + h(p,u)$ rövidítés. Ez a p kész állapotba kerülésekor, vagy az után mehetett végbe a $t'(p)$ már nem változó értéke mellett (3.4. Állítás), ez után a $t'(u)$ már csak csökkenhetett. Mivel $t'(u)$ véges, igaz az is, hogy $u \in K'$ vagy $u \in A'$.

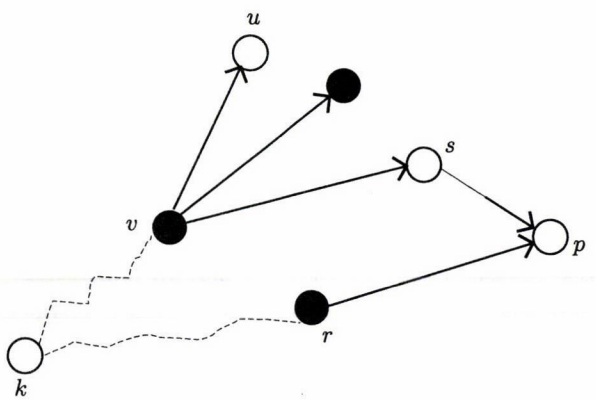
3.8. ÁLLÍTÁS. *Ha a $i \in K'$, akkor a hálózatban nincs a kezdőpontból az i-be vezető, a $t'(i)$-nél rövidebb út.*

Az 3.4. Állításból következően elegendő bizonyítani, hogy az i pont kész állapotba kerülésekor a minimumtulajdonság fennáll. Tegyük fel, hogy van olyan pont, amelyre ez nem igaz, jelölje a végrehajtás folyamán az első ilyet p (8. ábra). Tehát $c'(p) = r$, és $t'(p) = t'(r) + h(r,p)$ az aktív pontok között minimális, és létezik egy

olyan út a kezdőpontból $(\ldots, v, \ldots, s, p)$, amelynek w hosszára nézve igaz, hogy $w < t'(p)$. Mivel a p az első ilyen pont, az út nem az r-en keresztül vezet. Jelölje az út utolsó a p-t megelőző K'-beli pontját v.

Ha a v mutatója az s-et már túllépte, akkor $s \in A'$ és $t'(s) \leq t'(v) + h(v, s)$ (3.7. Állítás). Viszont a $v \in K'$ és a p definíciója miatt a $k - v$ út hossza $\geq t'(v)$, így $t'(p) > w \geq t'(v) + h(v, s) \geq t'(s)$ ami viszont ellentmond a $t'(p)$ minimumtulajdonságának.

Ha a v mutatója az s-et még nem lépte túl, akkor ha van olyan az s-et a v-nél megelőző aktív u pont (8. ábra), amelyet túllépett, akkor $t'(v) + h(v, s) \geq t'(v) + h(v, u) \geq t'(u)$ miatt van az ellentmondás. Ha viszont minden megelőző pont kész lenne, akkor az s-et közvetlenül megelőző is, és legkésőbb ennek kész állapotba kerülésekor sor került volna az s vizsgálatára, tehát a v mutatója az s-et már túllépte volna. Következésképpen a $t'(p)$ minimumtulajdonsága fennáll.



8. ábra

Ezzel (3.5. és 3.8. Állítás) *bebizonyítottuk, hogy az algoritmus helyes, vagyis egy minimális fát állít elő.*

4. Hatékonyság

Az algoritmus végrehajtása folyamán — az alapeljárással megegyezően — minden él pontosan egyszer kerül vizsgálatra (**c**′ lépés). Ez a tény az m mutató kezeléséből közvetlenül adódik. Vizsgáljuk meg az ilyen típusú algoritmusok műveletigényének meghatározó tényezőjét az aktivitás halmazok méretét, illetve ezek viszonyát az alapeljárást és az általunk definiált algoritmust illetően. Be fogjuk bizonyítani hogy a módosított algoritmus aktivitás halmaza minden lépésben szűkebb mint az eredetié. Az algoritmusok egy lépésén értjük a minimum-kiválasztástól az aktivitás halmaz ürességének ellenőrzéséig terjedő egy ciklusmenetet (**b**–**d** illetve

b'–d' szakaszok). A rövidség kedvéért az eredeti alapeljárást **D**-vel, az általunk definiált módosítottat **D'**-vel jelöljük.

4.1. ÁLLÍTÁS. *A két algoritmus minden lépésében:*
— *A két algoritmus ugyanazt a pontot választja ki és viszi át a kész halmazba, ugyanazon távolságadattal.*
A **D'** *algoritmus aktivitás halmaza része a* **D** *aktivitás halmazának.*

A bizonyítást teljes indukcióval végezzük. Nyilvánvaló, hogy az első lépés után az állítások igazak. Tételezzük fel, hogy egy adott $n \leq N$ sorszámú lépés előtt az állítások fennálltak, bebizonyítjuk, hogy az n sorszámú lépés után is fennállnak. Előrebocsátjuk, hogy a 3.5. Állítás és az indukciós feltevés miatt a lépés előtt sem az A' sem az A halmaz nem lehet üres. Először bebizonyítjuk, hogy a kiválasztott pontok a másik algoritmus aktivitás halmazában is szerepelnek, ugyanazzal a távolsággal és címkével.

Jelölje a **D**-ben választott pontot i, címkéjét $e = c(i)$. Mivel $e \in K$, így $e \in K'$ és a $t(e) = t'(e)$ minimális. Először bebizonyítjuk, hogy $i \in A'$ és $t'(i) = t(i)$. Ha a **D'**-ben az e mutatója nem lépett túl az i-n, akkor van olyan $j \in A'$ pont, amely az e-nél megelőzi az i-t. (Ellenkező esetben, mivel az $e \in K'$, maga az $i \in K'$ lenne, ellentmondásban az indukciós feltétellel.) De akkor $j \in A$ és mivel $e \in K$ és a **D**-ben az e kész állapotba kerülésekor minden e kezdőpontú él vizsgált a rövidítés szempontjából, és $h(e,j) \leq h(e,i)$, így $t(j) \leq t(e) + h(e,j) \leq t(e) + h(e,i) = t(i)$. A **D**-ben az i kiválasztásának módja miatt ez csak úgy lehet, hogy $h(e,j) = h(e,i)$ és $i < j$, ellentmondásban az élek rendezettségével. Tehát a **D'**-ben az i mutatója szükségképpen túllépett az i-n. Ebből viszont (3.7. Állítás) következik, hogy $i \in A'$ és $t'(i) \leq t'(e) + h(e,i) = t(e) + h(e,i) = t(i)$. Mivel viszont a $t(i)$ mint a következő kész pont távolsága minimális a hálózaton (a **D** helyessége miatt), a $t(i) \leq t'(i)$, tehát $t'(i) = t(i)$.

Jelölje a **D'**-ben választott pontot i', címkéjét $e' = c'(i')$. Az indukciós feltételből következően $i' \in A$. Mivel $e' \in K'$, így $e' \in K$ és a **D**-ben az e' kész állapotba kerülésekor minden kimutató él vizsgált: $t(i') \leq t(e') + h(e',i') \leq t'(e') + h(e',i') = t'(i')$. Mivel viszont a $t'(i')$-re mint a következő kész pont távolságára fennáll a minimumtulajdonság (3.8. Állítás), a $t'(i') \leq t(i')$, tehát $t'(i') = t(i')$.

Tehát mind az i, mind az i' aktív mindkét algoritmusban, így a választások minimumtulajdonsága és az előzőek miatt: $t(i) \leq t(i') = t'(i') \leq t'(i) = t(i)$, vagyis $t(i) = t(i')$ és $t'(i) = t'(i')$. Mindkét algoritmusban azonos távolságérték mellett a kisebb sorszámú pont választódik, így $i = i'$, amivel a 4.1. Állítás első részét bebizonyítottuk.

Be kell még bizonyítani, hogy az aktivitás halmazokra az $A' \subset A$ viszony a lépés végrehajtása után is fennáll. Az $i = i'$ törlése mindkét halmazból a tartalmazást nem változtatja. Hajtsuk végre az új kész elem belépéséből következő rövidítési vizsgálatokat. Az A-ban van minden olyan pont, amely valamely kész pontból kiinduló él végpontja, de maga még nem kész. A kész halmazok azonossága, valamint a **D'** belső eljárásának működése (5. ábra) és a 3.6. Állítás következményeképpen

az A'-be is csak ilyen pontok léphetnek be, tehát a tartalmazás továbbra is fennáll. Ezzel a 4.1. Állítást bebizonyítottuk.

Hogy az A' halmazok elemszáma a megfelelő A halmazokéhoz viszonyítva mennyivel csökken (páronként, összesen vagy átlagosan) az a konkrét hálózattól és kezdőponttól függ, kvantitatív becslést adni itt nem célunk. Az élszámmal való összefüggés jellegére azonban következtethetünk az alábbi állításból következően:

4.2. ÁLLÍTÁS. *A* **D**′ *eljárásban egy kész pont legfeljebb egy aktív pont címkéje lehet.*

A bizonyításhoz jelölje minden $p \in K'$ pontra $a(p)$ azon A'-beli pontok számát, amelynek címkéje p. Vizsgáljuk az $a(p)$ változását az eljárás egy lépésében a 4. és 5. ábra jelöléseivel. Megállapíthatjuk, hogy a belső eljárásban (5. ábra), ahol a 3.4. és 3.6. Állításokból következően az $x_1, \ldots, x_u$ pontok kész, az $y_1, \ldots, y_u$ pontok pedig aktív pontok, az $a(x_2), \ldots, a(x_u)$ értékek nem változnak. Ha $x_1 = x = i'$, akkor $a(x)$ legfeljebb 1 lesz, hiszen az i' mint új kész pont eddig nem szerepelhetett címkeként. Ha $x_1 = x = e$, akkor $a(x)$ értéke legfeljebb 1-gyel nő, viszont a belső eljárás hívása előtt az $a(x) = a(e)$ érték 1-gyel mindenképpen csökkent az i' kész állapotba kerülése miatt. Így minden $p \in K'$ pontra igaz, hogy a kész állapotba kerülésekor felvett $a(p)$ érték, ami maximum 1, csak csökkenhet az eljárásban.

Az alapeljárásra ez az állítás nyilvánvalóan nem áll, hiszen ott ha egy pont kész állapotban van, akkor kiinduló éleinek végpontjai közül mindazok aktívak, amelyek még nem készek. Így a 4.2 állításból következően az várható, hogy az egy pontból kiinduló élek átlagos számának növekedésével az A' átlagos elemszáma az A átlagos elemszámának egyre kisebb hányada lesz, tehát a különbség nő. Az alábbiakban néhány konkrét hálózattal és algoritmussal demonstráljuk ezt az állítást.

Hálózatok

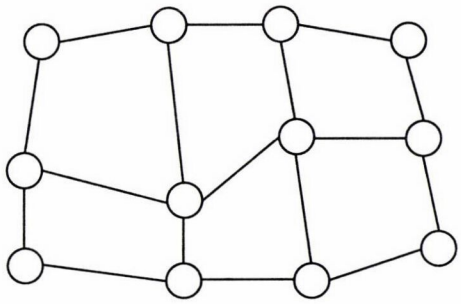
A hálózatokban az élhosszak az $[1, 10]$ intervallumban egyenletes eloszlású véletlenszám-generátorral lettek előállítva.

R2015, R3030: Négyzetrács jellegű (9. ábra) hálózatok, 20×15, illetve 30×30-as méretben.

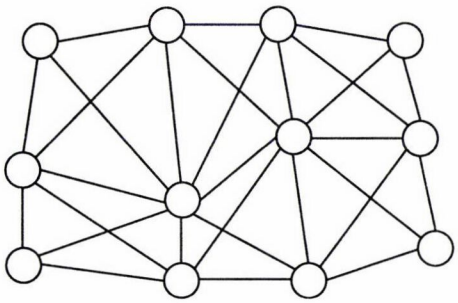
S2015, S3030: Átlós négyzetrács jellegű (10. ábra) hálózat, 20×15, illetve 30×30-as méretben.

V_3_4, V_3_8: 300 pontos, pontonként 4, illetve 8 kiinduló és végződő élet tartalmazó generált, szimmetrikus hálózat.

V_9_4, V_9_8: Mint az előzőek, de 900 ponttal.



9. ábra



10. ábra

Algoritmusok

DA: Alapeljárás az aktivitás halmazt ténylegesen halmazzal (2.1. típus) reprezentálva.

D'A: Módosított eljárás a **DA**-val megegyező halmazreprezentációval.

DB: Alapeljárás az aktivitás halmazt a távolság szerint növekvően rendezett listával (2.2. típus) reprezentálva.

D'B: Módosított eljárás a **DB**-vel megegyező halmazreprezentációval.

DC: Alapeljárás az aktivitás halmazt a távolság szerint növekvően részben rendezett kupaccal (2.3. típus) reprezentálva.

D'C: Módosított eljárás a **DC**-vel megegyező halmazreprezentációval.

A hálózatok minden kezdőpontjára előállítottuk a minimális fákat mindegyik algoritmussal. A kapott átlageredményeket a 4.1. táblázatban közöljük. A táblázat egyes sorainak részletes értelmezése:

Halmaz: Az *aktivitás halmazok* átlagos elemszáma pontonként (új kész pont kiválasztásánál).

Lépés: Egy fa meghatározásához szükséges *minimum-kiválasztási ill. besorolási és átsorolási lépések* átlagos száma.

%: A módosított eljárás adata az eredeti százalékában.

A táblázatban szándékosan nem közlünk futási idő értékeket. A futási idő, azonos szoftver — hardver környezet mellett is nagyon függ az algoritmus beprogramozási, kódolási módjától. Elvben jobb algoritmus, egy rosszabb hatásfokú kódolás mellett adhat gyengébb időeredményeket mint egy elvben rosszabb algoritmus egy jobb kódolás mellett.

A hatékonyság javulásának leginkább objektív mértéke az aktivitás halmaz méretének csökkenése (a táblázatban a „Halmaz %" sor), ami a példahálózatoknál, mint a táblázatban látható 18–38 százalékos mértékű. A halmazkezelés módjától függő mértékben ez a javulás jelentkezik áttételesen a besorolási és átsorolási lépésszám csökkenésében.

4.1. táblázat

Hálózat	R2015	R3030	S2015	S3030	V_3_4	V_9_4	V_3_8	V_9_8
Pont	300	900	300	900	300	900	300	900
Él	1130	3480	2194	6844	1200	3600	2400	7200
Él/Pont	3.66	3.86	7.31	7.60	4.00	4.00	8.00	8.00
DA,DB,DC Halmaz	22.39	36.35	36.48	57.03	75.43	224.96	113.54	337.38
D'A, D'B, D'C Halmaz	18.26	28.93	24.21	37.37	54.60	163.92	70.68	210.41
Halmaz %	**82**	**80**	**66**	**66**	**72**	**73**	**62**	**62**
DA Lépés	6719	32725	10944	51326	22629	202412	34063	303643
D'A Lépés	5478	26044	7263	33631	16380	147571	21204	189375
D'A Lépés %	**82**	**80**	**66**	**66**	**72**	**73**	**62**	**62**
DB Lépés	1939	9283	3704	17128	7311	67005	15046	136876
D'B Lépés	1711	8453	2736	12234	5437	50186	7498	66105
D'B Lépés %	**88**	**91**	**73**	**71**	**74**	**74**	**50**	**48**
DC Lépés	1727	5801	2053	6779	2184	8000	2491	8766
D'C Lépés	1620	5517	1803	5943	2033	7438	2175	7887
D'C Lépés %	**94**	**95**	**88**	**88**	**93**	**93**	**87**	**90**

A bemutatott elvi algoritmus több (2.2. és 2.3. típusú) gépi megvalósítása alkalmazásra került a szerző által készített közlekedési hálózatkezelő és tervező programrendszerekben, ilyen alkalmazások pl. a [6–9] munkák.

IRODALOM

[1] A. V. Aho, J. E. Hopcroft, J. D. Ullman, *Számítógép-algoritmusok tervezése és analízise,* Műszaki Könyvkiadó (Budapest, 1982).

[2] A. Bakó, Forgalomelosztás megoldása számítógéppel, *Alkalmazott Matematikai Lapok,* **6** (1980), 351–392.

[3] N. Deo, C. Pang, Shortest-Path Algorithms Taxonomy and Annotation, *Networks,* **14** (1984), 275–323.

[4] E. V. Denardo, B. L. Fox, Shortest — Route Methods 1. Reaching, Pruning, and Buckets, *Operations Research,* **27** (1979), 161–186.

[5] R. Dial, F. Glover, D. Karney, D. Klingman, A Computational Analysis of Alternative Algorithms and Labeling Techniques for Finding Shortest Path Trees, *Networks,* **9** (1979), 215–248.

[6] *Ergänzende Verkehrsprognose für die Grenzüberschreitenden Strassen,* Ingenieurbüro Kriebernegg (Österreich, Graz, 1992).

[7] *Traffic survey, forecast and sensitivity analysis of tolls for evaluating the tender of Motorway M5 on Hungary,* Bouygues (France) – Bauconsult (Hungary, Győr) (1993).

[8] *Az M3 autópálya forgalmi és díjbevételi tanulmánya* Object (Budapest) – Bauconsult (Győr) (1996).

[9] *A magyarországi gyorsforgalmi úthálózaton lebonyolódó utazások rendszerességének és legfontosabb szokásjellemzőinek meghatározása,* Bauconsult (Győr) 1997.

(Beérkezett: 1999. május 13.)

MARTON LÁSZLÓ
SZÉCHENYI ISTVÁN FŐISKOLA
SZÁMÍTÁSTECHNIKA TANSZÉK
9026 GYŐR, HÉDERVÁRI U. 3.
E-mail: marton@rs1.szif.hu

A LABEL-SETTING ALGORITHM FOR CALCULATING SHORTEST PATH TREES IN SPARSE NETWORKS

LÁSZLÓ MARTON

Shortest path analysis is a major analytical component of numerous quantitative transportation and communication models. Because of this, a number of algorithms have been developed for finding the shortest paths from one node to all other nodes in large sparse directed networks. This paper presents a labeling technique, an implementation of the general label-setting method. The study shows the correctness and the efficiency of the proposed method. In addition, we present computational results for random networks.

Alkalmazott Matematikai Lapok **19** (1999) 133–153

HEURISZTIKUS MÓDSZEREK A RELAXÁLT KÉTDIMENZIÓS TÉGLALAPPAKOLÁSI FELADATRA

DÓSA GYÖRGY

Veszprém

Egy ütemezéselméleti problémával, munkáknak egy erőforrással korlátozott ütemezésével foglalkozunk: Hogyan osszunk szét n munkát egy m hosszúságú időintervallumon úgy, hogy a felhasznált erőforrás maximuma minimális legyen. Speciális esetek az egydimenziós ütemezéselmélet feladata, a párhuzamos gépek ütemezése [13–15], illetve az egydimenziós ládapakolási feladat [3]. Több heurisztikus módszert vezetünk be, hatékonyságbecsléseket adunk, valamint kísérleti eredményekkel igazoljuk, hogy az új algoritmusok sok esetben a korábbiaknál jobb megoldást adnak.

1. Téglalappakolási feladatok

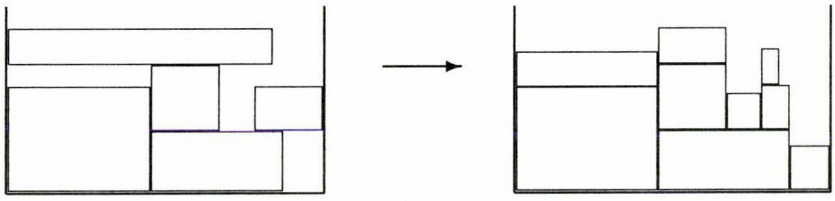
Az egydimenziós ládapakolási probléma a következő: Adott n darab tárgy, melyeknek súlyai $l_1, l_2, \ldots, l_n$, valamint olyan „ládák", amelyekbe m tömegű tárgy fér bele, (feltesszük, hogy $l_j \leq m$, $j = 1, \ldots, n$). Kérdés, hogy mennyi azon ládák minimális száma, amelyekbe az összes tárgy belefér. A feladatot átfogalmazhatjuk a következőképpen: Egy m egység szélességű, alul zárt, felül nyitott, téglalap alakú sávon belül kell elhelyeznünk n darab téglalapot, amelyeknek a szélessége l_j, és mindegyik magassága 1 egység. A téglalapokat úgy kell elhelyezni, hogy az l_j hosszú oldalaik a sáv alaplapjával párhuzamosak legyenek, és ne fedjék egymást. A kérdés az, hogy a téglalapoknak az előbbi módon való elhelyezéséhez minimálisan hány egységnyi magasságú sáv szükséges.

A ládapakolási feladat kombinatorikai értelemben vett duálja az ún. azonos párhuzamos berendezések ütemezésének problémája, ahol n számú munkát kell elvégeznünk m egyforma gép segítségével, ahol a j-edik munka elvégzésének időtartama bármely gép esetében w_j időegység ($j = 1, \ldots, n$). A munkák nem szakíthatók meg, vagyis amelyik munkát valamelyik gép elkezdi, azt be is kell fejeznie, másrészt nincs állásidő, vagyis ha egy munkát egy gép befejez, akkor azonnal elkezdheti a következő munkát. A feladat: Adjuk meg a munkáknak olyan ütemezését, hogy a legkésőbben befejeződő munka a lehető legkorábban fejeződjön be, vagyis a teljes átfutási idő minimális legyen. Ezt a következőképpen tudjuk átfogalmazni: Adott

egy m egység szélességű alul zárt, felül nyitott sáv. Helyezzünk el ebben n darab, 1 egység szélességű és w_j $(j = 1, \ldots, n)$ egység magasságú téglalapot úgy, hogy ne fedjék egymást, és a felhasznált maximális sávmagasság legyen minimális.

A kétdimenziós téglalappakolási feladatot az egydimenziós átfogalmazása alapján tudjuk megadni: Adott n darab $l_j \times w_j$ méretű téglalap $(j = 1, \ldots, n)$, ahol az első méret a vízszintes, a második a függőleges irányú kiterjedés, ezeket kell elhelyezni az m egység szélességű sávon úgy, hogy ne fedjék egymást, az oldalaik a sáv oldalaival párhuzamosak legyenek, és a felhasznált magasság minimális legyen. A téglalapok forgatását nem engedjük meg, számos gyakorlati alkalmazásban ugyanis a két méret különböző dolgot jelent. (Természetes feltételként $l_j \leq m$ most is teljesül minden j-re.) Ennek a feladatnak a duálja az a feladat, amikor adott a sáv magassága, és a felhasznált szélességet minimalizáljuk, azonban könnyen látható, hogy a téglalapok szélességének és magasságának a felcserélésével éppen az előző, kétdimenziós ládapakolási feladathoz jutunk.)

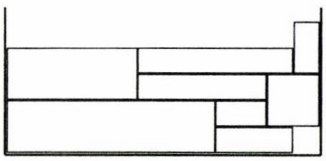
Végül a kétdimenziós téglalappakolási feladatnak egyfajta relaxációjával foglalkozunk: Képzeljük el, hogy az adott n darab $l_j \times w_j$ $(j = 1, \ldots, n)$ méretű téglalapot az m egység szélességű sávon már elhelyeztük, továbbra is úgy, hogy az oldalaik a sáv oldalaival párhuzamosak legyenek és ne fedjék egymást; de most megengedjük, hogy a téglalapok azon részei, ahol nem érintkezik másik téglalap „tetejével" függőlegesen „leessenek", amin azt értjük, hogy eltoljuk ezeket a részeket a sáv aljának az irányába mindaddig, amíg egy másik letett téglalap tetejébe, vagy a sáv aljába ütköznek. (1. ábra) A téglalapok előbb említett részeinek ilyen jellegű elmozdítását a téglalapok lefelé igazításának hívjuk. Kérdés, hogy így mennyi a minimálisan felhasználandó magasság a téglalapok elhelyezéséhez. Jelöljük a későbbiekben ezt a feladatot R2D-vel (relaxált kétdimenziós feladat).



1. ábra

Az R2D feladat felfogható úgy is, mint egy egyetlen erőforrással korlátozott ütemezési feladat: A téglalapok egy-egy munkát jelentenek, a j-edik munka l_j ideig tart, és az egyetlen erőforrásból w_j egységnyit vesz igénybe. Az erőforrásból minden pillanatban bármekkora mennyiség rendelkezésre áll, és a munkákat egy m egységnyi időhorizonton belül kell elvégezni. A kérdés az, hogy melyik munkát mikor kezdjük el, ha azt akarjuk, hogy az egy-egy időpontban felhasznált erőforrásmennyiség maximuma minimális legyen az időintervallumra vonatkozólag. Ez a relaxált feladat merül fel, amikor erőművek karbantartásakor a minimális tartalékkapacitást maximalizáljuk, mint a biztonság mértékét ([11], és [12]). Az alábbi

([10]-ből származó) 2. ábra azt a gyakran előforduló esetet illusztrálja, amikor a kétdimenziós, illetve relaxált feladat optimális megoldásának az értéke különböző. A sávszélesség 12. 8 téglalapot kell elhelyezni, amelyek méretei: $(8,2)$, $(5,2)$, $(3,1)$, $(2,1)$, $(5,1)$, $(6,1)$, $(2,2)$ és $(1,2)$. A relaxált feladat esetén ezek elférnek 4 egységnyi magasságon belül, viszont kétdimenziós esetben ez a magasság nem elég, az utolsó téglalap nem fér el. Ezek után, ha ezt külön nem mondjuk, csak az R2D feladattal foglalkozunk.



2. ábra

1.1. Jelölések

Legyen $\mathcal{R} = \{r_j(l_j, w_j), j = 1 \ldots n\}$ az adott n darab, $l_j \times w_j$ méretű téglalapból álló halmaz. Ezeknek megfelelően a j-edik munka elvégzésének időtartama l_j, az elvégzéséhez szükséges (időben állandó mennyiségű) erőforrás nagysága w_j. Feltesszük, hogy $l_j \leq m$, ahol m a sáv szélessége, valamint $\sum_{j=1}^{n} l_j > m$, különben a munkákat egymás után is el lehetne végezni. A munkák egy ütemezésén az $\mathcal{R}$ halmaz valamely $\mathcal{P} = \{P_0, P_1, \ldots, P_{m-1}\}$ partícióját értjük, ahol P_i azon téglalapok halmaza, amelyeknek a bal széle a sáv bal szélétől i egység távolságra van, vagyis azon munkák halmaza, amelyeket az i-edik időpontban kezdünk el elvégezni. A munkák sorrendje egy-egy rögzített időpontban az ütemezési feladat szempontjából közömbös. Legyen

$$S_i = \{r_\alpha | r_\alpha \in P_t,\ t \leq i < t + l_\alpha\}, \tag{1}$$

vagyis S_i azon munkák halmaza, amelyek már elkezdődtek, (vagy éppen elkezdődnek), de még nem fejeződtek be az i-edik időpontban. Legyen

$$W_i = \sum_{r_\alpha \in S_i} w_\alpha, \quad i = 0, \ldots m-1 \tag{2}$$

vagyis W_i az i-edik időpontban felhasznált erőforrás mennyisége. Ekkor a $\mathcal{P}$ ütemezés erőforrás-igényét a R2D ütemezés tulajdonságainak következtében így definiálhatjuk:

$$C(\mathcal{P}) = \max_{0 \leq i \leq m-1} W_i \tag{3}$$

A $\mathcal{P}^*$ ütemezés optimális, ha teljesül $C(\mathcal{P}^*) \leq C(\mathcal{P})$ az $\mathcal{R}$ halmaz tetszőleges $\mathcal{P}$ ütemezése esetén. Mivel véges sokféleképpen tudjuk az $\mathcal{R}$ halmaz elemeit m részre partícionálni, ilyen optimális ütemezés biztosan létezik, esetleg több is lehet.

1.2. Az optimumértékek viszonya

A $C(\mathcal{P}^*)$ értéket jelöljük C_{R2D}-del, ami tehát csak $\mathcal{R}$-től és az m számtól függ. Ha ugyanezzel az $\mathcal{R}$ téglalaphalmazzal és m sávszélességgel a kétdimenziós feladatot oldjuk meg, akkor a téglalapok elhelyezéséhez minimálisan szükséges magasságot, vagyis az optimális megoldást jelölje C_2. Annak az egydimenziós párhuzamos gépek ütemezési feladatnak az optimális megoldását jelölje C^p, amelyet úgy kapunk, hogy a téglalapokat függőlegesen l_j darab $1 \times w_j$ méretű szeletre felvágjuk, valamint annak az egydimenziós ládapakolási feladatnak az optimális megoldását jelölje C^b, amelyet úgy kapunk, hogy a téglalapokat vízszintesen w_j darab $l_j \times 1$ méretű szeletekre felvágjuk.

1. TÉTEL. *Jelölje C^p és C^b a megfelelő párhuzamos gépekre vonatkozó illete ládapakolási feladatok, C_2 és C_{R2D} pedig a kétdimenziós és relaxált kétdimenziós feladatok optimális megoldásait. Ekkor ezen értékek között a következő egyenlőtlenségek teljesülnek:*

$$\max\{C^p, C^b\} \leq C_{R2D} \leq C_2 \tag{4}$$

Bizonyítás. Csak a $C^b \leq C_{R2D}$ állítás nem triviális. Tekintsünk egy optimális R2D ütemezést. Ekkor bármelyik időpontban a felhasznált erőforrás nagysága legfeljebb C_{R2D}. A téglalapok vízszintes irányban történő elvágása után ez azt jelenti, hogy bármelyik időpontban legföljebb C_{R2D} számú munka van folyamatban. Ezért a $t_0 = 0$ időponttól kezdve folytatólagosan az éppen akkor kezdődő munkák mindig beletehetők legfeljebb C_{R2D} számú ládába. □

Ezután a feladatokra ütemezési feladatként is, és téglalap-pakolási feladatként is fogunk hivatkozni. Mivel az optimális ütemezés(ek) megkeresésének feladata már egydimenziós esetben is NP-teljes, gyakran heurisztikus módszerekkel próbálják az előbbi feladatokat megoldani. Egy heurisztikus megoldás értékelésében nagy segítséget jelent, ha „elég jó" alsó becsléssel rendelkezünk az optimum értékét illetően: így meg tudjuk becsülni, hogy a heurisztikus módszer által szolgáltatott megoldás mennyire közelíti meg az optimumot, ugyanis az természetesen az alsó becslés és a heurisztikus megoldás értéke között van. Másrészt sok pontos megoldást adó (nem polinomiális lépésszámú) algoritmus pontosan egy alsó becslés, és egy heurisztikus megoldás generálásával kezdődik, és akkor működik „gyorsan", ha az optimumhoz eléggé közeliek ezek az alsó, illetve felső becslések.

A következő fejezetben alsó becslésekkel foglalkozunk, a 3. fejezetben bevezetünk néhány heurisztikus módszert, amelyek korábbi kétdimenziós heurisztikáknak az R2D téglalap-pakolási feladatra való alkalmazásai, valamint egy új algoritmust. E két fejezetben becsléseket adunk arra, hogy az egyes alsó becslések a legrosszabb esetben „milyen messze" lehetnek az optimumtól, illetve a heurisztikus megoldások a legrosszabb esetben legföljebb hányszorosai lehetnek az optimumnak. Érdemes megjegyezni, hogy a legrosszabb esetben kapott arányoknál a gyakorlatban sokkal jobb (1-hez közeli) arányt kapunk. A negyedik fejezet numerikus eredményeket taralmaz: néhány ezer véletlenszerűen generált feladat esetében például kiszámít-

juk a heurisztikus megoldás/alsó becslés törtek átlagát. Ez az arány alig valamivel nagyobb 1-nél, (persze különböző feladatosztályok, heurisztikák illetve becslések esetén más és más), ami azt mutatja hogy az alsó becslések is és a heurisztikus megoldások is nagyon jók: közel vannak egymáshoz, így az optimumhoz is. Még egyszer jegyezzük meg, hogy amiatt vagyunk kénytelenek ezt a kerülő utat alkalmazni, vagyis azt, hogy az alsó becslések „jóságát" a heurisztikus megoldáshoz való közelségükkel igazoljuk, és viszont, mert a pontos megoldás értékét általában nem ismerjük.

2. Alsó becslések

2.1. Néhány alsó becslés

Az ebben a fejezetben közölt alsó becslések érvényesek az $R2D$, és a kétdimenziós feladatra is. Legyen L egy tetszőleges alsó becslés, vagyis legyen tetszőleges $\mathcal{R}$ téglalaphalmaz esetén $L(\mathcal{R}) \leq C_{R2D}(\mathcal{R})$. Ekkor a $\frac{L(\mathcal{R})}{C_{R2D}(\mathcal{R})}$ törtek infimumát az L alsó becslés elméleti hatékonyságának nevezzük, és $H(L)$-lel jelöljük, vagyis

$$H(L) = \inf\left\{\frac{L(\mathcal{R})}{C_{R2D}(\mathcal{R})}\right\},$$

ahol $\mathcal{R}$ tetszőleges téglalaphalmaz. Graham 1966-os [8] dolgozatában közli azt az egydimenziós ütemezési feladatra vonatkozó alsó becslést, amely szerint a felhasznált sávmagasság legalább akkora, mint a legmagasabb téglalap magassága, másrészt legalább akkorra, mint a téglalapok összterületének m-ed része. Ez a becslés általánosítható:

2. TÉTEL. *Az R2D feladat esetében a felhasznált sávmagasság legalább akkora, mint a téglalapok magassága, valamint legalább akkorra, mint a téglalapok összterületének m-ed része:*

$$C_{R2D} \geq L_1 := \max\left\{\frac{\sum_{j=1}^{n} l_j \times w_j}{m},\ \max\{w_j,\ j = 1 \ldots n\}\right\} \tag{5}$$

Bizonyítás. Ha a téglalapokat teljesen egyenletesen sikerülne elosztani, akkor lenne az összterület m-edrésze a magasságuk, másrészt bármelyik téglalap magassága alsó korlát a felhasznált sávmagassághoz. □

Megjegyzés. Könnyen látható, hogy az előbbi maximum két tagját külön-külön véve az első tag elméleti hatékonysága $\frac{1}{m}$, a második tagé pedig 0.

3. TÉTEL. *Az L_1 alsó becslés elméleti hatékonysága legalább $\frac{1}{3}$, vagyis $H(L_1) \geq \frac{1}{3}$.*

Bizonyítás. Az [1]-ben szereplő $C_2 \leq \frac{2 \cdot T}{m} + w_{\max}$ becslésből, (ahol T a téglalapok össz-területe, $w_{\max}$ pedig a legnagyobb magasság) látható, hogy $C_2 \leq 3 \cdot L_1$, ebből $\frac{L_1}{C_{R2D}} \geq \frac{L_1}{C_2} \geq \frac{1}{3}$ adódik. □

Megjegyzés. [4] szerint speciálisan az egydimenziós esetben az L_1-nek megfelelő becslés elméleti hatékonysága pontosan $\frac{1}{2}$, ezek szerint a R2D és kétdimenziós pontos becslések értékei $H(L_1)$-re $\frac{1}{3}$ és $\frac{1}{2}$ közötti számok.

A következő tétel egy relaxációs elvet fejez ki:

4. TÉTEL. *Ha az eredeti feladat téglalapjai közül*
a, *valahányat elhagyunk,*
b, *egynek, vagy többnek az egyik (vagy mindkét) méretét csökkentjük,*
c, *akárhányat valamelyik oldalával párhuzamosan kettévágunk; akkor az új feladat C'-vel jelölt optimumértékére teljesül: $C' \leq C_{R2D}$.*

Bizonyítás. Az előbbi helyekre most is lepakolhatók a (megmaradt) téglalapok, így az erőforrás felhasználását egyik időpontban sem növeltük. □

5. TÉTEL. *Legyen $t_j = \left\lfloor \frac{l_j}{k_1} \right\rfloor$, $u_j = \left\lfloor \frac{w_j}{k_2} \right\rfloor$, $j = 1, \ldots, n$, ahol k_1 és k_2 tetszőleges rögzített egész számok, amelyekre teljesül, hogy $1 \leq k_1 \leq \max\{l_j, j = 1, \ldots, n\}$ és $1 \leq k_2 \leq \max\{w_j, j = 1, \ldots, n\}$. Ekkor*

$$C_{R2D} \geq L_2 := \left\lceil \frac{\sum_{j=1}^{n} t_j \cdot u_j}{\left\lfloor \frac{m}{k_1} \right\rfloor} \right\rceil \cdot k_2 \tag{6}$$

Bizonyítás. Helyettesítsük az $r_j(l_j \times w_j)$ téglalapokat az $s_j(t_j \cdot k_1 \times u_j \cdot k_2)$ téglalapokkal $(i = 1, \ldots, n)$. Így az előző tétel alapján az új feladat optimumértéke az előzőnél nem nagyobb. Az $s_j(t_j \cdot k_1 \times u_j \cdot k_2)$ téglalapokat vágjuk fel $\sum_{j=1}^{n} t_j \cdot u_j$ számú $s'(k_1 \times k_2)$ méretű (egyforma) téglalapra. Ezekből a sáv szélességében legfeljebb $\left\lfloor \frac{m}{k_1} \right\rfloor$ darab helyezhető el; így kell

$$\left\lceil \frac{\sum_{j=1}^{n} t_j \cdot u_j}{\left\lfloor \frac{m}{k_1} \right\rfloor} \right\rceil$$

sor az elhelyezésükhöz. Minden sorban a téglalapok magassága k_2, amiből adódik az állítás. □

Megjegyzés. Az előbbi L_2 alsó becslés erősebb becslés L_1-nél, ugyanis $k_1 = k_2 = 1$ esetén éppen az összterület adódik, $k_1 = 1$, $k_2 = \max\{w_j, j = 1 \ldots n\}$ esetén pedig a becslés értéke legalább k_2. Ezek szerint $H(L_2) \geq H(L_1)$. Másrészt például ha a téglalapok egyformák, akkor a becslés megegyezik az optimális megoldás értékével.

1. Példa. Legyen $m = 3$, $n = 2$, $R = \{(2,5),(2,11)\}$. Ekkor $L_1 = 11$, míg az L_2 becslés ennél lényegesen jobb alsó becslést szolgáltat: ha például $k_1 = 2$ és $k_2 =$

5, akkor a (6) kifejezés jobb oldalán álló becslés értéke 15. Ugyanezt az értéket szolgáltatja a $k_1 = 1$ és $k_2 = 5$ választás is. L_2 értéke 15, míg az optimális megoldás értéke 16, ugyanis a téglalapok nem férnek el egymás mellett, ezért magasságaik összeadódnak.

6. TÉTEL. *Rendezzük a téglalapokat magasságaik szerinti csökkenő sorrendbe. Legyen j_0 az az index, amelyre az első j_0 darab téglalap még befér egymás mellett a ládába, de a következő már nem, vagyis $\sum_{j=1}^{j_0} l_j \leq m$, $\sum_{j=1}^{j_0+1} l_j > m$. Ekkor*

$$C_{R2D} \geq L_3 := \max\{L_2, w_{j_0} + w_{j_0+1}\} \tag{7}$$

Bizonyítás. Ha nincs olyan időpont, amikor az első j_0 munka közül legalább kettő egyszerre tartana, akkor a $j_0 + 1$-edik téglalapnak megfelelő munkához van olyan időpont, amikor legalább egy, a sorrendben előtte levő munkával egyidőben folyamatban van. Ha pedig az első j_0 számú munka közül legalább kettő egyszerre folyamatban van, akkor ezek együttes időtartama is legalább $w_{j_0} + w_{j_0+1}$. □

Megjegyzés. Az egydimenziós feladat esetén $H(L_3) = \frac{2}{3}$, [4].

2.2. Egy speciális eset

Tegyük fel, hogy a munkáknak megfelelő téglalapok „elférnek két sorban", vagyis létezik $\mathcal{R}$-nek olyan $\mathcal{R}_1 \cup \mathcal{R}_2$ partíciója, amelyre $\sum_{r_j \in \mathcal{R}_i} l_j \leq m$, $i = 1, 2$. Továbbra is feltesszük azt, hogy $\sum_{j=1}^{n} l_j > m$, vagyis egy sor nem elég az elhelyezésükhöz. Belátható, hogy a R2D feladat ebben az esetben ekvivalens a kétdimenziós feladattal, valamint az is teljesül, hogy a feladat még mindig az NP-teljes feladatosztályba tartozik.

7. TÉTEL. *Az előbbi speciális esetben az L_1 alsó becslés elméleti hatékonysága $\frac{1}{2}$, vagyis $H(L_1) = \frac{1}{2}$.*

Bizonyítás. Mivel két sor elég a téglalapok elhelyezéséhez, az optimális megoldás értéke legfeljebb a $\max\{w_j, j = 1 \ldots n\}$ érték kétszerese lehet, ezért $H(L_1) \geq \frac{1}{2}$. Másrészt ha az $\mathcal{R}$ téglalaphalmaz a következő: $n = m + 1$, $\mathcal{R} = \{r_j(1 \times m), j = 1, \ldots, n\}$, akkor $L_1 = \max\left\{\frac{(m+1)m}{m}, m\right\} = m + 1$, az optimális megoldás értéke pedig $2m$. Így a $H(L_1) = \frac{1}{2}$ pontos érték adódik. □

8. TÉTEL. *Ha van olyan optimális megoldás, ahol bármelyik időpontban legfeljebb két munka van egyszerre folyamatban, akkor feltehető, hogy az egyik sorban lévő téglalapok sorrendje magasság szerinti csökkenő és balra vannak igazítva, a másik sorban lévőké pedig növekvő, és jobbra vannak igazítva.*

Bizonyítás. Azt mutatjuk meg, hogy az egyes sorokban levő téglalapok halmazán nem változtatva, csak más sorrendbe rakva őket a fenti a legjobb elrendezés. Teljes indukcióval bizonyítunk, a második sorban lévő téglalapok száma szerint. Ha ez a szám egy, akkor az állítás teljesül. Tegyük fel, hogy k-ra az állítás igaz. Helyezzük el optimálisan a téglalapokat két sorban, tegyük fel, hogy $k + 1$ téglalap

van a második sorban, és az előbbi elrendezésben vannak. Állítjuk, hogy ez az elrendezés nem javítható. A második sor legkisebb téglalapja legyen T, ez van a sor bal szélén. Hagyjuk el T-t. Ha a teljes átfutási idő nem csökkent, akkor az állítás az indukciós feltétel miatt igaz. Ha csökkent, akkor az előbb a teljes átfutási idő éppen a T téglalapnál vétetett föl. Legyen a T téglalap bal szélénél az alatta lévő sorban lévő téglalap T'. Az alsó-, és a felső sor között tetszőleges elhelyezés esetén $\sum_{r_j \in \mathcal{R}} l_j - m$ szélességű átfedés van, és akkor járunk a legjobban, ha ide mindkét sorból a legkisebb téglalapokat válogatjuk. (Ha valamelyik nem fér bele teljesen, akkor a beférő részét vesszük.) Ezen átfedésen belül a T' téglalap fölött legalább $w(T)$ magasságú téglalap van, vagyis az előbbi elrendezésnél nincs jobb. □

9. TÉTEL. *Tegyük fel, hogy a téglalapokat csökkenő magasság szerinti sorrendben elhelyezve is beférnek két sorba, vagyis létezik olyan j_0 index, amelyre $\sum_{j=1}^{j_0} l_j \le m$, $\sum_{j=j_0+1}^{n} l_j \le m$. Ekkor $H(L_3) = \frac{2}{3}$.*

Bizonyítás. Helyezzük el a téglalapokat az előbbi módon, de alternálva, vagyis az első sort balról jobbra, a következőt jobbról balra töltsük fel. Így kapunk egy közelítő megoldást, amelynek értéke legyen C'. A maximális felhasznált sávmagasság legfeljebb $w_1 + w_{j_0+1}$, mert az alsó sorban a bal oldali, a felső sorban a jobb oldali téglalapnak maximális a magassága, ezért $C' \le w_1 + w_{j_0+1}$. Két esetet különböztetünk meg. Ha $w_{j_0} + w_{j_0+1} \le w_1$, akkor $w_{j_0} \ge w_{j_0+1}$ miatt $w_{j_0+1} \le \frac{1}{2} w_1$, és így $C_2 \le C' \le w_1 + w_{j_0+1} \le \frac{3}{2} w_1 \le \frac{3}{2} L_1 \le \frac{3}{2} L_3$. A másik esetben $w_{j_0} + w_{j_0+1} > w_1$. Ekkor viszont $C_2 \le C' \le w_1 + w_{j_0+1} < w_{j_0} + w_{j_0+1} + w_{j_0+1} \le \frac{3}{2}(w_{j_0} + w_{j_0+1}) \le \frac{3}{2} L_3$. Ezek szerint $H(L_3) \ge \frac{2}{3}$. Másrészt legyen $n = m + 3$, az első $m - 1$ téglalap magassága legyen m, az utolsó három téglalap magassága legyen $\frac{m}{2}$, és valamennyiük szélessége legyen 1 egység. Ekkor $L_3 = \max\left\{m + \frac{1}{2}, \frac{m}{2} + \frac{m}{2}\right\} = m + \frac{1}{2}$, az optimális megoldás értéke pedig $\frac{3}{2}m$. Így $H(L_3) \le \frac{2}{3}$, vagyis a két eredményt összefoglalva $H(L_3) = \frac{2}{3}$. □

2. Példa. Legyen $m = 15$, $n = 5$, $R = \{(1,9),(4,8),(7,7),(2,6),(3,4)\}$. Ekkor $L_1 = L_2 = 9$, míg $L_3 = 10$.

2.3. Két további alsó becslés

Ebben a fejezetben feltesszük, hogy a téglalapok magasság szerinti csökkenő sorrendbe vannak rendezve, vagyis $w_1 \ge w_2 \ge \ldots \ge w_n$.

10. TÉTEL. *Legyen*

$$k = \left\lceil \frac{\sum_{r_j \in \mathcal{R}} l_j}{m} \right\rceil .$$

Ekkor C_{R2D} legalább akkora, mint a sorrendben az utolsó k számú téglalap magasságának az összege, vagyis

$$C_{R2D} \ge L_4 := w_{n-k+1} + \ldots + w_n. \tag{8}$$

Bizonyítás. A téglalapok elhelyezéséhez legalább k „sor" szükséges, vagyis lesz olyan időpont, amikor legalább k számú munka egyszerre folyamatban van. □

3. Példa. Legyen $m = 6$, $n = 4$, és álljon az R téglalaphalmaz két 3×3-as, egy 3×2-es és egy 4×3-as téglalapból. Ekkor $L_1 = L_2 = L_3 = 6$, míg $L_4 = 8$.

Az előző alsó becslés szerint ha mindegyik időpontban „elég sok" munka van egyszerre folyamatban, akkor így is adódik egy alsó korlát C_{R2D}-re. Készítsük el az $\mathcal{R}_1 = \{r_j(l_j \times 1),\ j = 1, \ldots, n\}$ téglalaphalmazt, amelyet tehát úgy kapunk, hogy az $\mathcal{R}$-beli téglalapok magasságát 1 egységre változtatjuk. Tekintsük a munkáknak egy tetszőleges ütemezését. Ha nincs olyan időpont, amikor egyszerre p-nél több munka lenne folyamatban, akkor az $\mathcal{R}_1$-beli téglalapok bepakolhatók p számú, m méretű ládába, vagy másképpen:

11. TÉTEL. *Ha az $\mathcal{R}_1$-beli téglalapok nem pakolhatók be p számú, m méretű ládába, akkor a munkáknak bármelyik ütemezése esetén van olyan időpont, amikor p-nél több munka van egyszerre folyamatban. Így*

$$C_{R2D} \geq w_{n-p} + w_{n-p+1} + \ldots + w_n. \qquad \square$$

4. Példa. Legyen $m = 9$, $n = 5$, $R = \{(4,3),(4,3),(4,3),(3,3),(3,3)\}$. Könnyen látható, hogy a téglalapok két sorban nem férnek el, ezért a 11. Tételbeli alsó becslés értéke 9, a korábbi alsó becslések értéke azonban csak 6. Az optimális megoldás értéke is 9.

2.4. Egydimenziós alsó becslések alkalmazása

Vágjuk el függőlegesen az $\mathcal{R}$-beli téglalapokat úgy, hogy a j-edik, $l_j \times w_j$ méretű téglalapból l_j darab $1 \times w_j$ méretű szelet legyen, a kapott téglalaphalmazt jelölje $\mathcal{R}^p$. Ekkor a relaxált feladat C^p optimális megoldására teljesül $C^p \leq C_{R2D}$, így a párhuzamos gépek ütemezése esetén kapott bármely C' alsó korlátra a R2D feladatra vonatkozóan is alsóbecslést kapunk. (Ugyanígy ezek az alsó becslések közvetlenül is alkalmazhatók, ha a munkák elvégzéséhez szükséges idők megegyeznek.) Néhány ilyen alsó becslés [13]-ban található. Jelöljük néhány, párhuzamos gépekre vonatkozó alsó becslést $P(i)[R^p, m]$-mel, ahol $i \in I$. (Itt $|I|$ számú alsó becslésről van szó.) Most vágjuk el vízszintes irányban az R-beli téglalapokat, a kapott téglalaphalmaz R^b, ennek elemszáma n_b, (ahol $n_b = \sum_{j=1}^{n} w_j$. A j-edik, $l_j \times w_j$ méretű téglalapból w_j darab $l_j \times 1$ méretű szelet lesz). Az így kapott feladatra az előbbi alsó becslésekből egy újabbat nyerhetünk:

12. TÉTEL. *Jelölje a vízszintes irányban elvágott R-beli téglalapokat R^b. Legyen $P(i)[R', m']$ néhány párhuzamos gépekre vonatkozó alsó becslés, ahol R' a téglalaphalmaz, és m' a sávszélesség. Ekkor teljesül a következő egyenlőtlenség, ahol C^b azt a legkisebb magasságot jelenti, amelyen belül R^b elemei elhelyezhetők.*

$$C^b \geq \min\left\{C \,\middle|\, \max_{i \in I} P(i)[R^b, C] \leq m\right\} \tag{9}$$

Bizonyítás. Ha $P(i)[R^b, C] > m$, akkor a téglalapok nem férnek el a C egység magas, m egység széles sávon, ezért a ládák száma C-nél több kell hogy legyen. □

(Ugyanez a módszer fordított sorrendben is működik, néhány ládapakolási alsó becslést alkalmazva, azokból párhuzamos gépekre vonatkozóakat kapunk, és ezek mindegyike alkalmazható az R2D feladathoz.)

3. Heurisztikus módszerek

3.1. A $\mathcal{BL}$ algoritmus

Az optimális ütemezés(ek) megkeresésének feladata már egydimenziós esetben is NP-teljes, ezért gyors, közel-optimális ütemezéseket adó algoritmusokat vezetünk be az R2D feladat megoldására. R. L. Graham 1966-ban közölte az LPT algoritmust (LPT = Longest Processing Time) egyforma párhuzamos gépek ütemezésére [8]. Graham algoritmusát általánosította [1] a kétdimenziós esetre, az általánosított algoritmust $\mathcal{BL}$ (=Bottom Left) algoritmusnak nevezve el. Ez a következőket végzi: Először a munkákat valamilyen sorrendbe rendezzük. A soron következő téglalapot úgy helyezzük el, hogy a lehető leglejjeb helyezkedjen el. Az így szóba jövő helyek közül a legelső időpontra ütemezük a téglalapot, vagyis balra igazítjuk. Ebből az algoritmusból természetes módon úgy kaphatunk R2D algoritmust, hogy a téglalapokat lefelé is igazítjuk. Az egyik lehetőség az, hogy az összes téglalapot elhelyezzük a kétdimenziós szabály szerint, és a legvégén igazítjuk a már elhelyezett téglalapokat lefelé. Egy másik lehetőség, hogy rögtön, egy-egy téglalapot közvetlenül az elhelyezése után lefelé igazítunk, és ezután helyezzük el a következő téglalapot a minimális magasságban balra igazítva, majd ezt is lefelé igazítjuk, és így tovább. Ez utóbbi algoritmus formális leírása a következő:

A R2D feladatra vonatkozó $\mathcal{BL}$ algoritmus

1. Legyen m az időintervallum hossza, $P_i = \emptyset$, $W_i = 0$ $(i = 0, \ldots, m-1)$.
2. Rendezzük a téglalapokat valamilyen sorrendbe, legyen $j := 1$.
3. Legyen $R(i) = \max_{i \le k \le i+l_j-1} W_k$, $(i = 0, \ldots, m - l_j)$.
   Legyen $i_0 = \arg\min \{R(i) : 0 \le i \le m - l_j\}$.
4. Legyen $P_{i_0} = P_{i_0} \cup \{r_j\}$, és legyen $W_k = W_k + w_j$ $(k = i_0, \ldots, i_0 + l_j - 1)$, vagyis ütemezzük a j-edik munkát az i_0-adik időpontra.
5. Legyen $j = j + 1$. $j \le n$ esetén menjünk a 3. lépésre, egyébként vége.

A kétdimenziós $\mathcal{BL}$ algoritmussal elméleti hatékonyságával kapcsolatban [1] a következő alapvető eredményeket közli: Legyen $C(\mathcal{BL})$ a kétdimenziós algoritmus által kapott megoldás értéke, C^* az optimális megoldás értéke. Ekkor a $\frac{C(\mathcal{BL})}{C^*}$ arány tetszőlegesen nagy lehet, ha a téglalapokat a szélességeik szerinti növekvő sorrendben, vagy pedig a magasságaik szerinti csökkenő sorrendben helyezzük el. A cikk ezen tételek bizonyítására egy-egy példát közöl, ahol az előbbi arány egy tetszőleges, előre magadott számnál nagyobb. Ezek a példák az R2D feladat, és

annak a $\mathcal{BL}$ algoritmussal történő megoldására is közvetlenül alkalmazhatók: Az előbbi sorrendek esetén a $\frac{C(\mathcal{BL})}{C_{R2D}}$ arány is tetszőlegesen nagy lehet, ugyanis az előbbi téglalapokat az R2D szabály szerint előbb elhelyezve, majd (akár rögtön ezután, akár a legvégén) lefelé igazítva, ugyanoda kell őket tenni, mint a kétdimenziós esetben, és ugyanazt a felhasznált sávmagasságot kapjuk, mint a kétdimenziós esetben. Az előzőek alapján érvényes a következő

13. TÉTEL. *Tetszőleges $M > 0$ valós számhoz létezik olyan $\mathcal{R}^1$ téglalaphalmaz, amelynek az elemeit a szélességeik szerinti növekvő sorrendben elhelyezve $\frac{C(\mathcal{BL})}{C_{R2D}} > M$ teljesül, hasonlóképpen létezik olyan $\mathcal{R}^2$ halmaz, amelynek az elemeit a magasságaik szerinti csökkenő sorrendben elhelyezve $\frac{C(\mathcal{BL})}{C_{R2D}} > M$ teljesül.* □

Más a helyzet, ha a sorrend a csökkenő szélesség szerinti. Ekkor kétdimenziós esetben a $\frac{C(\mathcal{BL})}{C^*}$ arány legrosszabb esetben 3 lehet, és ez a becslés éles. Ha a téglalapokat legvégül, valamennyiük elhelyezése után igazítjuk lefelé, akkor ugyanez az éles felső becslés adódik az R2D feladat esetében. Az [1]-beli bizonyítása viszont nem alkalmazható az R2D esetre akkor, ha közvetlenül az elhelyezésük után igazítjuk a téglalapokat lefelé. Ebben az esetben a $\frac{C(\mathcal{BL})}{C^*} \leq 4$ felső becslést bizonyítjuk a következő fejezetben, de sokkal általánosabb keretek között.

3.2. A $\mathcal{BL}$ algoritmus általánosításai

Nevezzük $\mathcal{B}$-nek a $\mathcal{BL}$ algoritmusnak azt a változatát, ahol a következő téglalapot a lehető leglejjebb helyezzük el, de balra igazítás nélkül. Ekkor a 3. pont a következőképpen változik:

3. Legyen $R(i) = \max_{i \leq k \leq i+l_j-1} W_k$, $(i = 0, \ldots, m - l_j)$, $r = \min\{R(i) \mid 0 \leq i \leq m - l_j\}$. Legyen $I_0 = \{k \mid R(k) = r,\ 0 \leq k \leq m - l_j\}$ és legyen $i_0 \in I_0$ tetszőleges index.

Nevezzük $\mathcal{BS}$-nek (Bottom Side) a $\mathcal{BL}$ algoritmusnak azt a változatát, ahol a következőt végezzük: ha több helyen vétetik fel a 3. pontbeli minimum, akkor helyezzük a téglalapot oda, ahol a sáv valamelyik széléhez a lehető legközelebb van. Ekkor a 3. lépés a következőképpen változik tehát:

3. Legyen $R(i) = \max_{i \leq k \leq i+l_j-1} W_k$, $(i = 0, \ldots, m - l_j)$, $r = \min\{R(i) \mid 0 \leq i \leq m - l_j\}$. Legyen $I_0 = \{k \mid R(k) = r,\ 0 \leq k \leq m - l_j\}$. Legyen $i_1 = \min\{k \in I_0\}$, $i_2 = \max\{k \in I_0\}$. $i_1 \leq m - l_j - i_2$ esetén legyen $i_0 = i_1$, egyébként legyen $i_0 = i_2$.

A $\mathcal{BL}$ illetve $\mathcal{BS}$ algoritmusok mindegyike a $\mathcal{B}$ algoritmus speciális esete, amire érvényes a következő

14. TÉTEL. *Ha a $\mathcal{B}$ algoritmus elvégzése során a téglalapokat csökkenő szélesség szerint helyezzük el, akkor teljesül a következő egyenlőtlenség:*

$$\frac{C(\mathcal{B})}{C_{R2D}} \leq 4. \tag{10}$$

Bizonyítás. Nagyobb általánosságban bizonyítunk, pontosabban belátjuk a következő lemmát:

LEMMA. *Tegyük fel, hogy a téglalapokat egy L lista szerinti sorrendben helyezzük el, a sávon belül tetszőleges helyre, rögtön ezután lefelé igazítva őket, az utolsóként elhelyezett téglalap szélessége minimális, az utolsó téglalapot a sáv aljához a lehető legközelebb helyezzük el a lefelé igazítása előtt, továbbá a maximális felhasznált sávmagasság egyenlő az utolsóként elhelyezett téglalap tetejének a sáv aljától mért távolságával. Ekkor az ütemezés C' erőforrásigénye az L_1 alsó becslésnek legfeljebb négyszerese.*

A Lemma bizonyítása. Jelöljük h^*-gal azt a magasságot, ahova az utolsó, n-edik téglalapot helyezzük, még mielőtt lefelé igazítanánk. Az így kapott ütemezés erőforrás-igényének maximuma, vagyis a maximálisan használt sávmagasság ekkor $C' = h^* + w_n$. Húzzunk a h^* magasságban egy vízszintes vonalat. Ennek a sáv baloldali határával vett metszéspontja legyen az A, a jobboldalival vett metszéspontja a B pont. Ha sikerül belátnunk, hogy a h^* magasságig a sáv területének legalább egyharmad része ki van töltve, akkor ebből már következik az állítás, ugyanis $w_n \leq \frac{h^*}{3}$ esetén

$$\frac{C'}{L_1} = \frac{h^* + w_n}{L_1} \leq \frac{h^* + w_n}{\frac{h^*}{3}} \leq \frac{h^* + \frac{h^*}{3}}{\frac{h^*}{3}} = 4, \tag{11}$$

ha pedig $w_n > \frac{h^*}{3}$, akkor ebből

$$\frac{C'}{L_1} = \frac{h^* + w_n}{L_1} \leq \frac{h^* + w_n}{w_n} \leq \frac{3w_n + w_n}{w_n} = 4. \tag{12}$$

Most vizsgáljuk meg a h^* magasságban meghúzott AB szakaszt. Ez átmetszhet már letett téglalapokat, vagy azok egyes részeit, illetve kitöltetlen részeken halad keresztül. Jelöljük az előbbi pontok halmazát H_1-gyel, illetve H_0-val. Mivel az utolsó téglalapot nem tudtuk elhelyezni a h^* magasságnál lejjebb, (a lefelé igazítása előtt), az AB szakaszon nincs legalább l_n hosszúságú kitöltetlen (tehát H_0-beli) rész. Most legyen $P \in H_1$ tetszőleges pont, és legyen R_1 azon téglalapok halmaza, amelyeknek megfelelő munkák folyamatban vannak a P-nek megfelelő t időpontban. Ezek mindegyike legalább l_n hosszú ideig folyamatban van, így a $[t - l_n, t + l_n]$ időintervallum alapú és h^* magasságú T téglalap legalább félig ki van töltve. Jegyezzük meg, hogy ez akkor is igaz, hogy ha ez az időintervallum „kilóg" a $[0, m-1]$ időintervallumból, (vagyis hogyha $t - l_n < 0$ vagy ha $t + l_n > m - 1$). Ekkor viszont az előbbi $[t - l_n, t + l_n]$ időintervallumnak a sávba eső részére igaz az, hogy legalább $l_n \times h^*$ területnyi része kitöltött. A bizonyítást ezután részekre osztjuk aszerint, hogy az m sávszélesség az l_n értéknek hányszorosa.

1. $1 \leq \frac{m}{l_n} \leq 3$. Ekkor a T téglalapnak a sávba eső részéből legalább $l_n \times h^*$ területnyi rész kitöltött, ennek az $m \times h^*$ terület legföljebb háromszorosa lehet.
2. Legyen $\frac{m}{l_n} = 3k$, ahol k pozitív egész. Osszuk fel az AB szakaszt k darab $3 \cdot l_n$ hosszúságú $A_\alpha B_\alpha$ intervallumra $\alpha = 1 \ldots k$, (ahol $A_1 = A$, $B_k = B$, és $A_\alpha = B_{\alpha-1}$minden α-ra), majd minden $A_\alpha B_\alpha$ intervallumnak a középső l_n hosszú részéből válasszunk egy-egy H_1-beli pontot, azaz legyen $P_\alpha \in H_1$, $P_\alpha \in A_\alpha B_\alpha$, $\alpha = 1 \ldots k$. Ilyen pontok léteznek, mert az AB szakasz akármelyik l_n hosszú intervallumának van H_1-gyel közös pontja. Ekkor az előzőek alapján az $A_\alpha B_\alpha \times h^*$ téglalapok mindegyike legalább egyharmadáig ki van töltve.
3. Legyen $3k \leq \frac{m}{l_n} < 3k+1$. Osszuk fel ugyanúgy az AB szakaszt $3 \cdot l_n$ hosszúságú $A_\alpha B_\alpha$ intervallumokra. Az első $k-1$ számú $3l_n$ szélességű részben a kitöltöttség az előzőek szerint legalább egyharmados. Ezután a maradék $(4-x) \cdot l_n$ szélességű intervallumra koncentrálunk, (ahol $0 < x \leq 1$ teljesül). Ezt az intervallumot jelöljük CD-vel, mérjünk fel a C ponttól jobbra $2 \cdot l_n$ egységet, így nyerjük a C_1 pontot, majd D- től balra is $2 \cdot l_n$ egységet, így kapjuk a D_1-et. (A pontok sorrendje tehát C, D_1, C_1, D.) A $CC_1 \times h^*$ és $D_1D \times h^*$ téglalapok mindegyike legalább feléig ki van töltve, így a CC_1 intervallumhoz tartozik legalább $l_n \times h^*$ területegység, a D_1D intervallumhoz is tartozik legalább ennyi. A középső D_1C_1 intervallum hossza $x \cdot l_n$, így ha az ehhez tartozó sáv teljesen ki van töltve, akkor is marad a $C_1D \times h^*$ téglalapban legalább $(1-x)$-szer $l_n \times h^*$ területegység. Ezek szerint a CD intervallum CC_1 részéhez legalább $l_n \times h^*$ területegység tartozik, a C_1D intervallumhoz legalább $(1-x)$-szer $l_n \times h^*$ területegység, így a CD intervallumhoz összesen legalább $(2-x) \cdot l_n \times h^*$ területegység. A CD intervallum fölötti rész kitöltöttsége ezek szerint legalább $\frac{2-x}{4-x}$, és könnyen látható, hogy ez az arány $x = 1$ esetén minimális, és akkor éppen egyharmad.
4. $3k+1 \leq \frac{m}{l_n} < 3k+3$ esete. Mérjünk fel az AB szakaszon balról $3(k-1)l_n$ egységet, itt a kitöltöttség az előző pont szerint legalább egyharmados. Ezen túl marad egy legalább $4 \cdot l_n$ hosszú, de $6 \cdot l_n$-nél rövidebb rész a szakaszból. Mérjünk fel itt balról is és jobbról is $2 \cdot l_n$ egységet, az ezekhez tartozó sávrészek kitöltöttsége legalább $\frac{1}{2}$, így a maradék kitöltöttsége is legalább $\frac{2}{6}$.

A tétel bizonyítása ezután a következő: Tegyük fel, hogy a téglalapokat a szélességük szerinti csökkenő sorrendben helyezzük el. Legyen a j-edik az a téglalap, amelyik esetén a maximális a felhasznált sávmagasság. Ekkor a Lemma erre a j-edik, mint utolsóként elhelyezett téglalapra alkalmazható, az eddig felhasznált sávmagasság az első j darab téglalaphoz tartozó L_1 értéknek legfeljebb a négyszerese. Ezután a maradék téglalapokat elhelyezve a felhasznált sávmagasság nem változik, az L_1 alsó becslés értéke pedig csak növekedhet, így a $\frac{C'}{L_1}$ arány nem nőtt. □

[5] közöl egy példát, ahol monoton csökkenő szélesség szerinti sorrend esetén a $\frac{C'}{L_1}$ arány legalább $2.5 - \varepsilon$, illetve a $\frac{C'}{C_{R2D}}$ arány legalább $\frac{7}{3} - \varepsilon$. Ezek szerint a $\frac{C'}{C_{R2D}}$

arány esetén az éles becslés $\frac{7}{3}$ és 4 közötti szám, a $\frac{C'}{L_1}$ arány esetében pedig 2.5 és 4 közötti szám. Nyitott kérdés az is, hogy hogyan változnak ezek a felső becslések, ha a téglalapokra további megkötéseket teszünk, például ha a szélességeik is és a magasságaik is monoton csökkennek. [1] ennek csak azzal a speciális esetével foglalkozik, amikor a téglalapok négyzetek.

3.3. Az $\mathcal{FFDH}$ algoritmus

A kétdimenziós $\mathcal{FFDH}$ (First Fit Decreasing Height) algoritmus [2] a téglalapokat a magasságaik szerinti csökkenő sorrendben helyezi el egy-egy szintre egymás mellé balra igazítva, a következő téglalapot mindig a legelső olyan szintre helyezi el, ahova befér. Ha már nem fér be egyetlen korábbi szintre sem, akkor közvetlenül a már létező szintek fölött egy új szintet nyit, és ide rakja az aktuális téglalapot. Legyen i a szintek száma, t_α az α-adik szinten lévő téglalapok szélességeinek összege, és tegyük fel, hogy az α-adik szint s_α magasságban van.

Az $\mathcal{FFDH}$ algoritmus

1. Legyen $i = 1$, $t_1 = 0$, $s_1 = 0$.
2. Rendezzük a téglalapokat a magasságaik szerinti csökkenő sorrendbe, legyen $j := 1$.
3. $t_\alpha + l_j > m$ $(\alpha = 1, \ldots, i)$ esetén legyen $t_{i+1} = 0$, $s_{i+1} = s_i + w_j$, $j = j + 1$.
4. Legyen $\alpha_0 = \min\{\alpha \mid t_\alpha + l_j \leq m\}$. A j-edik téglalap az α_0-adik szintre kerül, $t_{\alpha_0} := t_{\alpha_0} + l_j$.
5. Legyen $j = j + 1$. $j \leq n$ esetén menjünk a 3. lépésre, egyébként vége.

Az algoritmus megfelelő módosítással természetszerűleg R2D algoritmusnak is tekinthető. Így érvényes a [3] dolgozatban közölt elméleti hatékonyság-becslés. Ezek szerint $\frac{C(\mathcal{FFDH})}{C_{R2D}} \leq 2,7$. Ha az egyes szinteket nem mindig balról jobbra, hanem alternálva balról jobbra és jobbról balra töltjük fel, akkor ugyan az elméleti felső korlát nem változik, de gyakran jobb megoldást kaphatunk. (Ennél jobb lehetőség lenne az, hogy nem egyszerűen alternálva töltjük fel a szinteket, hanem a következőképpen: Ha a bal szélen az aktuális összmagassága a téglalapoknak nagyobb mint a jobb szélen, akkor a következő szintet jobbról indulva töltjük fel, ellenkező esetben pedig balról jobbra. Ha az utolsó sáv feltöltése után létrejövő baloldali illetve jobboldali összmagasság különbségét akarnánk minimalizálni NP-teljes feladatot kapnánk, így megelégszünk annyival, hogy csak a soron következő szintnek a balról, vagy jobbról való feltöltéséről döntünk.)

3.4. Egy új algoritmus ($\mathcal{MIX}$), a $\mathcal{BL}$ és $\mathcal{FFDH}$ algoritmusok keveréke

A címben szereplő algoritmusoknak a következő keverékét készítjük el: Rendezzük valamilyen sorrendbe a téglalapokat. Helyezzük el a lehető leglejjebb a következő téglalapot, ha már az előző szinten nem fér el. Ennek a szintnek a magasságában helyezzünk el annyi téglalapot, amennyit csak lehetséges, balról jobbra feltöltve ezt a megkezdett szintet, és ezt a két lépést iteráljuk. Az algoritmust

nevezzük $\mathcal{MIX}$ algoritmusnak. Ez így egy kétdimenziós algoritmus, a lefelé igazításra megint többféle lehetőség kínálkozik: 1. Ahogy egy téglalapot elhelyezünk, rögtön ezután lefelé igazítjuk. 2. Ahogy egy-egy szinten elhelyeztünk maximális számú téglalapot, ezeket ezután egyszerre lefelé igazítjuk. 3. Csak az összes téglalap elhelyezése után igazítunk lefelé. Könnyen látható, hogy az [1]-beli példa esetén a $\frac{C(\mathcal{MIX})}{C_{R2D}}$ arány tetszőlegesen nagy lehet, ha a téglalapok sorrendje a magasságaik szerinti csökkenő. Csökkenő szélesség esetén a korábbi Lemma alkalmazható, és így az előbbi arány legfeljebb 4. A becslés ebben az esetben sem tűnik élesnek, az arány sejtésünk szerint legföljebb 3. Az algoritmus formális leírása:

A $\mathcal{MIX}$ algoritmus

1. Legyen m az időintervallum hossza, $P_i = \emptyset$, $W_i = 0$ $(i = 0, \ldots, m-1)$.
2. Rendezzük a téglalapokat az L lista szerinti sorrendben, legyen $j := 1$.
3. Legyen $s = \min_{0 \le i \le m-l_j} \max_{i \le k \le i+l_j-1} W_k$, vagyis s a következő szint magassága.
4. Legyen $i_0 = \min \left\{ i : \max_{i \le k \le i+l_j-1} W_k = s,\ i \in \{0, \ldots, m-l_j\} \right\}$.
5. $i_0 > -\infty$ esetén ütemezzük a j-edik munkát az i_0-adik időpontra, vagyis legyen $P_{i_0} = P_{i_0} \cup \{r_j\}$, és legyen $W_k = W_k + w_j$, $(k = i_0, \ldots, i_0 + l_j - 1)$. Legyen $j = j+1$, $j > n$ esetén vége, egyébként menjünk újra a 4. pontra.
6. $i_0 = -\infty$ esetén erre a szintre már nem fért be téglalap. Legyen $j = j+1$, $j \le n$ esetén menjünk a 3. pontra, (új szint keresése), egyébként vége.

Jegyezzük még meg, hogy ha a szinteket alternálva balról jobbra, illetve jobbról balra töltjük fel, akkor sok esetben jobb megoldást kapunk. (Itt is alkalmazható az előző paragrafus végén leírt bonyolultabb módszer is.) A második ponbeli L lista lehet a magasságok vagy szélességek szerinti csökkenő sorrend. A következő paragrafusban látni fogjuk, hogy sok feladatosztály esetén az új $\mathcal{MIX}$ algoritmus a másik két algoritmusnál lényegesen többször ad jobb megoldást.

4. Numerikus eredmények

4.1. A BL és BS algoritmusok összehasonlítása

Az első táblázat a $\mathcal{BL}$ és $\mathcal{BS}$ algoritmusok összehasonlítását tartalmazza. A sáv szélessége m, a téglalapok száma n, a téglalapok szélességét az $[x_1, x_2]$, a magasságukat pedig az $[y_1, y_2]$ intervallumból választottuk egyenletes eloszlás szerint (a kapott számokat lefelé kerekítve). A 13. Tétel szerint csökkenő magasság szerinti ütemezés esetén a $\frac{C(\mathcal{BL})}{C^*}$ arány tetszőlegesen nagy lehet, míg a 14. Tétel szerint csökkenő szélesség szerinti ütemezés esetén ez az arány legfeljebb 4. Mégis, meglepő módon nem mindig a csökkenő szélesség szerinti ütemezés ad jobb eredményt. A sorokban lévő többi szám azt mutatja, hogy az algoritmus által kapott megoldás

tízezer független kisérletet végezve hányszor volt a négy között minimális értékű.

	m	n	x_1, x_2	y_1, y_2	csökk. szélesség		csökk. magasság	
					BL	BS	BL	BS
1.	29	27	5,6	11,18	9745	9745	1512	1797
2.	35	50	2,7	1,8	120	98	8815	9050
3.	35	50	5,7	10,15	2275	4576	1520	6218
4.	40	50	6,7	13,15	4	6192	2288	9279
5.	21	50	4,7	13,15	188	2216	1401	9470

Az első esetben a csökkenő szélesség szerinti ütemezés esetén csaknem mindig minimális eredményt adott, a $\mathcal{BL}$ és $\mathcal{BS}$ algoritmusok mindegyike. A következő példa esetén fordított a helyzet, itt a csökkenő magasság szerinti ütemezés ad általában jobb eredményt, és a $\mathcal{BS}$ algoritmus némileg jobb a $\mathcal{BL}$ algoritmusnál. Ez azért érdekes, mert előzőleg láttuk, hogy csökkenő magasság szerinti sorrend esetén a $\frac{C(\mathcal{BL})}{C_{R2D}}$ és $\frac{C(\mathcal{BS})}{C_{R2D}}$ arányok tetszőlegesen nagyok is lehetnek. A 3. példa az előbbi feladatosztálytól csak a téglalapok szélességeinek és magasságainak megválasztásában különbözik. Itt érdekes módon az előzőtől lényegesen eltér az eredmény: A $\mathcal{BS}$ algoritmus mindkét sorrend esetén lényegesen jobb a $\mathcal{BL}$-nél, ezek között is jobb a csökkenő magasság szerinti sorrend. A 4. feladatosztály esetén az előbbi tendencia erősödött fel, a $\mathcal{BS}$ algoritmus lényegesen jobb a $\mathcal{BL}$ algoritmusnál. Az utolsó sor példája esetében pedig a négy algoritmus közül a csökkenő magasság szerinti ütemezéssel végzett $\mathcal{BS}$ algoritmus az eseteknek durván 95%-ában volt a legjobbak között, és legalább 60%-ban egyedüli legjobb volt.

4.2. Mennyire jók az alsó becslések?

Ebben a részben azt vizsgáltuk, hogy az alsó becslések mennyire közelítik meg a heurisztikus megoldás értékét, ahol csak a (korábbi) $\mathcal{BL}$ és $\mathcal{FFDH}$ algoritmusokat alkalmaztuk. Az első tesztfeladat esetén a sávszélességet $m = 19$-nek választottuk, a téglalapok szélességeit az $[1, 5]$, a magasságokat az $[1, 9]$ intervallumból választottuk egyenletes eloszlás szerint. A téglalapok száma először 10, majd 20, és végül 50. A három egymás alatti sorban a következő értékek szerepelnek: Hányszor volt az alsó becslés értéke egyenlő a jobbik heurisztikus megoldással, és persze akkor az optimummal is; hányszor volt a jobbik heurisztikus megoldásnak legalább 0.9-szerese, illetve az alsó becslés/heurisztikus megoldás törtek átlaga százalékban kifejezve, 100 független kisérletet elvégezve.

$m = 19, \quad l_i \in [1,5], \quad w_i \in [1,9]$					
n	L_1	L_2	L_3	L_4	
10	53	67	81	88	$= Heu$
	77	92	96	99	$\geq 0.9Heu$
	93,80	96,48	97,92	98,78	$\frac{L}{Heu}$
20	7	10	11	12	$= Heu$
	76	79	79	79	$\geq 0.9Heu$
	92,47	92,85	92,93	93,05	$\frac{L}{Heu}$
50	13	13	13	13	$= Heu$
	100	100	100	100	$\geq 0.9Heu$
	97,06	97,06	97,06	97,06	$\frac{L}{Heu}$

$n = 10$ esetén már az L_1 becslés is elég jó értéket ad (az esetek felében optimális megoldás), ez a tulajdonság a téglalapok számának növelésével viszont romlik. Érdekes módon elég nagy téglalapszám esetén ($n = 50$) az eseteknek csak a töredékében szolgáltat L_1 pontos becslést, viszont minden esetben legalább 0.9-szerese az optimumnak. Vagyis egy bizonyos téglalapszámon túl az L_1 becslésen a többi becslés már nem javít. Az alábbi táblázat két olyan speciális esettel foglalkozik, amikor az előzőektől eltérően lényegesen javítottak a következő becslések az L_1 becslés értékén. Az egyik eset az, amikor a téglalapok majdnem egyformák, és ekkor már az L_2 becslésnek közel kell lennie az optimumhoz, a másik, amikor a téglalapok összszélessége kisebb mint a sávszélesség kétszerese, és ekkor azt várjuk, hogy az L_3 illetve L_4 becslések adnak majnem pontos eredményt.

n	m			L_1	L_2	L_3	L_4	
				0	0	0	24	$= Heu$
20	12	[4,5]	[3,4]	16	16	16	100	$\geq 0.9Heu$
				87,51	87,51	87,51	96,99	$\frac{L}{Heu}$
20	60	[1,9]	[4,8]	3	54	68	92	$= Heu$
				47	83	88	94	$\geq 0.9Heu$
				87,02	94,97	96,43	98,78	$\frac{L}{Heu}$

Érdekes módon az első esetben nem L_2, hanem az L_4 becslés adott „jó" eredményt. A másik esetben a várakozásnak megfelelően az L_4 becslés általában jobb az előzőeknél.

4.3. A korábbi és a $\mathcal{MIX}$ algoritmusok összehasonlítása

Az alábbi két táblázat a [3]-ban szereplő tesztek mintájára következik. Az elsőben a téglalapok szélességét és magasságát egyenletes eloszlás szerint választottuk az $[x_1, x_2]$ illetve az $[y_1, y_2]$ intervallumból, kerekítéssel. A $\mathcal{BL}$ esetén csökkenő szélesség szerinti, az $\mathcal{FFDH}$ és $\mathcal{MIX}$ esetében csökkenő magasság szerinti a téglalapok sorrendje. A $\mathcal{MIN}$ algoritmus a korábbi heurisztikus módszerek közül az aktuálisan legjobb értéket adó algoritmust jelöli. Az egyes sorokban lévő számok azt mutatják, hogy hányszor érte el a heurisztikus megoldás a legjobb alsó becslést (ami az L_4 becslés, ekkor ez a heurisztikus megoldás biztosan optimális megoldás is egyben); hányszor volt ennek legföljebb 1.05-szerese; illetve a heurisztikus megoldás/alsó becslés arányok átlagát.

m	n	x_1, x_2	y_1, y_2	BL	FFDH	MIX	MIN	
29	27	1,4	1,8	427	5528	9711		$= L_4$
				427	5528	9711		$\leq 1.05 \cdot L_4$
				1.249	1.083	1.029	1.028	C/L_4
72	93	1,8	1,9	0	4950	9788		L_4
				451	9583	9991		$\leq 1.05 \cdot L_4$
				1.113	1.031	1.011	1.010	C/L_4
40	50	6,7	12,15	8	3756	9587		L_4
				2510	9976	9983		$\leq 1.05 \cdot L_4$
				1.152	1.081	1.073	1.072	C/L_4

Látható, hogy a $\mathcal{MIX}$ algoritmus általában jobb a többinél. A $\frac{C}{L_4}$ arány minden esetben kisebb a $\mathcal{MIX}$ oszlopában, a legnagyobb különbség abban van, hogy sokkal többször adja a négy algoritmus közötti legkisebb megoldásértéket ez az algoritmus. A legjobb heurisztikus megoldás és a legjobb alsó becslés átlagosan csak néhány százalékkal tér el egymástól. Több feladatosztály esetén nagy százalékban a legjobb heurisztikus megoldás (általában a $\mathcal{MIX}$) biztosan optimális volt, emellett más esetekben is lehet hogy a legjobb heurisztikus megoldás elérte az optimumot, csak az alsó becslés pontatlansága ezt nem tudta kimutatni. A következő táblázat esetében egy m egység szélességű, y egység magasságú téglalapot véletlenszerűen felosztottunk legföljebb *xpar* számú függőleges sávra, majd ezután ezeket egymástól függetlenül vízszintesen legföljebb *ypar* számú téglalapra, és az

így kapott téglalaphalmazzal dolgoztunk.

	m	y	$xpar$	$ypar$	BL	FFDH	MIX	
1.	100	100	7	2	5	10000	10000	L_4
					28	10000	10000	$\leq 1.05 \cdot L_4$
					1.472	1	1	C/L_4
2.	100	100	7	5	375	1397	8730	L_4
					989	3947	9682	$\leq 1.05 \cdot L_4$
					1.285	1.170	1.084	C/L_4
3.	100	100	15	4	8	447	9718	L_4
					47	1849	9983	$\leq 1.05 \cdot L_4$
					1.371	1.189	1.062	C/L_4
4.	100	100	16	8	0	75	9972	L_4
					42	130	9997	$\leq 1.05 \cdot L_4$
					1,229	1,155	1,026	C/L_4

Most már az L_1 becslés is az optimummal egyenlő, így az L_4 is. A kapott adatok szerint az első feladatosztályban még túl kevés részre vágjuk szét a nagy téglalapot, a csökkenő szélesség szerinti $\mathcal{FFDH}$ és $\mathcal{MIX}$ algoritmusok minden esetben, míg $\mathcal{BL}$ viszont szinte sohasem ad optimumot. Ha növeljük a vágások számát, már nem mindig adnak optimumot az előzőek, de egyre inkább kitűnik a $\mathcal{MIX}$ fölénye a többivel szemben. A 4. feladatosztály esetén már csaknem mindig a $\mathcal{MIX}$ az egyedüli legjobb algoritmus a négy között: most az $\mathcal{FFDH}$ is végérvényesen lemaradt vele szemben. A $\mathcal{MIX}$ a $\frac{C}{L_4}$ arány tekintetében is messze jobb a másik három algoritmusnál. (Sajnos azonban a $\mathcal{MIX}$ algoritmus sem mindig ad optimális megoldást, vagyis a nagy téglalap magasságát.) Jegyezzük meg, hogy nem csak a 100×100-as méretű téglalapnál, hanem általánosságban is az előbbieket tapasztaltuk. Megállapíthatjuk, hogy sok esetben a legjobb heurisztikus megoldás, (ami az előbbi feladatosztályokban szinte mindig a csökkenő szélesség szerint végrehajtott $\mathcal{MIX}$ algoritmus), optimális, vagy ahhoz közeli megoldást adott.

Még annyit jegyezzünk meg, hogy a heurisztikus megoldás/alsó becslés törtre vonatkozó felső becslést kapunk úgy, ha a heurisztikus módszer hatékonysági becslésében szereplő konstanst elosztjuk az alsó becslés elméleti hatékonyságával. Az így kapott számnál a heurisztikus megoldás/alsó becslés arány szupremuma valójában kisebb is lehet, de erre vonatkozó vizsgálatot jelen cikkel kapcsolatos kutatásaink során nem végeztünk.

Köszönetnyilvánítás. A szerző megköszöni Vizvári Béla értékes segítségét a feladat megfogalmazásában, és a cikk szerkezetének a kialakításában.

IRODALOM

[1] B. S. Baker, E. G. Coffman, R. L. Rivest, Orthogonal packings in two dimension, *SIAM J. Comput,* **4** (1980), 846–855.

[2] E. E. Coffman, Jr., M. R. Garey, D. S. Johnson, R. E. Tarjan, Performance bounds for level-oriented two-dimensional packing algorithms, *SIAM J. Comput,* **4** (1980), 808–826.

[3] E. E. Coffman, Jr., M. R. Garey, D. S. Johnson, An application of bin-packing to multiprocessor scheduling, *SIAM J. Comput.,* **7** (1978), 1–17.

[4] M. Dell'Amico, S. Martello, "Optimal Scheduling of Tasks on Identical Parallel Processors", Research Report, DEIS, University of Bologna, OR/90/7.

[5] Dósa Gy., „Alsó becslések a másféldimenziós téglalappakolási feladatra", Preprint, Univ. of Veszprém, No. 72 (1998).

[6] D. K. Friesen, Tighter bounds for the multifit processor scheduling algorithms, *SIAM J. Comput.,* **13** (1984), 170–181.

[7] D. K. Friesen, M. A. Langston, Evaluation of a MULTIFIT-based scheduling algorithm, *J. Algorithms,* **7** (1986), 35–59.

[8] R. L. Graham, Bounds for certain multiprocessor anomalies, *Bell System Tech. J.,* **45** (1966), 1563–1581.

[9] R. L. Graham, Bounds on multiprocessor timing anomalies, *SIAM J. Appl. Math.,* **17** (1969), 416–429.

[10] M. Hujter, "On the dynamic storage allocation problem", Manuscript, Computer and Automation Institute Hung. Acad. Sci. (1990).

[11] Racsmány Anna, *Ütemezéselmélet* (Egyetemi jegyzet, MKKE, 1981).

[12] Zs. Soós, L. Varga, "Greedy Algorithm for Maintenance Scheduling in Electric Power System" in: *Proc. of the Fifth ECMI Conf.,* Ed. M. Heiliö (B. G. Teubner and Kluwer, 1991), 115-118.

[13] B. Vizvári, R. Demir, R., "It is Difficult to Find a Difficult Problem for the Scheduling of Identical Parallel Machines", Department of Industrial Engineering of Bilkent University, Research Report, IEOR-9212, 1992.

[14] B. Vizvári, R. Demir, "A Coloumn Generation Algorithm to Schedule Identical Parallel Machines", Rutgers University, Research Report, RRR-99-93, September 1994.

[15] B. Vizvári, *Bevezetés a termelésirányítás matematikai elméletébe* (Egyetemi jegyzet, ELTE, 1991).

(Beérkezett: 1999. március 9.)

DÓSA GYÖRGY
VESZPRÉMI EGYETEM
8201 VESZPRÉM
PF. 158
E-mail: dosagy@almos.vein.hu

HEURISTICAL METHODS FOR THE RELAXED TWO-DIMENSIONAL RECTANGLE-SCHEDULING PROBLEM

GYÖRGY DÓSA

We are dealing with a problem of one-resource-constrained scheduling: How can a number of n tasks be distributed on the time-interval $[0, m]$ so that the maximum of the amount of the resource used be minimum. There are two special cases: the scheduling of identical parallel machines, and the bin-packing problem. We apply two earlier methods and introduce a new heuristics. Upper bounds and numerical aspects are also investigated. The results show that in many cases the new method is better than the previously existing ones.

Alkalmazott Matematikai Lapok **19** (1999) 155–168

MULTIFIT TÍPUSÚ MÓDSZEREK PÁRHUZAMOS GÉPEK ÜTEMEZÉSÉRE

DÓSA GYÖRGY

Veszprém

Egy fontos ütemezéselméleti problémával, egyforma párhuzamos gépek ütemezésével foglalkozunk: Hogyan osszunk szét n munkát $m < n$ gép között úgy, hogy a teljes átfutási idő a legkisebb legyen, vagyis a legkésőbben befejeződő munka a lehető legkorábban fejeződjön be. Az [1]-ben szereplő Multifit algoritmust általánosítjuk: a logaritmikus keresés keretét a First Fit Decreasing algoritmus helyett más algoritmusokra alkalmazzuk. Pontos becslést adunk az elméleti hatékonyságra, és új algoritmusokat vezetünk be. Kísérleti eredményekkel igazoljuk, hogy az új algoritmusok sok esetben jobb megoldást adnak a régieknél.

1. Bevezetés

Az ütemezéselmélet gyakran vizsgált feladata az egyforma párhuzamos gépek ütemezésének problémája, a $P \parallel C_{max}$ probléma. Ez a következő: Adott egy feladathalmaz: $\mathcal{T} = \{T_1, T_2, \ldots, T_n\}$, ahol T_i $(i = 1 \ldots n)$ olyan munkát jelent, amelyet m egyforma gép valamelyikével kell elvégezni. A T_i munka elvégzésének ideje, vagy röviden hosszúsága $l(T_i)$. A munkák elvégzése tehát bármely gépen ugyanannyi időbe kerül. Megszakítást nem engedünk meg, vagyis ha egy gép valamelyik munkát elkezdi, akkor azt be is kell fejeznie. A munkák egy ütemezésén a $\mathcal{T}$ halmaz valamely $\mathcal{P} = \{P_1, P_2, \ldots, P_m\}$ partícióját értjük. (Ekkor az i-edik gép végzi a P_i-ben lévő feladatokat.) Feltesszük, hogy nincsenek várakozási idők: amint egy gép befejezi valamelyik általa elvégzendő munkát, és még van olyan, amit őneki kell elvégezni, akkor azt rögtön el is kezdi. Az egy-egy gépen elvégzendő munkák sorrendje az ütemezési feladat szempontjából közömbös. A $\mathcal{P}$ ütemezés (teljes) átfutási idejét a következőképpen definiáljuk:

$$\mathcal{L}(P) = \max_{1 \le i \le m} l(P_i) \tag{1}$$

ahol $\mathcal{T}$ tetszőleges X részhalmaza esetén $l(X) = \sum_{T \in X} l(T)$, vagyis $l(P_i)$ az i-edik gépen a legkésőbb befejeződő munka befejezési idejét jelenti, ha feltesszük, hogy a gépek a 0 időpontban kezdenek dolgozni.

A $\mathcal{P}^*$ ütemezés optimális, ha teljesül $\mathcal{L}(P^*) \leq \mathcal{L}(\mathcal{P})$ a $\mathcal{T}$ halmaz tetszőleges $\mathcal{P}$ ütemezése esetén. Mivel véges sokféleképpen tudjuk a $\mathcal{T}$ halmaz elemeit m részre partícionálni, ilyen optimális ütemezés biztosan létezik, esetleg több is lehet. Az $\mathcal{L}(\mathcal{P}^*)$ értéket jelöljük C^*-gal, ami tehát csak $\mathcal{T}$-től és az m számtól függ.

A feladat téglalappakolási feladatként is megfogalmazható: Itt a munkáknak egy-egy $l(P_i)$ magasságú, egy egység szélességű téglalap felel meg, ezeket kell egy m egység szélességű, alul zárt, felül nyitott sávon elhelyezni átfedés nélkül úgy, hogy minimális magasságot használjunk fel a sávból. A téglalapok oldalai a sáv oldalaival párhuzamosak, és függőlegesen állnak, vagyis a forgatásukat nem engedjük meg, számos gyakorlati alkalmazásban ugyanis a két méret különböző dolgot jelent.

Az optimális ütemezés(ek) megkeresésének feladata NP-teljes, így ehelyett gyors (polinomiális), és hatékony, vagyis közel-optimális ütemezéseket adó algoritmusokat keresünk. A cikk szerkezete a következő. A 2. paragrafusban két korábbi algoritmust mutatunk be, amelyeknek közös általánosításaival foglalkozunk a későbbiekben. A harmadik paragrafusban közöljük az új, módosított Multifit-típusú algoritmus-családot, és erre pontos hatékonyságbecslést adunk. A 4. paragrafusban bevezetünk néhány új, Multifit típusú algoritmust, és numerikus vizsgálatokkal foglalkozunk.

2. Korábbi eredmények

2.1. Az LPT algoritmus

R. L. Graham 1966 -os cikkében [4] közölte az LPT algoritmust (LPT=Longest Processing Time). Ez a következő: Először a munkákat monoton csökkenő sorrendbe rakjuk a végrehajtásukhoz szükséges idő szerint. A munkákat ebben a sorrendben ütemezzük, mindig a lehetséges legkorábbi időpontra, vagyis a legkorábban felszabaduló gépre tesszük a következő, még nem ütemezett munkát. Formálisan:

Az LPT algoritmus

0. Legyen $P_i = \emptyset$ $(i = 1 \ldots m)$, $j = 1$.
1. Legyen $i_0 = \arg\min \{l(P_i)\}$.
2. Legyen $P_{i_0} = P_{i_0} \cup \{T_j\}$, $j = j + 1$.
3. $j \leq n$ esetén menjünk az 1. lépésre, egyébként vége.

Jelöljük egy tetszőleges $\mathcal{A}$ algoritmus által meghatározott ütemezés átfutási idejét $\mathcal{L}(A)$-val. (Valójában ez az érték a gépek m számától, valamint a $\mathcal{T}$ feladathalmaztól is függ.) Bevezetjük a következő mennyiséget:

$$R_m(\mathcal{A}) = \sup_{\mathcal{T}} \left\{ \frac{\mathcal{L}(A)}{C^*} \right\}, \tag{2}$$

ahol $\mathcal{T}$ tetszőleges feladathalmaz. Ekkor Graham algoritmusára teljesül a következő

1. TÉTEL. $R_m(\mathrm{LPT}) = \left(\frac{4}{3} - \frac{1}{3m}\right)$. □

Az $R_m(\mathcal{A})$ szám azt „méri", hogy az $\mathcal{A}$ algoritmus által adott ütemezés átfutási ideje legfeljebb hányszorosa az optimális ütemezés átfutási idejének. Graham algoritmusa esetében ez a szám $\frac{4}{3} - \frac{1}{3m}$. A „sup" helyett a fenti képletben „max" is állhatna: van olyan $\mathcal{T}$ feladathalmaz, amely esetén $\mathcal{L}(\mathrm{LPT}) = \left(\frac{4}{3} - \frac{1}{3m}\right) \cdot C^*$.

2.2. A MULTIFIT algoritmus

A Multifit algoritmust több szerző javasolta [1]-ben. Itt logaritmikus keresés történik, minden lépésben az FFD (=First Fit Decreasing) ládapakolási algoritmust alkalmazva valamilyen C láda-mérettel. (A ládapakolási feladat esetében adott n számú, T_j méretű tárgy ($j = 1 \ldots n$), és ezeket akarjuk valahogyan elhelyezni m számú, C méretű ládában úgy, hogy az egy-egy ládában lévő tárgyak összmérete nem haladhatja meg a láda C kapacitását.) Az FFD algoritmus is a végrehajtási idő csökkenő sorrendje szerint ütemezi a munkákat, vagy a ládapakolás nyelvén kifejezve a ládákba a tárgyakat a méretük monoton csökkenő sorrendje szerint helyezi el. A soron következő tárgyat az indexsorrend szerinti legelső olyan ládába teszi, ahova befér. Legyen C a ládák mérete, m a ládák száma. Ekkor az FFD algoritmus formális leírása a következő:

Az FFD algoritmus

0. Legyen $P_i = \emptyset$ $(i = 1 \ldots m)$, $j = 1$.
1. Legyen $i_0 = \min\{i \geq 1,\ l(P_i) + l(T_j) \leq C\}$, $i_0 > m$ esetén vége.
2. Legyen $P_{i_0} = P_{i_0} \cup \{T_j\}$, legyen $j = j + 1$.
3. $j \leq n$ esetén menjünk az 1. lépésre, egyébként vége.

Ha az algoritmus az 1. lépésnél ér véget, akkor túl kicsinek bizonyult az algoritmus számára a C ládaméret, nem sikerült a tárgyakat a ládákba bepakolni. Legyen $\mathrm{FFD}[\mathcal{T}, C, m] = 1$, ha az FFD algoritmusnak sikerül bepakolnia a $\mathcal{T}$-ben lévő tárgyakat m számú C méretű ládába, egyébként pedig legyen $\mathrm{FFD}[\mathcal{T}, C, m] = 0$. (Itt az algoritmus általánosíthatósága kedvéért eltértünk az eredeti jelöléstől.) Legyen

$$(3) \qquad C_1 = \max\left\{\frac{l(\mathcal{T})}{m},\ \max_{1 \leq i \leq n} l(T_i)\right\} \qquad C_2 = \max\left\{2 \cdot \frac{l(\mathcal{T})}{m},\ \max_{1 \leq i \leq n} l(T_i)\right\}$$

Ekkor $C < C_1$ esetén $\mathrm{FFD}[\mathcal{T}, C, m] = 0$, valamint $C \geq C_2$ esetén $\mathrm{FFD}[\mathcal{T}, C, m] = 1$ biztosan teljesül [1] szerint. Legyen

$$r_m(\mathrm{FFD}) = \inf\{r \mid \mathrm{FFD}[\mathcal{T}, r \cdot C^*, m] = 1,\ \forall \mathcal{T}\}$$

Ekkor teljesül a következő alapvető

2. TÉTEL. *Tetszőleges* $\mathcal{T}$ *és* $r \geq r_m$ *esetén* $\mathrm{FFD}[\mathcal{T}, r \cdot C^*, m] = 1$. □

Ez azt jelenti, hogy ha a ládaméret legalább $r_m \cdot C^*$, akkor az algoritmus az összes tárgyat elhelyezi a ládákban. A Multifit algoritmus végrehajtásához először megállapítunk egy alsó és egy felső korlátot a teljes átfutási időre, ezeket jelöljük C_L-lel, illetve C_U-val. Az előbbiek szerint például a $C_L = C_1$, $C_U = C_2$ választás megfelelő. Ezután a $[C_L, C_U]$ intervallum C felezőpontját választjuk a láda méretének, és megpróbáljuk az FFD algoritmus szerint a tárgyakat elhelyezni a ládákban. Ha sikerül, vagyis mindegyik tárgy bekerül valamelyik ládába, akkor C_U szerepét C veszi át, ha nem sikerül, akkor pedig C_L-t helyettesítjük C-vel, és ezt a lépést (egy előre meghatározott k számszor) iteráljuk. Ekkor a Multifit algoritmus formálisan a következőképpen írható le:

A Multifit $[k]$ algoritmus

1. Legyenek $C_L \leq C^* \leq C_U$ az átfutási idő alsó illetve felső becslései, valamint legyen $i := 1$.
2. $i > k$ esetén vége, egyébként $C := (C_U + C_L)/2$
3. Ha FFD $[\mathcal{T}, C, m] = 1$, akkor $C_U := C$,
4. Ha FFD $[\mathcal{T}, C, m] = 0$, akkor $C_L := C$,
5. $i := i + 1$. Menjünk a 2. lépésre.

A Multifit algoritmus elméleti hatékonyságáról [1] az 1.22 felső korlátot bizonyította, vagyis $R_m\big(\text{Multifit}\,[k]\big) \leq 1.22 + 1/2^k$, [2] az előbbi konstanst 1.2-re javította, a módszert javítva $\frac{72}{61}$ pontos felső becslést adott [3].

3. Az általánosított Multifit algoritmus

Az FFD algoritmust fogjuk kicserélni egy később meghatározandó F algoritmussal, amelyre egy általános keretet adunk meg. A tárgyakat a méretük szerinti csökkenő sorrend szerint ütemezzük. Először kiválasztjuk azokat a ládákat, ahova a következő tárgy befér, eztán egy R értékelő függvény segítségével kiválasztjuk ezen ládák közül azt, ahol az értékelő függvény értéke a legnagyobb; és ide tesszük a soron következő tárgyat. Ha ez a maximális függvényérték több helyen is fölvétetik, akkor válasszuk a legkisebb indexű ládát. Formálisan:

Az F algoritmus

1. Legyen $P_i = \emptyset$ $(i = 1 \ldots m)$.
2. Rendezzük a tárgyakat LPT sorrendben, legyen $j := 1$.
3. Legyen $I = \big\{i : 1 \leq i \leq m, l(P_i) + l(T_j) \leq C\big\}$, $I = \emptyset$ esetén vége.
4. Legyen $R(i_0) = \max\big\{R(i) : i \in I\big\}$.
5. Legyen $P_{i_0} = P_{i_0} \cup \{T_j\}$, legyen $j = j + 1$.
6. $j \leq n$ esetén menjünk a 3. lépésre, egyébként vége.

Példák

1. Legyen $R(i) = -l(P_i)$. Ebben az esetben éppen az LPT algoritmust kapjuk, a j-edik tárgy a legelső olyan ládába kerül, ahol a ládában már benne levő tárgyak összmérete minimális.
2. Legyen $R(i) = 1$. Ebben az esetben pedig az FFD algoritmust kapjuk, hiszen a következő tárgy a legelső olyan ládába kerül, ahova befér.

Ezek szerint az általános F algoritmusunk a korábban említett mindkét algoritmusnak általánosítása, pontosabban mindkettő „belefér" ebbe az általános keretbe.

3. TÉTEL. *Tetszőleges $\mathcal{T}$ feladathalmaz, R értékelő függvény és $C < C^*$ esetén* $F[\mathcal{T}, C, m] = 0$.

Bizonyítás. Nyilvánvaló. □

KÖVETKEZMÉNY. *Tetszőleges $\mathcal{T}$ feladathalmaz, R értékelő függvény és $C < C_1$ esetén* $F[\mathcal{T}, C, m] = 0$.

Bizonyítás. Ez abból következik, hogy $C_1 \leq C^*$. □

4. TÉTEL. *Tetszőleges $\mathcal{T}$ feladathalmaz, R értékelő függvény és $C \geq C_2$ esetén* $F[\mathcal{T}, C, m] = 1$.

Bizonyítás. Tegyük fel, hogy $C \geq C_2$, és $F[\mathcal{T}, C, m] = 0$. Ez azt jelenti, hogy az F algoritmus során elérkezünk egy olyan állapothoz, hogy maradt még a ládákban el nem helyezett tárgy, de ez már semelyik ládába nem fér be. Feltehető, hogy már csak egyetlen tárgy maradt elhelyezetlenül, különben a többit elhagyjuk a $\mathcal{T}$ feladathalmazból, és a maradékra még mindig teljesül a kezdeti feltétel. Ekkor tehát

$$l(P_i) + l(T_n) > C \quad i = 1, \ldots, m$$

ahol T_n az utolsó, el nem helyezett tárgy. Az egyenlőtlenségeket összegezve:

$$\sum_{i=1}^{m} l(P_i) + m \cdot l(T_n) > m \cdot C.$$

Ha a bal oldalhoz $l(T_n)$-t hozzáadunk, (ami nemnegatív), az egyenlőtlenséget m-mel osztva a következőt kapjuk:

$$\frac{l(\mathcal{T})}{m} + l(T_n) > C \geq C_2 \geq \frac{2 \cdot l(\mathcal{T})}{m}. \tag{4}$$

Ebből kapjuk: $l(T_n) > \frac{l(\mathcal{T})}{m}$, ami lehetelen, hiszen $l(T_n) \leq l(T_i)$ $i = 1, \ldots, n$ esetén az LPT sorrend miatt, így az egyenlőtlenségeket összegezve $n \cdot l(T_n) \leq \sum_{i=1}^{n} l(T_i) = l(\mathcal{T})$, s így n-nel osztva kapjuk $l(T_n) \leq \frac{l(\mathcal{T})}{n} < \frac{l(\mathcal{T})}{m}$. □

Ezek alapján most is meg tudunk adni alsó és felső korlátot a C átfutási időre, amelynél kisebb, illetve amelynél nagyobb értékek esetén $F[\mathcal{T}, C, m] = 0$, illetve $F[\mathcal{T}, C, m] = 1$ biztosan teljesül. Az algoritmusunk általános kerete így a következő lesz:

A GMF $[k]$ (általánosított Multifit) algoritmus

1. Legyenek C_L és C_U az átfutási idő alsó, és felső becslései: $C_L \leq C^* \leq C_U$, legyen $i := 1$.
2. $i > k$ esetén vége, egyébként $C := (C_U + C_L)/2$
3. Ha $F[\mathcal{T}, C, m] = 1$, akkor $C_U := C$,
4. Ha $F[\mathcal{T}, C, m] = 0$, akkor $C_L := C$,
5. $i := i + 1$. Menjünk a 2. lépésre.

F1 Feltevés. Legyenek $\alpha, \beta \in R$ olyan számok, amelyekre teljesül, hogy tetszőleges $X \subseteq \mathcal{T}$ esetén az $l(X) \leq \alpha$ és $l(X) \leq \beta$ feltételek közül vagy mindkettő, vagy egyik sem teljesül. (Ez azt jelenti, hogy ha valamely tárgyak befférnek egy α méretű ládába, akkor a β méretű ládába is befférnek, és fordítva.) Ekkor az α vagy β ládaméretek esetén alkalmazva az $\mathcal{F}$ algoritmust, a tárgyaknak ugyanazt az elhelyezését kapjuk.

Megjegyzés. Az előbbi F1 feltevés nyilvánvalóan igaz akkor, ha az R értékelő függvény értéke nem függ a ládamérettől, csak attól, hogy mely tárgyak vannak már az egyes ládákban. Emiatt F1 mindkét korábbi, speciális esetként tárgyalt algoritmus esetében fennáll, (nevezetesen az LPT és a Multifit algoritmus esetében is), hiszen az értékelő függvény értéke nem függ a ládamérettől. A továbbiakban feltesszük, hogy teljesül az F1 feltevés.

1. Definíció. Tetszőleges F ládapakolási algoritmus esetén definiáljuk a következő mennyiséget:

$$r_m(F) = \inf \left\{ r \mid F[\mathcal{T}, r \cdot C^*, m] = 1, \forall \mathcal{T} \right\}$$

Az r_m szám tehát az a legkisebb szorzó, amennyivel megszorozva a C^* értéket a kapott méretű m számú ládába az F algoritmus biztosan be tudja pakolni a téglalapokat. Könnyen látszik, hogy r_m definíciója értelmes, hiszen $C^* \geq \frac{1}{m} \cdot l(\mathcal{T})$, ugyanis ha a gépekre egyenletesen sikerülne elhelyezni a munkákat, akkor kapnánk ezt az értéket, ekkor viszont $m \cdot C^* \geq l(\mathcal{T})$, vagyis $r = m$ választással minden munka az első gépre kerül, vagyis nem üres számhalmaz infimumát vesszük.

5. TÉTEL. *Tegyük fel, hogy teljesül az F1 feltevés. Ekkor tetszőleges $\mathcal{T}$ és $r \geq r_m$ esetén $F[\mathcal{T}, r \cdot C^*, m] = 1$.*

Bizonyítás. Először belátjuk, hogy $F[\mathcal{T}, r_m \cdot C^*, m] = 1$. Tegyük fel ezzel ellentétben, hogy az F algoritmus nem képes a tárgyakat a $C = r_m \cdot C^*$ méretű, m számú ládába bepakolni, vagyis legalább egy tárgy kimarad a ládákból. Legyen $C_\alpha = \min \left\{ l(X), l(X) > C, X \subseteq \mathcal{T} \right\}$. (Emlékeztetünk arra, hogy $l(X) = \sum_{T \in X} l(T)$ az

X-beli tárgyak összméretét jelenti, tetszőleges $X \subseteq \mathcal{T}$ részhalmaz esetén.) Ekkor tetszőleges C' esetén, ahol $C \leq C' < C_\alpha$, az $F1$ feltevés teljesülése folytán az F algoritmus a tárgyakat a C' méretű ládákba sem képes bepakolni, (mert minden ládába pontosan ugyanazok a tárgyak kerülnek, mint az előbb), ez pedig ellenmond r_m értelmezésének. Most lássuk be, hogy $F[\mathcal{T}, r \cdot C^*, m] = 1$ teljesül tetszőleges $r > r_m$ esetén. Tegyük fel tehát, hogy $F[\mathcal{T}, r \cdot C^*, m] = 0$, valamely $\mathcal{T}$ feladathalmaz esetén, legyen $r = \alpha \cdot r_m$, ekkor $\alpha > 1$. Tekintsünk egy tetszőleges optimális pakolást m darab, C^* méretű ládába. Növeljük meg a ládák méretét $\alpha \cdot C^*$-ra, és töltsük ki az összes ládában a megmaradt helyeket elegendően kicsiny tárgyakkal úgy, hogy valamennyi láda teljesen meg legyen töltve, és az új tárgyak közül mindegyik mérete legyen kisebb a korábban létező tárgyak közül bármelyiknek a méreténél. A kapott feladathalmazt jelöljük $\bar{\mathcal{T}}$-vel. Ekkor $\bar{C}^* = \alpha \cdot C^*$, ahol $\bar{C}^*$ az új feladathalmaz esetén a teljes átfutási idő értéke. Hajtsuk végre az F algoritmust az $r \cdot C^*$ ládamérettel és $\bar{\mathcal{T}}$ feladathalmazzal. Mivel az újólag hozzávett tárgyak mérete az előzőleg létezőknél kisebb, ezért az F algoritmus előbb a régebbi tárgyakat próbálja elhelyezni a ládákban, azokat viszont nem tudja mindet elhelyezni, hiszen feltettük, hogy $F[\mathcal{T}, r \cdot C^*, m] = 0$, ekkor viszont ebből $F[\bar{\mathcal{T}}, r \cdot C^*, m] = 0$ következik. Mivel $r \cdot C^* = \alpha \cdot r_m \cdot C^* = r_m \cdot \bar{C}^*$, kapjuk: $F[\bar{\mathcal{T}}, r_m \cdot \bar{C}^*, m] = 0$, ez pedig ellentmond az előzőleg már belátott résznek. □

6. TÉTEL. $R_m\big(\mathrm{GMF}\,[k]\big) \leq r_m(F) + \left(\frac{1}{2}\right)^k$

Bizonyítás. Megegyezik az eredeti esetben történő bizonyítással [3]. □

Ezek után az $r_m(F)$ konstansra szeretnénk „jó" felső becslést kapni, hiszen így egyben az $R_m\big(\mathrm{GMF}\,[k]\big)$ expanziós faktort is meg tudtuk becsülni.

3.1. Hatékonysági becslés

Ebben a részben belátjuk, hogy az általános F algoritmusunk elméleti hatékonyságára tetszőleges $m \geq 2$ esetén teljesül az $r_m(F) \leq \frac{4}{3} - \frac{1}{3 \cdot m}$ becslés, ha az F1 és a később ismertetendő F2 feltevések teljesülnek. A becslés éles, hiszen megegyezik a Graham által kapott, az LPT algoritmusra vonatkozó konstanssal, ami ennek az algoritmusnak speciális esete, és ott a becslés éles. A tétel bizonyítását több lépésen keresztül végezzük el.

2. Definíció. Legyen p, q tetszőleges pozitív egész szám, ahol $p \geq q$. Egy (p/q) ellenpéldán olyan $(\mathcal{T}, M)$ rendezett párt értünk, (ahol $\mathcal{T}$ egy feladathalmaz, M pedig egy pozitív szám, a ládák száma), amelyekre teljesülnek

$$F[\mathcal{T}, p, M] = 0, \qquad C^*[\mathcal{T}, M] \leq q, \tag{5}$$

vagyis az F algoritmus nem képes a tárgyakat M számú p méretű ládába elhelyezni, de van q-nál nem nagyobb teljes átfutási idővel rendelkező ütemezés.

3. Definíció. Minimális (p/q) ellenpéldán olyan $(\mathcal{T}, M)$ rendezett párt értünk, amely minimális a következő értelemben:

a, A $(\mathcal{T}, M)$ pár (p/q) ellenpélda
b, tetszőleges a, -t kielégítő $\mathcal{T}'$ esetén $|\mathcal{T}'| \geq |\mathcal{T}|$
c, tetszőleges $1 \leq m < M$ esetén $r_m \leq p/q$.

7. LEMMA. *Legyen $(\mathcal{T}, M)$ minimális (p/q) ellenpélda. Ekkor egyetlen, a sorrendben utolsó feladat marad ütemezetlenül, vagyis $F[\mathcal{T} \setminus \{T_n\}, p, M] = 1$. Az utolsó tárgy már egyik ládába sem fér.*

Bizonyítás. Ellenkező esetben az ellenpéldánk nem lenne minimális. □

4. Definíció. Azt mondjuk, hogy $X, Y \subseteq \mathcal{T}$ esetén X dominálja Y-t, ha létezik olyan $f : Y \to X$ injektív leképezés, amelyre $l(f(y)) \geq l(y)$ tetszőleges $y \in Y$ esetén.

Ezután feltesszük, hogy teljesül a következő

F2 Feltevés. Tegyük fel, hogy az F algoritmust m ládára alkalmazva a tárgyaknak a következő elhelyezését kapjuk: $\mathcal{P} = \langle P_1, \ldots, P_m \rangle$. Ekkor az F algoritmust $m-1$ ládára alkalmazva a $\mathcal{T}' = \mathcal{T} \setminus P_\alpha$-beli tárgyaknak a következő elhelyezését kapjuk: $\mathcal{P}' = \langle P_1, \ldots, P_{\alpha-1}, P_{\alpha+1}, \ldots, P_m \rangle$ tetszőleges $1 \leq \alpha \leq m$ egész szám esetén, vagyis a többi ládába ugyanazok a tárgyak kerülnek, mint az előbb.

Megjegyzés. Ez az előbbi feltevés is teljesül mind az LPT, mind az eredeti Multifit algoritmus esetében is.

8. TÉTEL (Főtétel). *Legyen $m \geq 2$ tetszőleges egész. Tegyük fel, hogy az* F *algoritmusra teljesülnek az* F1 *és* F2 *feltevések. Ekkor $r_m(F) \leq \frac{4}{3} - \frac{1}{3 \cdot m}$. A becslés ilyen általánosságban nem javítható.*

Bizonyítás. A bizonyítást az alábbi Lemmákon keresztül végezzük el, közben feltesszük, hogy a 8. Tétel feltevései teljesülnek, (vagyis F1 és F2 érvényben van).

9. LEMMA. *Legyen $(\mathcal{T}, M)$ minimális (p/q) ellenpélda. Legyen $i, j \in \{1, \ldots, M\}$. Legyen $\mathcal{P} = \langle P_1, \ldots, P_M \rangle$ az* F *algoritmus által történő elhelyezése a tárgyaknak, (ahol az utolsó tárgy kimarad a ládákból), $\mathcal{P}^* = \langle P_1^*, \ldots, P_M^* \rangle$ pedig egy tetszőleges optimális pakolás. Ekkor a P_i halmaz nem dominálja a P_j^* halmazt.*

Bizonyítás. Tegyük fel, hogy az állítás nem igaz, vagyis $\exists f : P_j^* \to P_i$ injektív függvény, amelyre $l(f(y)) \geq l(y)$ $\forall y \in P_j^*$ esetén. Legyen $\mathcal{T}' = \mathcal{T} \setminus P_i$. Ekkor az $F2$ feltevés teljesülése folytán az F algoritmus a $\mathcal{T}'$-beli tárgyakat pontosan ugyanúgy helyezi el $M-1$ ládába, mint az előbb. Másrészt tekintsük a $\mathcal{P}^*$ ütemezést, és konstruáljunk ebből egy $\mathcal{P}'$ ütemezést úgy, hogy minden $y \in P_j^*$ tárgyat cseréljünk ki az ő $f(y)$ képével. Ekkor a P_j'-beli tárgyak mindegyike az P_i halmazhoz tartozik, $l \neq j$ esetén a P_l'-beli tárgyak mérete csak csökkenhetett, ezért az P_i-beli tárgyakat elhagyva $C^*[\mathcal{T}', M-1] \leq q$, másrészt $F[\mathcal{T}', p, M-1] = 0$, ez pedig ellentmond annak, hogy a $(\mathcal{T}, M)$ pár minimális (p/q) ellenpélda. □

10. LEMMA. $|P_j^*| \geq 2$, *minden $1 \leq j \leq M$ esetén.*

Bizonyítás. Ha $P_j^* = \{x\}$ lenne valamilyen $x \in \mathcal{T}$ esetén, akkor ez az x tárgy benne van valamelyik P_i ládában, de ekkor $\{P_i\}$ dominálja $\{P_j^*\}$-ot, ami ellentmond az előző Lemmának. □

11. LEMMA. $|P_i| \geq 2$, *minden* $1 \leq i \leq M$ *esetén.*

Bizonyítás. Tegyük fel, hogy $P_i = \{x\}$ valamely i-re. Ekkor $l(x) + l(T_n) > p$, ahol $l(T_n)$ az utolsó, ezért legrövidebb, nem ütemezett tárgy hossza. Ebből $l(x) + l(y) > p > q$ következik tetszőleges $y \neq x$, $y \in \mathcal{T}$ esetén, így az x tárgy mellé az optimális ütemezésben sem kerülhetett be a ládájába más tárgy, ami ellentmond az előző Lemmának. □

12. LEMMA. *Legyen* $(\mathcal{T}, M)$*tetszőleges* (p/q) *ellenpélda, legyen* $\alpha = l(T_n)$. *Ekkor* $\alpha > \frac{M}{M-1}(p-q)$.

Bizonyítás. A T_n tárgy már nem fér be a P_i ládába, ezért $l(P_i) + \alpha > p$ minden $1 \leq i \leq M$ esetén, ezt i-re összegezve kapjuk:

$$\sum_{i=1}^{M} l(P_i) + M\alpha > Mp. \tag{6}$$

Másrészt

$$\sum_{i=1}^{M} l(P_i) + \alpha \leq Mq,$$

hiszen az optimális ütemezés során minden tárgy befért az M számú q kapacitású ládába. Ezeket rendezve:

$$(M-1)\alpha > M(p-q),$$

ebből pedig a kívánt állítás adódik. □

Az $r_m \leq \frac{4}{3} - \frac{1}{3 \cdot m}$ *becslés bizonyítása.* Legyen m a ládák száma, legyen $p = 4m-1$, és $q = 3m$. Állítjuk, hogy ezekre a p és q számokra nem létezik (p/q) ellenpélda, vagyis ebből már következik, hogy $r_m \leq \frac{4m-1}{3m} = \frac{4}{3} - \frac{1}{3 \cdot m}$. Tegyük fel, hogy az állítással ellentétben a $(\mathcal{T}, m)$ pár (p/q) ellenpélda. A 12. Lemma alapján most $\alpha > \frac{m}{m-1}(p-q) = \frac{m}{m-1}(m-1) = m$. Így semelyik optimális ládába sem fér kettőnél több tárgy, (mert ha legalább három tárgy lenne valamely ládában, akkor az összméretük meghaladná a $3m$-et.) Így a 10. lemmát fölhasználva $|P_j^*| = 2$ minden j-re, és ezért $n = 2m$. Ekkor viszont $|P_i| = 2$ minden i-re, hiszen a 11. Lemma szerint $|P_i| \geq 2$ szintén teljesül, ekkor viszont nem maradt ki tárgy a ládákból, ami ellentmondás. □

Megjegyzés. Az előző Tétel szerint a Graham által kapott $R_m(\mathrm{LPT}) = \frac{4}{3} - \frac{1}{3 \cdot m}$ hatékonysági becslés sokkal általánosabb algoritmikus keretben is teljesül.

4. Néhány új algoritmus, numerikus eredmények

4.1. Új algoritmusok

1. Cseréljük ki az FFD algoritmust a BFD (Best Fit Decreasing) algoritmusra. A BFD algoritmus a tárgyakat csökkenő méret szerinti sorrendben helyezi el, a soron következő tárgy abba a ládába kerül, ahova befér, és amelyik láda ezáltal a lehető leginkább megtelik. Ha az általánosított algoritmikus keretünk esetén az R értékelő függvény a következő: $R(i) = l(P_i)$, akkor éppen ezt az algoritmust kapjuk. Az alábbi példa mutatja, hogy van olyan $\mathcal{T}$ feladathalmaz, amikor BFD, illetve van olyan, amikor az FFD algoritmus ad jobb megoldást: A $\mathcal{T} = [15, 8, 8, 5, 4, 2, 2]$ esetén $\mathrm{FFD}[\mathcal{T}, 22, 2] = 1$, de $BFD[\mathcal{T}, 22, 2] = 0$, a BFD algoritmus csak 23 egység magasságba képes bepakolni a téglalapokat. Másrészt $\mathcal{T} = [15, 8, 8, 5, 4, 2]$ esetén $\mathrm{FFD}[\mathcal{T}, 21, 2] = 0$, az FFD algoritmusnak 22 egység magasságra lenne szüksége, (éppen úgy, mint az előbb), azonban $BFD[\mathcal{T}, 21, 2] = 1$. Könnyen belátható, hogy az $F1$ és $F2$ tulajdonságok teljesülnek, (az $F1$ azért teljesül, mert R nem függ a ládamérettől), ezért az algoritmus elméleti hatékonysága legfeljebb $\frac{4}{3} - \frac{1}{3 \cdot m}$.

Néhány új ládapakolási algoritmust is bevezetünk. Mindegyik csökkenő méret szerint helyezi el a tárgyakat.

2. Legyen l tetszőleges egész, amelyre $1 \leq l \leq n$. helyezzük el az első l számú tárgyat az LPT szabály szerint, a többit pedig a BFD algoritmus szerint. Nevezzük ezt $\mathcal{F}(l)$ algoritmusnak. Nyilván $l = n$ esetén az LPT, $l = 1$ esetén a BFD algoritmust kapjuk, egyébként vagyis $l \neq 1$, $l \neq n$ esetén pedig új algoritmust kapunk. Könnyen látható, hogy az $F1$ és $F2$ feltevések teljesülnek, ezért érvényes az általános esetre vonatkozó hatékonysági becslés.

3. Legyen $1 \leq l \leq n$ rögzített szám. A tárgyakat az LPT szabály szerint helyezzük el addig, amíg mindegyik ládában lesz legalább l darab tárgy, ezután a maradék tárgyakat a BFD algoritmus szerint.

Az előző három algoritmus esetében az $F1$ és $F2$ feltevések azért teljesülnek, mert azok teljesülnek az LPT és a BFD algoritmusok esetében, és az előbbiek ez utóbbiakból lettek összekombinálva egy előre rögzített átváltási kritérium szerint.

4. A következő algoritmus alapötlete az, hogy a soron következő tárgyat helyezzük oda, ahol a ládák alsó vagy felső széléhez a lehető legközelebbre kerül. Pontosabban ezen azt értjük, hogy az elhelyezendő téglalap aljának a nagy téglalap aljától, vagy pedig az elhelyezendő téglalap tetejének a C magasságú nagy téglalap tetejétől mért távolsága legyen minimális; vagy ami ugyanezt jelenti, az elhelyezendő téglalap tetejének a nagy téglalap aljától, vagy pedig az elhelyezendő téglalap aljának a C magasságú nagy téglalap tetejétől mért távolsága legyen maximális. Ekkor az R értékelő függvény a következő: $R(i) = \max\{l(P_i) + l(T_j), C - l(P_i)\}$. Ekkor azonban belátható, hogy $F1$ nem teljesül. Ezért az algoritmust egy kicsit módosítjuk: Legyen $C_\alpha = \min\{l(X), l(X) > C, X \subseteq \mathcal{T}\}$, (az 5. Tétel bizonyításában szereplő szám.) Helyettesítsük a C számot C_α-val: $R(i) = \max\{l(P_i) + l(T_j), C_\alpha - l(P_i)\}$. Ekkor már R nem függ a ládamérettől, ezért $F1$ teljesül, és be-

látható, hogy $F2$ is teljesül, és ezért érvényes a hatékonysági becslés. Nevezzük az előbbi algoritmust F_{max} algoritmusnak.

4.2. Numerikus eredmények

Jelöljük az előbbi ládapakolási algoritmusokkal végrehajtott általános Multifit típusú algoritmusokat a belső algoritmusok szerint a következő jelölésekkel: LPT, FFD (az eredeti Multifit belső algoritmusa), BFD, $F(l)$ a 2. algoritmus, $F_{\max}$ a 4. algoritmus. Az első táblázatban az LPT, a FFD, és a BFD algoritmusokkal végrehajtott Multifit típusú módszert hasonlítottuk össze. A teszt során 10 000 egymástól független kísérletet végeztünk különböző feladatosztályokban, és azt számoltuk, hogy melyik algoritmus hányszor adott a három között legjobb eredményt. A téglalapok hosszúságait a „par" intervallumból választottuk egyenletes eloszlás szerint, kerekítéssel.

	m	n	*par*	LPT	FFD	BFD
1.	3	15	[1,30]	5808	8541	8620
2.	3	15	[1,60]	4467	7876	8029
3.	3	15	[15,30]	9214	3435	3435
4.	5	20	[1,30]	3932	9287	9394
5.	5	20	[1,60]	2736	8806	9049
6.	5	20	[40,60]	9994	9	9
7.	5	19	[30,60]	5608	5781	5781

A BFD-vel végrehajtott Multifit algoritmus egyetlen feladatosztályban sem volt rosszabb, mint a hagyományos Multifit, bizonyos esetekben azonban lényegesen jobb volt. Az 1. feladatosztályban elég nagy a különbség, a BFD javára. A következő feladatosztály esetén a BFD előnye megmarad, viszont az LPT algoritmus a téglalapok hosszainak a növekedése miatt kevésbé hatékony mint az előbb. A 3. feladatosztály az előzőektől abban különbözik, hogy a téglalapok hosszabbak lettek, a minimális hosszúság a maximálisnak a fele. Most az LPT algoritmus lényegesen megelőzte a másik kettőt, és érdekes módon e másik két algoritmus egyenlő számban adott minimális eredményt, csaknem azonosan működnek. A következő három feladatosztályban, ahol $m = 5$ és $n = 20$, az előző háromhoz hasonló eredményeket kaptunk. A különbség az előző esettel ($m = 3$ és $n = 15$) szemben az, hogy az LPT és a másik két algoritmus által kapott eredmények közötti rés nagyobb lett. Az utolsó feladatosztályban a tárgyak száma nem osztható a gépek számával. Ennek az a hatása az algoritmusokra, hogy az LPT algoritmus a másik kettőhöz képest kevésbé jó mint az előbb, és megint azt kaptuk, hogy a BFD illetve az FFD algoritmusok lényegében egyformán működtek. A teszt eredményeit összefoglalva általánosságban az mondható, hogy ha az FFD jobban működik az LPT algorit-

musnál, akkor a BFD algoritmus még az FFD-nél is jobb; ellenkező esetben pedig a BFD és FFD algoritmusok lényegében azonosan működnek.

A következő tesztekben az LPT, BFD és $F(l)$, illetve az LPT, BFD és $F_{\max}$ algoritmusokat hasonlítottuk össze. A teszt során 10 000 független kísérletet végeztünk különböző feladatosztályokban, és azt számoltuk, hogy melyik algoritmus hányszor adott a három között legjobb eredményt, hányszor volt ennek legfeljebb 1.05-szerese, illetve a $\frac{C}{C_{lb}}$ értékek átlagát, ahol C a heurisztikus megoldás által meghatározott átfutási idő, C_{lb} pedig egy javított alsó becslés, a [8]-ban szereplő alsó becslések maximuma.

		LPT	BFD	$F(l)$	l/n	
	$m = 2$	0	5563	7440	0.8	min
1.	$n = 81$	10000	10000	10000		$1.05 \cdot$ min
	[10,20]	1.0074	1.0023	1.0011		C/C_{lb}
	$m = 4$	9587	38	9975	0.77	min
2.	$n = 16$	9943	38	9996		$1.05 \cdot$ min
	[40,60]	1.0051	1.062	1.0049		C/C_{lb}
	$m = 3$	9258	4	9997	0.83	min
3.	$n = 15$	9997	4	9999		$1.05 \cdot$ min
	[120,170]	1.0035	1.056	1.0030		C/C_{lb}

		LPT	BFD	$F_{\max}$	
	$m = 10$	0	4080	8841	min
4.	$n = 32$	0	4080	8841	$1.05 \cdot$ min
	[11,28]	1.121	1.036	1.024	C/C_{lb}
	$m = 2$	0	85	9941	min
5.	$n = 93$	9918	9999	10000	$1.05 \cdot$ min
	[60,70]	1.009	1.007	1.0009	C/C_{lb}
	$m = 3$	0	90	9950	min
6.	$n = 61$	143	5485	10000	$1.05 \cdot$ min
	[60,70]	1.029	1.024	1.014	C/C_{lb}

A tesztek szerint a vizsgált feladatosztályokban az $F(l)$, illetve az $F_{\max}$ algoritmusok mind a minimumok elérésének számában, mind a $\frac{C}{C_{lb}}$ arányok tekintetében

jobbak, illetve lényegesen jobbak a másik két algoritmusnál. Érdemes megjegyezni, hogy itt nem az FFD, hanem az annál hatékonyabbnak bizonyult BFD-t előzte meg a két új algoritmust. Az első négy eset extremális abban az értelemben, hogy az LPT és a BFD algoritmusok közül az egyik a másikhoz képest elenyésző számban adott jó eredményt. Az utolsó két feladatosztályban pedig mindkettő elenyésző számban adott jó eredményt az $F_{\max}$ algoritmussal szemben.

Köszönetnyilvánítás. A szerző megköszöni Vizvári Béla segítségét a cikk szerkezetének a kialakításában, és a cikk felépítésében.

IRODALOM

[1] E. E. Coffman, Jr., M. R. Garey, D. S. Johnson, An application of bin-packing to multiprocessor scheduling, *SIAM J. Comput.*, **7** (1978), 1–17.

[2] D. K. Friesen, Tighter bounds for the multifit processor scheduling algorithms, *SIAM J. Comput.*, **13** (1984), 170–181.

[3] D. K. Friesen, M. A. Langston, Evaluation of a MULTIFIT-based scheduling algorithm, *J. Algorithms*, **7** (1986), 35–59.

[4] R. L. Graham, Bounds for certain multiprocessor anomalies, *Bell System Tech. J.*, **45** (1966), 1563–1581.

[5] R. L. Graham, Bounds on multiprocessor timing anomalies, *SIAM J. Appl. Math.*, **17** (1969), 416–429.

[6] Racsmány Anna, *Ütemezéselmélet* (Egyetemi jegyzet, MKKE, 1981).

[7] Zs. Soós, L. Varga, "Greedy Algorithm for Maintenance Scheduling in Electric Power System" in: *Proc. of the Fifth ECMI Conf.*, Ed. M. Heiliö (B. G. Teubner and Kluwer, 1991), 115-118.

[8] B. Vizvári, R. Demir, R., "It is Difficult to Find a Difficult Problem for the Scheduling of Identical Parallel Machines", Department of Industrial Engineering of Bilkent University, Research Report, IEOR-9212, 1992.

[9] B. Vizvári, R. Demir, "A Coloumn Generation Algorithm to Schedule Identical Parallel Machines", Rutgers University, Research Report, RRR-99-93, September 1994.

[10] B. Vizvári, *Bevezetés a termelésirányítás matematikai elméletébe* (Egyetemi jegyzet, ELTE, 1991).

(Beérkezett: 1999. július 20.)

DÓSA GYÖRGY
VESZPRÉMI EGYETEM
8201 VESZPRÉM
PF. 158
E-mail: dosagy@almos.vein.hu

GENERALIZED MULTIFIT-TIPE METHODS FOR SCHEDULING PARALLEL IDENTICAL MACHINES

GYÖRGY DÓSA

We investigate a very known NP-complete problem of the scheduling-theory: scheduling of parallel independent machines, that is: How to distribute n tasks among $m < n$ machines as to minimize the overall finishing time. We give a common generalization of two heuristical methods called LPT due to Graham [1] and the method called Multifit. We change the algorithm FFDH (which is an inner part of Multifit) with other algorithms, whereas the theoretical upper bound of the algorithm is the same as the upper bound of LPT. Numerical aspects also investigated: the results show that the new method often give better solutions as the others.

Alkalmazott Matematikai Lapok **19** (1999) 169–183

EGY TRANSZPORTMODELL ALKALMAZÁSA A GYÁL TÉRSÉGÉBEN LÉTESÍTENDŐ HULLADÉKLERAKÓ ESETLEGES TALAJSZENNYEZŐ HATÁSÁNAK VIZSGÁLATÁRA (ESETTANULMÁNY)

KÉRI GERZSON, ORSOVAI IMRE ÉS RAPCSÁK TAMÁS

Budapest

Új üzemek létesítése során ma már alapvető követelmény – a maximális biztonságra való törekvés mellett – a létesítés megkezdése előtt tervet kidolgozni váratlan üzemzavarok és azok következtében fellépő károk kezelésére (havariaterv). Egy hulladéklerakó esetén az egyik lehetséges havariaeset a lerakó szigetelésének megsérülése, és ennek következtében szennyezőanyagoknak a talajba kerüléséből eredő környezetszennyezés. Mivel a szigetelésre nagyon biztonságos technológiát alkalmaznak, ezért szennyezőanyagok talajvízzel történő kijutásának a valószínűsége rendkívül kicsi, de teljes bizonyossággal nem zárható ki.

A szennyezőanyagok talajban történő terjedésére alkalmazható, irodalomból ismert modelleket használva fel tudjuk becsülni, hogy olyan esetben, ha egy adott időpontban mégis megsérülne a hulladéklerakó szigetelése, akkor ennek következményeként különböző irányokban és távolságokban milyen mértékben szennyeződne a talaj. Ez az esettanulmány a Gyál térségében létesítendő regionális hulladéklerakóra vonatkozó számításokról számol be. Számításainkban a szennyezőanyag-koncentrációnak térben és időben való változását követjük, természetesen nem valódi, hanem feltételezett adatokkal, hiszen a lerakó még csak ezután fog megépülni, és remélhetőleg az elkészülése után sem fog szennyezőanyag a lerakón kívüli térségbe kerülni.

1. Bevezetés

Az A.S.A. Környezetvédelem és Hulladékgazdálkodás Magyarország Kft. megbízásából az MTA SZTAKI Operációkutatás és Döntési Rendszerek Osztálya és a vele alvállalkozóként együttműködő GEOÖKOTERV Környezetföldtani Kutató és Tervező Kft. 1997-ben elkészített egy környezeti hatástanulmányt, amely egy regionális hulladéklerakó Gyál térségében történő megvalósításával kapcsolatban felmerülő környezeti szempontokat elemzi és értékeli. Ezt megelőzően ugyancsak az MTA SZTAKI és a GEOÖKOTERV készítette el az ugyanerre a térségre vonatkozó kommunális hulladék elhelyezésének kérdéseit vizsgáló döntéselőkészítő tanulmányt.

A környezeti hatástanulmány 9. fejezete a tervezett hulladéklerakó üzemeltetése során esetleg fellépő havariaesetek (azaz katasztrófának még nem minősíthető,

de a környezetre veszélyes káros jelenségek) előfordulásának lehetőségeit tárgyalja. A hulladéklerakó építése és üzemeltetése során alkalmazandó technológia biztosítja, hogy bármilyen havariaeset bekövetkezésének a valószínűsége nagyon kicsi, ennek ellenére a tanulmány kitér az elképzelhető havariaesetekre, és vázolja az ilyen, kis valószínűséggel esetleg mégis bekövetkező havariaesetek kezelésének (a hiba megszüntetésének és a káros következmények felszámolásának vagy mérséklésének) módját.

A lehetséges havariaesetek közül a hatástanulmány három fajtát emel ki. Ezek: a lerakott hulladék begyulladása, rendkívüli csapadék vagy földrengés okozta rézsűcsúszás, ill. a szigetelés meghibásodása esetén fellépő környezeti talajszennyezés. Ebben a cikkben az utóbbival foglalkozunk, nevezetesen: a talajszennyeződés egy matematikai modelljével, annak megoldási módszereivel és a konkrét esetre vonatkozó matematikai számításokkal. Meg kell itt jegyeznünk, hogy egy hulladéklerakóból származó szennyezőanyagoknak a külső talajba kerüléséből és tovaterjedéséből akkor származhatnak komolyabb problémák, ha a szennyezés élővizeket veszélyeztet. A szennyeződéstől főleg az ivóvíztermelő kutakat kell félteni.

Bár a modell paramétereinek a konkrét esetnek megfelelő számszerűsítése bizonyos nehézségekbe ütközött, ugyanis a modell számításainak az elvégzéséhez szükséges adatok többsége nem állt rendelkezésünkre, mégis úgy ítéltük meg a kérdést, hogy az ilyen adatok vonatkozásában a leggyakoribb jellemző értékekkel számolva is tájékoztató értékű, tehát használható eredményeket kapunk, ezeket azonban a jelenség természeténél fogva az általában megkívánt mérnöki pontosságnál kevésbé pontosnak, tehát valóban csak tájékoztató jellegű adatoknak kell tekintenünk.

Elméletileg megvolna a lehetőség arra, hogy kísérletek és mérések sorának az elvégzésével, majd a mért adatoknak (pl. szennyezőanyag-koncentrációknak) és a modell alapján számított adatoknak az összehasonlításával a modell paramétereit kalibráljuk, ehhez azonban a rendelkezésre álló idő is kevés volt, és a kísérleti fúrások és mérések anyagi fedezete sem állt rendelkezésre.

A szennyezőanyag-terjedés különböző modelljeit részletesen tárgyalja Kovács Balázs és Szabó Imre [1] munkája. A szerzők felsorolják a transzportfolyamat különböző komponenseit (konvekció, diffúzió, diszperzió, adszorpció és degradáció), ezek összeillesztésével megfogalmazzák az általános transzportegyenlet egy-, ill. kétdimenziós változatát, és ismertetik a transzportegyenletek jó néhány megoldási módszerét. Ezért a tervezett Gyál környéki hulladéklerakó esetére nem volt szükség külön modell és algoritmus kidolgozására, választhattunk az [1] munkában szereplő modellek és megoldási módszerek közül. Ennélfogva a jelen publikáció lényeges elméleti újdonságot nem tartalmaz, csupán néhány apró részletben tértünk el az [1] munkában javasolt számítási módszerektől. Mégis úgy gondoltuk, hogy tanulságos lehet az elméletnek a konkrét esetre vonatkozó számításait egy esettanulmányban ismertetni, a modell számításaival járó buktatók ecsetelésével együtt.

2. A transzportegyenletek és paramétereik

A szennyezőanyag-terjedés törvényszerűségeit matematikailag – közelítő módon – a transzportegyenletek írják le. Ezek ismeretével lehetőség nyílik arra, hogy meghatározzuk az esetleges szennyezőforrások hatását és a szennyezőanyag koncentrációjának térben és időben való eloszlását. Ennek alapján pedig hatékonyabbá válhat a szennyezőanyag további terjedésének megállítása és a szennyeződés eltávolítása. A transzportegyenletek megoldására analitikus módszereket, véges differencia módszereket, végeselem módszereket, valamint különböző szemianalitikus eljárásokat lehet alkalmazni, amint ezt Kovács Balázs és Szabó Imre az [1] munkában kifejtik.

A transzportegyenletek matematikai értelemben a C (szennyezőanyag-)koncentrációra vonatkozó másodrendű parciális differenciálegyenletek.

2.1. Az egydimenziós transzportegyenlet alakja:

$$\frac{\partial C}{\partial t} + \frac{v}{R} \cdot \frac{\partial C}{\partial x} = \frac{\alpha_L v}{R} \cdot \frac{\partial^2 C}{\partial x^2} - \lambda C, \tag{2.1}$$

ahol v a szivárgás sebessége a pórusokban, α_L a longitudinális diszperzivitás, R a késleltetési tényező, λ a bomlási együttható, t az idő-, x a szivárgás irányában meghatározott térkoordináta.

2.2. A kétdimenziós transzportegyenlet alakja:

$$\frac{\partial C}{\partial t} + \frac{v}{R} \cdot \frac{\partial C}{\partial x} = \frac{\alpha_L v}{R} \cdot \frac{\partial^2 C}{\partial x^2} + \frac{\alpha_T v}{R} \cdot \frac{\partial^2 C}{\partial y^2} - \lambda C, \tag{2.2}$$

ahol v, α_L, R, λ, t és x ugyanazok, mint az egydimenziós esetben, α_T a transzverzális diszperzivitás, y pedig a második, tehát a szivárgás irányára merőleges térkoordináta.

A transzportegyenletek részletes levezetése [1]-ben megtalálható. A konkrét esetre leginkább megfelelőnek a kétdimenziós transzportegyenletet találtuk, amelyre analitikus megoldási módszert alkalmaztunk. A konkrét alkalmazást illetően feltételeztük, hogy az uralkodó szivárgás vízszintes irányú és iránya megállapítható. A két-, ill. háromdimenziós derékszögű koordinátarendszer x tengelyét ebben az irányban jelöljük ki, az y tengelyt pedig a vízszintes síkban, az erre merőleges irányban. A függőleges irányú szivárgást viszonylag rövid útvonala miatt elhanyagolhatjuk. A szivárgás fő trendjét meghatározó konvekció (fizikailag vagy kémiailag oldott anyagoknak a pórusokban való tömeges áramlása) a szennyezőanyag koncentrációjának változását eredményezi, ezért a transzportegyenletekben a t szerinti és az x szerinti elsőrendű parciális derivált közötti lineáris kapcsolattal írható le. Ha a transzportfolyamatnak csak a konvekciós összetevőjét vennénk

figyelembe, akkor tehát a jobb oldal nélküli, azaz zéró jobb oldallal felírt (2.1) vagy (2.2) egyenlet adódna. Ezt az alapvető áramlási trendet módosítja a diszperzió, amely egyrészt a különböző töménységű oldatok közötti kiegyenlítődésre irányuló részecskemozgáson, másrészt a szivárgási sebesség lokális eltérésein alapul, és másodrendű deriváltat (deriváltakat) tartalmazó tag(ok) belépését eredményezi a transzportegyenletbe. Attól függően, hogy csak az x irányú diszperziót vesszük figyelembe, vagy pedig az x és az y irányút, beszélünk egy-, ill. kétdimenziós transzportegyenletről. A másodrendű tagokban szorzóként szereplő α_L longitudinális és α_T transzverzális diszperzivitás olyan mechanikai paraméterek, melyek pontos értelmezésével itt nem kívánunk foglalkozni. Lényegében ugyanez vonatkozik az R késleltetési tényezőre is, amely egy 1-nél nagyobb, dimenzió nélküli faktor: a közeg sűrűségének, a hézagtérfogatnak és a megkötődési-visszaoldódási folyamat konstansainak a függvénye. A szennyezőanyag koncentrációjának mértékét csökkentheti rádioaktív, ill. kémiai és biológiai jellegű bomlás. (Ezt a λ-t tartalmazó tag fejezi ki a transzportegyenletekben.)

Az esetleges szennyezés konkrét lefolyását illetően feltételeztük, hogy a szennyezőanyag kiszivárgásának időtartama alatt naponta azonos M tömegű szennyeződés jut ki a környezetbe. Mivel a szennyezőanyag szivárgásának a sebessége nagyon kicsi, ezért nem eredményez jelentős számítási hibát, ha a kiszivárgó szennyeződést a valóságtól eltérően lökésszerűnek tekintjük, vagyis úgy számolunk, mintha naponta egy ízben M tömegű pillanatnyi szennyeződés jutna ki a környezetbe. Az emiatt fellépő számítási hiba elhanyagolható a paraméterek pontatlanságából eredő hibához képest.

3. A kétdimenziós transzportegyenlet egyszerűbb eseteinek analitikus megoldása

Tekintsük először azt a legegyszerűbb esetet, amikor M tömegű pillanatnyi szennyezés jön létre a $t = 0$ időpontban az $x = 0$, $y = 0$ helyen. Ebben az esetben a megoldás:

$$(3.1)\quad C(x,y,t) = \frac{M}{4\pi m n_0 v t \sqrt{\alpha_L \alpha_T}} \cdot \exp\left(- \frac{\left(x - \frac{vt}{R}\right)^2}{\frac{4\alpha_L v t}{R}} - \frac{y^2}{\frac{4\alpha_T v t}{R}} \right) \cdot \exp(-\lambda t),$$

ahol m a víztartó vastagsága, n_0 a szabad hézagtérfogat (a talaj hézagainak, üregeinek térfogataránya, dimenzió nélküli faktor), v a szivárgás sebessége a pórusokban, α_L a longitudinális, α_T a transzverzális diszperzivitás, R a késleltetési tényező, λ a bomlási együttható. (Lásd: [1, 154. és 159. oldal].)

Ha a feltételeken mindössze annyit módosítunk, hogy a pillanatnyi szennyezés a $t = \tau$ időpontban keletkezik, akkor a megoldás:

$$C(x,y,t) = \frac{M}{4\pi m n_0 v(t-\tau)\sqrt{\alpha_L \alpha_T}} \cdot \exp\left(- \frac{\left(x - \frac{v(t-\tau)}{R}\right)^2}{\frac{4\alpha_L v(t-\tau)}{R}} - \frac{y^2}{\frac{4\alpha_T v(t-\tau)}{R}} \right) \cdot \exp\big(-\lambda(t-\tau)\big). \tag{3.2}$$

Térjünk most vissza a 2. szakasz végén már vázolt feltételezett esethez, kicsit még konkrétabban. A tervezett hulladéklerakó belterületéről havaria folytán kikerülő szennyezőanyag terjedését vizsgálva feltételezzük, hogy a hiba keletkezésétől annak észleléséig tartó kritikus időszakban T napig naponta azonos M tömegű csurgalékvíz kerül a környezetbe. A szennyezőanyag-koncentráció térben és időben való eloszlását ekkor úgy kapjuk, hogy a (3.2) jobb oldalán levő kifejezést összegezzük $\tau = 0, 1, 2, \ldots, T-1$-re. Ily módon azt kapjuk, hogy a szennyezőanyag-koncentráció eloszlása a t időpontban a

$$\frac{C(x,y,t)}{M} = \frac{\exp(-\lambda t)}{4\pi m n_0 v \sqrt{\alpha_L \alpha_T}} \cdot \sum_{\tau=0}^{\min(t,T)-1} \frac{\exp(\lambda\tau)}{t-\tau} \exp\left(- \frac{\left(x - \frac{v(t-\tau)}{R}\right)^2}{\frac{4\alpha_L v(t-\tau)}{R}} - \frac{y^2}{\frac{4\alpha_T v(t-\tau)}{R}} \right) \tag{3.3}$$

formulával fejezhető ki. (A τ időpontban keletkező szennyezés természetesen nem befolyásolja az ezt megelőző időpontokra vonatkozó koncentráció-eloszlást, ezért csak a t-nél kisebb τ értékekre kell összegezni.)

4. A számítások gyakorlati megvalósítása

A $\frac{C(x,y,t)}{M}$ viszonyszám értékét egy 2000 méter hosszú és 200 méter széles sávban számítjuk ki különböző t időpontokra. A leginkább veszélyeztetett terület ugyanis a talajvíz-áramlás középvonalát tartalmazó sáv. A szóbanforgó sáv geometriai leírása:

$$0 \le x \le 2000, \qquad -100 \le y \le 100$$

A számítások eredményét, vagyis az egységnyi M értékre számított koncentráció megoszlását, annak időben való változását a szóbanforgó téglalap alakú sávban grafikus ábrázolásban mutatjuk az 1–12. ábrákon.

A konkrét esetre vonatkozó számításokat a $\frac{C(x,y,t)}{M}$ koncentráció viszonyszám különböző helyeken felvett értékeinek a (3.3) formula alapján történő meghatározására FoxPro kódban végeztük. A számításhoz szükséges adatoknak a következő értékeket tekintettük:

A víztartó vastagsága: $m = 20$ m. (Agyagréteg mélysége.)

Szabad hézagtérfogat: $n_0 = 0{,}15$. (Becsült érték.)

Szivárgás sebessége: $v = 1{,}5$ m/nap. (Becsült érték.)

Bomlási együttható: $\lambda = 0{,}0001$ (Becsült érték.)

Késleltetési tényező: $R = 1{,}2$. (Becsült érték.)

Longitudinális diszperzivitás: $\alpha_L = 25$ m. (Becsült érték.)

Transzverzális diszperzivitás: $\alpha_T = 0{,}5$ m. (Becsült érték.)

Szivárgás időtartama: $T = 90$ nap. (A monitoring rendszer legrosszabb esetben ennyi idő múlva észreveszi a hibát.)

A becsült érték alapján felvett adatok értékét a szabad hézagtérfogat, a szivárgási sebesség, a késleltetési tényező és a longitudinális diszperzivitás esetében az [1, 74. old.] táblázatában megadott „jellemző" értékekkel becsültük. A bomlási együttható esetén csekély mértékben eltértünk a „jellemző" értékként feltüntetett zérus értéktől, ez azonban érzésünk szerint csak nagyon jelentéktelen mértékben befolyásolhatja a számítás eredményét. A transzverzális diszperzivitásra nem találtunk jellemző értéket az említett táblázatban, volt azonban annyi támpontunk, hogy ennek értéke rendszerint egy vagy két nagyságrenddel kisebb a longitudinális diszperzivitás értékénél.

Az első menetben azt számítottuk ki, hogy a feltételezett hiba forrásától a talajvíz áramlási irányában 200, 400, 600, 800, ill. 1000 méter távolságba mikorra ér el a depóniáról származó szennyvíz a maximális koncentrációban, és annak mekkora az értéke a fenti transzportparaméterek esetén. A számítás eredményét az alábbi táblázat tartalmazza:

távolság	a koncentráció csúcsértékénél a szivárgás megszűnése után eltelt napok száma	legnagyobb koncentráció viszonyszám
200	90	0,0029
400	240	0,0014
600	400	0,00092
800	560	0,00067
1000	710	0,00053

A harmadik oszlopban a $C(x, y, t)/M$ értékek szerepelnek, tehát az oszlop értékei azt mutatják, hogy havaria esetén a depónia területéről egy nap alatt kiszivárgó csurgalékvíz tömegének mekkora része jelenik meg az adott helyen mérhető talajvíz egy köbméterében. Ha a számítást egyenként lebontanánk a csurgalékvízzel kikerülő káros anyagokra, akkor a táblázatunkban szereplő értékeket be kellene szorozni az egyes komponenseknek a csurgalékvízre vonatkozó koncentrációjával (1-nél jóval kisebb értékekkel). A károsanyag-komponensek pontos összetételét előre nem ismerhetjük, de anélkül is jól megbecsülhető a számítások alapján, hogy a depóniától 1000 méterre és ezt meghaladó távolságban az egyes komponensek csak milliomod és ezred közötti koncentrációban kerülhetnek a transzport-folyamat során a talajba.

A következő lépésben kiszámítottuk, hogy a korábban említett 2000 méter hosszú és 200 méter széles sávban különböző időpontokban milyen lenne a szennyezőanyag-koncentráció eloszlása egy havaria esetén. A kiválasztott időpontok: a meghibásodás kezdetétől számított minden 100-adik nap, az 1200-adik napig bezárólag (vagyis a hiba megszüntetése utáni 10-edik, 110-edik, 210-edik stb. nap). Ennek megfelelően a számítás eredményét 12 térhatású ábrával szemléltetjük (1–12. ábrák). A függőleges dimenzió a szennyezőanyag-koncentráció értékeit mutatja azzal a feltételezéssel, hogy naponta 1 köbméter mennyiségű csurgalékvíz szivárog ki 90 napon keresztül.

Megjegyzés. Az ábrák szemügyre vétele és tanulmányozása során vegyük figyelembe az ábrák bal szélén látható függőleges skála-beosztást, amely ábráról-ábrára változik. (A használt szoftver minden esetben automatikusan beállítja). Ezért az ábrákon megjelenő hullámfelületek „amplitudója" a valóságban az idő múlásával fokozatosan csökken (l. a lenti táblázatban).

Mivel a térhatású ábrákról pontosan nem olvashatók le a koncentrációk csúcsértékei, azért a következő táblázatban feltüntetjük az erre vonatkozó adatokat:

ábraszám	a szivárgás kezdete/megszüntetése után eltelt napok száma	csúcs-koncentráció értéke	helye
1.	100/ 10	0,0090	50 m
2.	200/ 110	0,0027	200 m
3.	300/ 210	0,0017	300 m
4.	400/ 310	0,0012	450 m
5.	500/ 410	0,00093	550 m
6.	600/ 510	0,00076	700 m
7.	700/ 610	0,00063	800 m
8.	800/ 710	0,00055	950 m
9.	900/ 810	0,00048	1050 m
10.	1000/ 910	0,00042	1200 m
11.	1100/1010	0,00038	1300 m
12.	1200/1110	0,00034	1450 m

Megállapítható, hogy a szennyezőanyag koncentráció csúcsértéke – amikor eléri az 1000 méteres távolságot – elég csekély mértékű, ezt követően pedig lassan, de biztosan tovább csökkenő tendenciát mutat.

A számított adatok eredményeként a következőket állapíthatjuk meg: Tudjuk, hogy a tervezett lerakó térségében a talajvízáramlás Ny–DNy irányú. A lerakóhoz legközelebb (kb. 800–1000 méterre eső) élővizek közül nagyjából ebben az irányban a 14. sz. csatorna húzódik. Abban a rendkívül csekély valószínűségű esetben, ha a szigetelőfólia meghibásodik, a kikerülő szennyezőanyag (a monitoring rendszer mintavételi gyakorisága alapján a legrosszabb esetet, 90 napig tartó szivárgást feltételezve) az 1 km-re levő kis csatornához kb. 800 nap alatt, mintegy 2000-szeres hígulásban jut el, tehát az élővízre nem jelent veszélyt. A lerakótól ÉNy-ra elhelyezkedő Gyáli patak, valamint a keletre és ÉK-re elhelyezkedő gyáli mélyfúrású ivóvíztermelő kutak biztonságát pedig gyakorlatilag semmiféle veszély nem fenyegeti a tervezett lerakó részéről.

Köszönetnyilvánítás: A szerzők köszönetüket fejezik ki az A.S.A. Környezetvédelem és Hulladékgazdálkodás Magyarország Kft.-nek egyrészt a tanulmányok készítésére vonatkozó megbízásért, másrészt azért, mivel a megbízás során olyan igénnyel lépett fel, amely a transzportmodellek alkalmazhatóságának a vizsgálatára, és mellékeredményként e cikk elkészítésére vezetett. Köszönetet mondunk Martin Attila projekt fejlesztési igazgatónak is a munka elkészítéséhez adott segítségért, hasznos információkért és értékes tanácsaiért.

IRODALOM

[1] Kovács Balázs, dr. Szabó Imre, Hulladékelhelyezés IV. *A szennyezőanyagok terjedése. A modellezés elmélete és gyakorlata*, „Ipar a környezetért" alapítvány, 1995.

Igénybe vett szoftverek:

A számítások gépi programjai a Microsoft FoxPro for Windows (2.6 verzió) fejlesztő környezetben készültek. A melléklet ábráinak készítése a Golden Software Inc. „Surfer Version 6.04" szoftverjével készült.

(Beérkezett: 1998. január 8.)

KÉRI GERZSON
MTA SZTAKI
BUDAPEST

ORSOVAI IMRE
GEOÖKOTERV
BUDAPEST

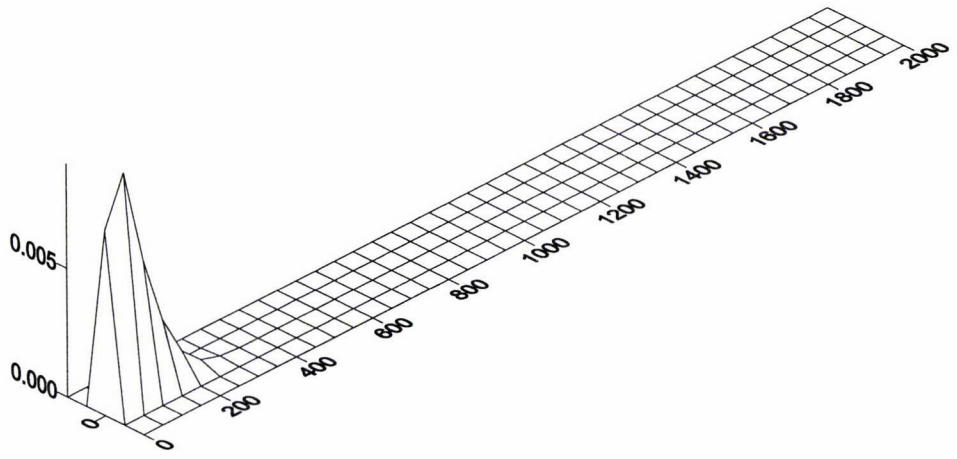
RAPCSÁK TAMÁS
MTA SZTAKI
BUDAPEST

APPLICATION OF A TRANSPORT MODEL TO EXAMINE THE POSSIBLE SOIL POLLUTION OF A WASTE-MATERIAL DEPOSITORY (A CASE-STUDY)

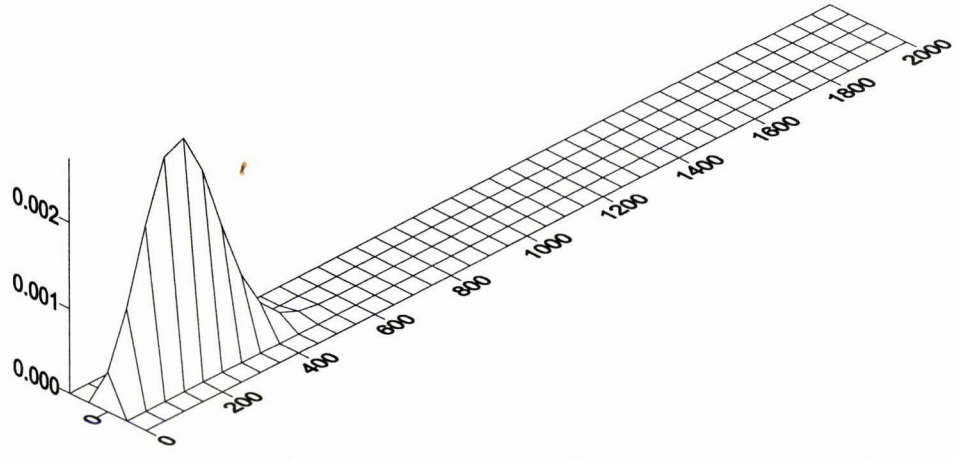
GERZSON KÉRI, IMRE ORSOVAI AND TAMÁS RAPCSÁK

In our days, when establishing new industrial units, before starting the foundation — besides efforts in the interest of the maximal safety — the elaboration of a plan (damage-plan) for managing the losses in consequence of break-downs is a fundamental requirement. When building a waste-material depository, one of the possible damage plans must concern the damage of the water-proofing and in its consequence, the environmental pollution caused by the waste material penetrating into the soil. Since for water-proofing highly safe technologies are used, the probability of the waste materials' getting into the soil with the ground water is extremely limited, but cannot be excluded entirely.

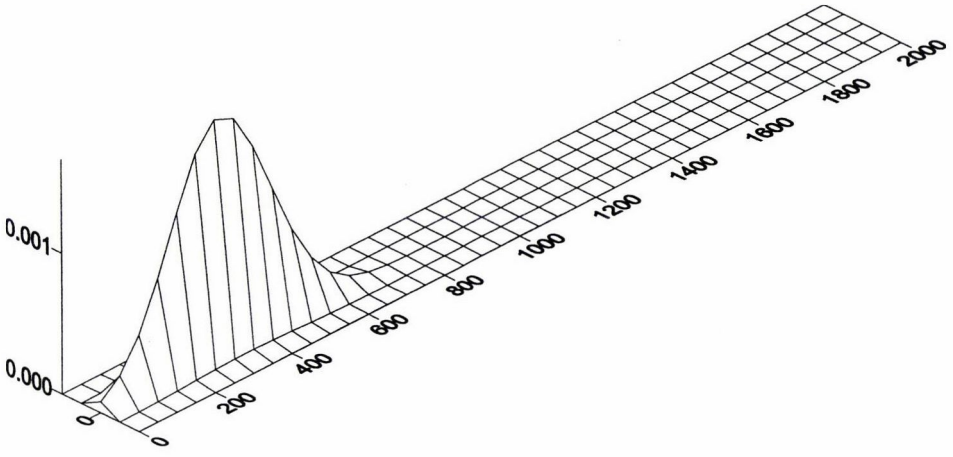
By using the models, known from the literature, that can be applied for tracing the spread of waste-material in the soil, in case the water-proofing of the waste-material depository should be damaged in a given time, the distance and the degree of the soil's pollution in different directions in consequence of the damage can be assessed. The case-study is about the calculations regarding a regional waste-material depository to be built in the vicinity of a village named Gyál. The change of the concentration of waste-materials in space and time is concentrated on, naturally, with not real but hypothetical data, since the depository will be built in the near future only. Hopefully, no waste-material will get into the soil either during the operation.


0.005
0.000
0
0
200
400
600
800
1000
1200
1400
1600
1800
2000

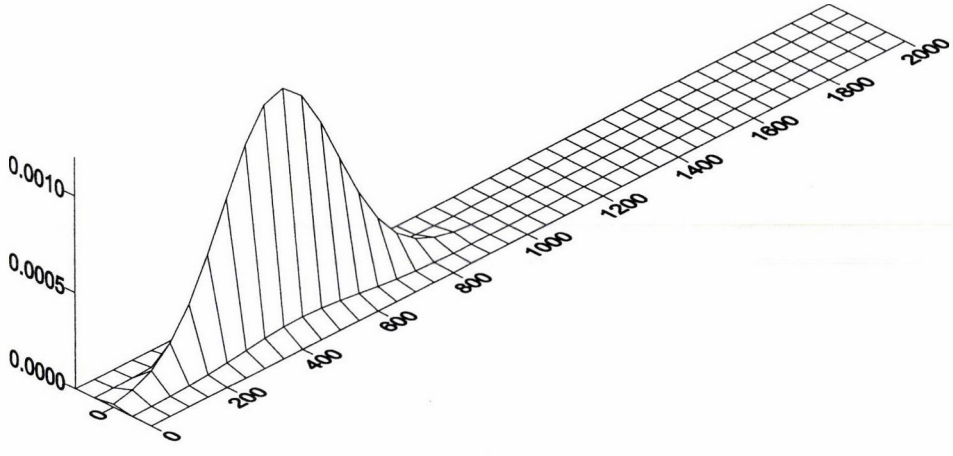
1. ábra


0.002
0.001
0.000
0
0
200
400
600
800
1000
1200
1400
1600
1800
2000

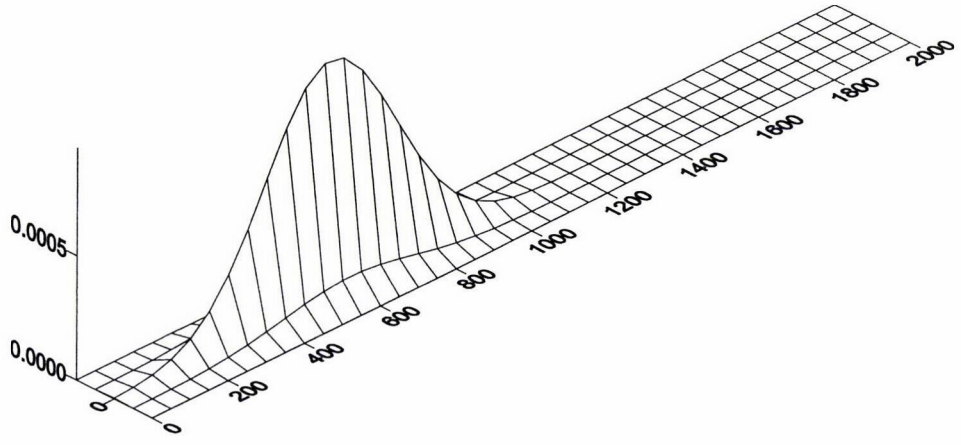
2. ábra


0.001
0.000
0
0
200
400
600
800
1000
1200
1400
1600
1800
2000

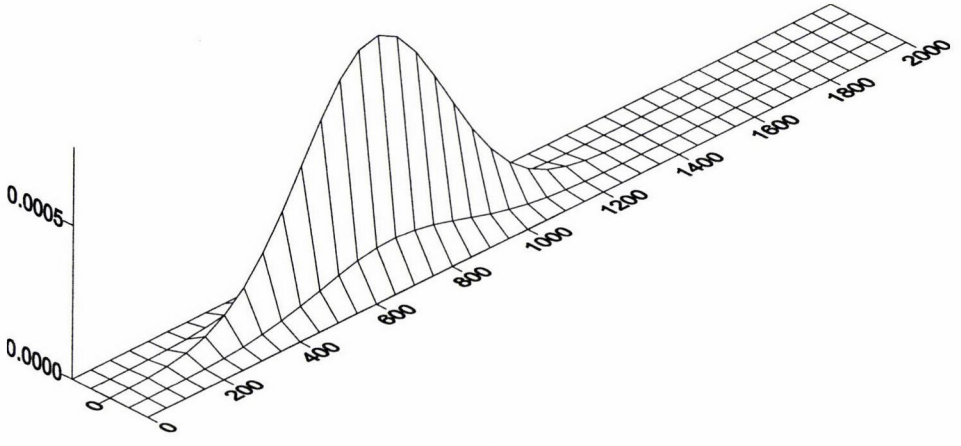
3. ábra


0.0010
0.0005
0.0000
0
0
200
400
600
800
1000
1200
1400
1600
1800
2000

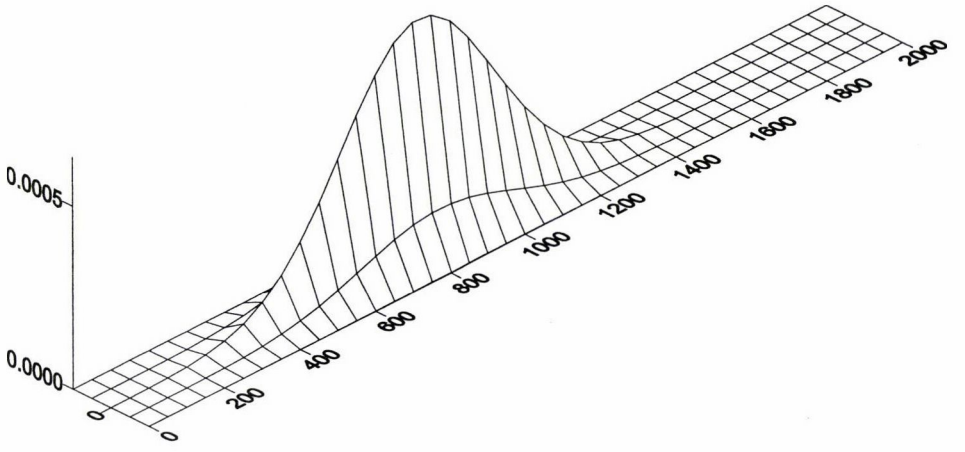
4. ábra


0.0005
0.0000
0
0
200
400
600
800
1000
1200
1400
1600
1800
2000

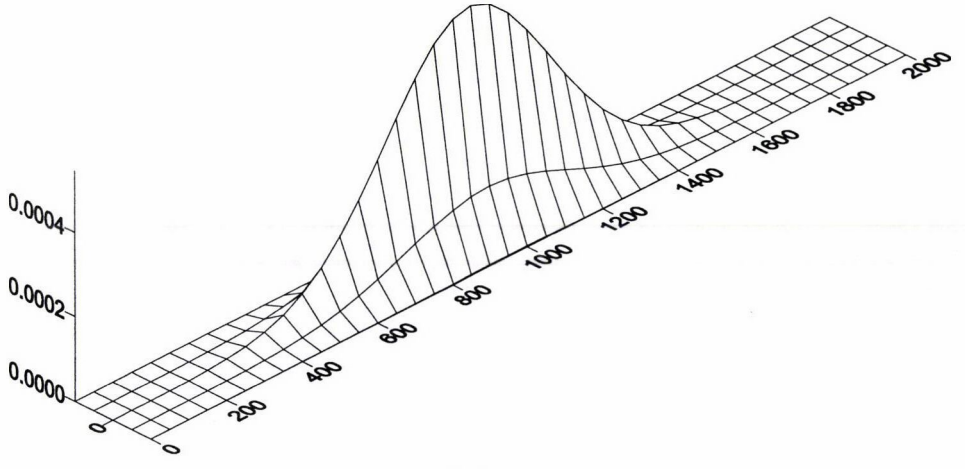
5. ábra


0.0005
0.0000
0
0
200
400
600
800
1000
1200
1400
1600
1800
2000

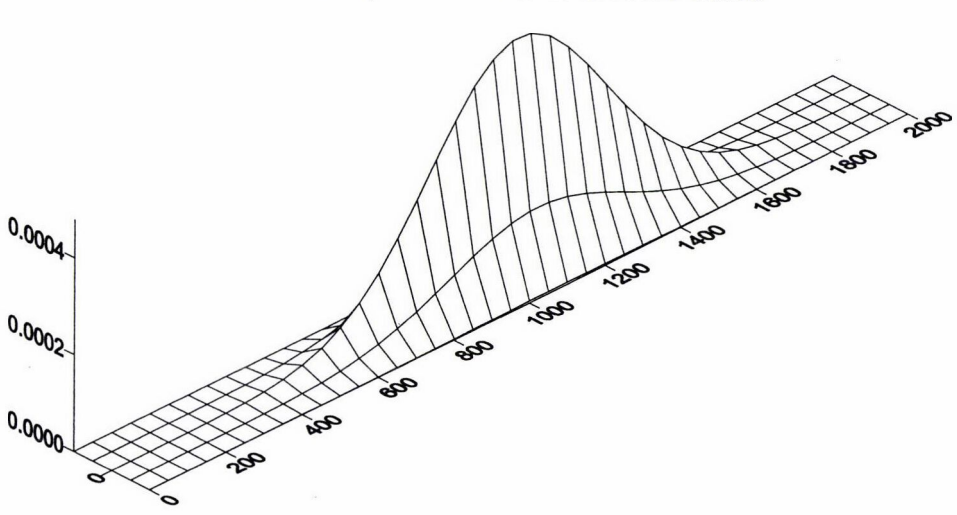
6. ábra


0.0005
0.0000
0
0
200
400
600
800
1000
1200
1400
1600
1800
2000

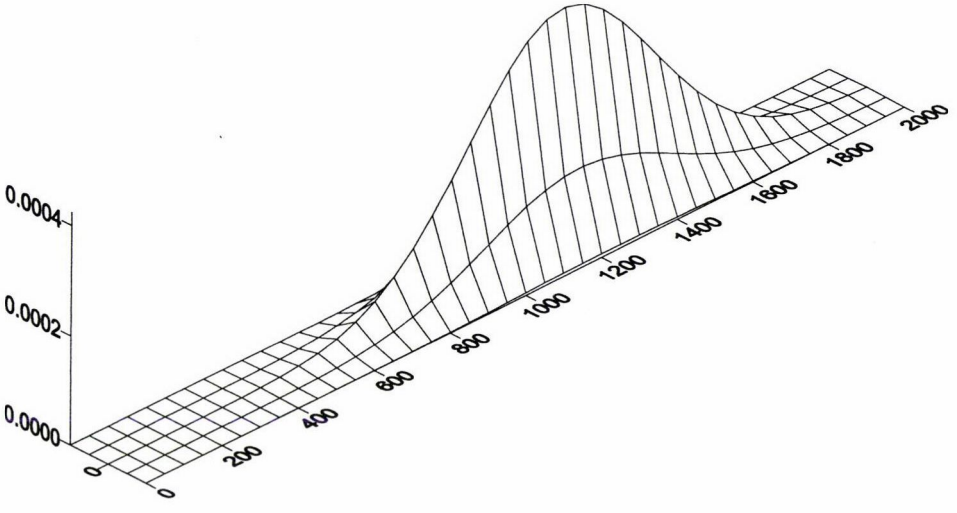
7. ábra


0.0004
0.0002
0.0000
0
0
200
400
600
800
1000
1200
1400
1600
1800
2000

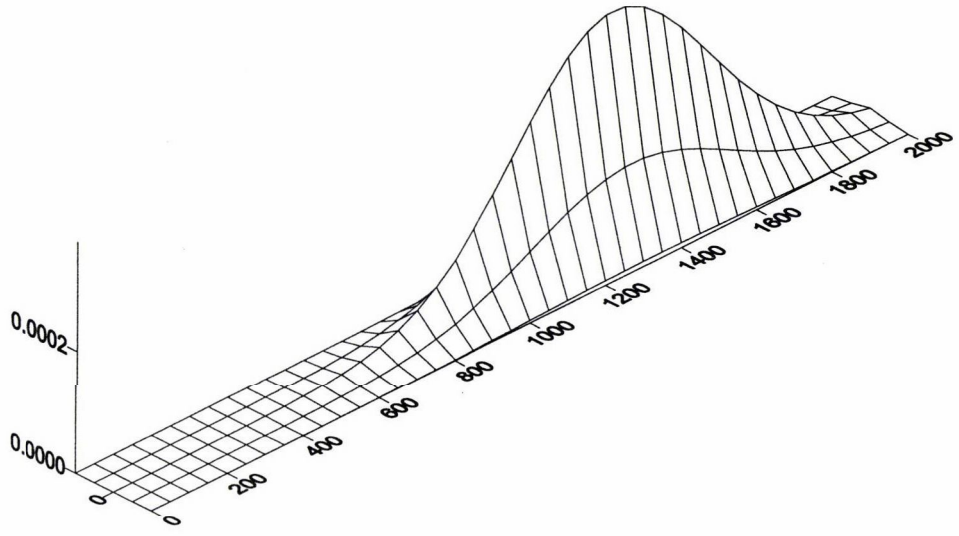
8. ábra



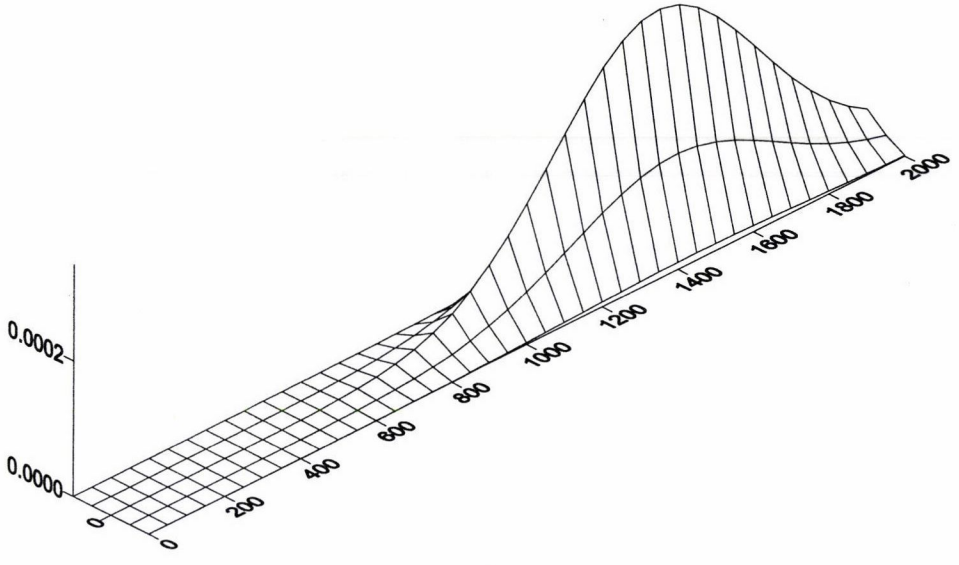
9. ábra



10. ábra



11. ábra



12. ábra

Alkalmazott Matematikai Lapok **19** (1999) 185–198

RUGALMAS-KÉPLÉKENY ANYAGÚ SÍKBELI KERETEK ELSŐ-, MÁSOD- ÉS HARMADRENDŰ ELMÉLETTEL TÖRTÉNŐ SZÁMÍTÁSA MATEMATIKAI PROGRAMOZÁSSAL

NÉDLI PÉTER*

Budapest

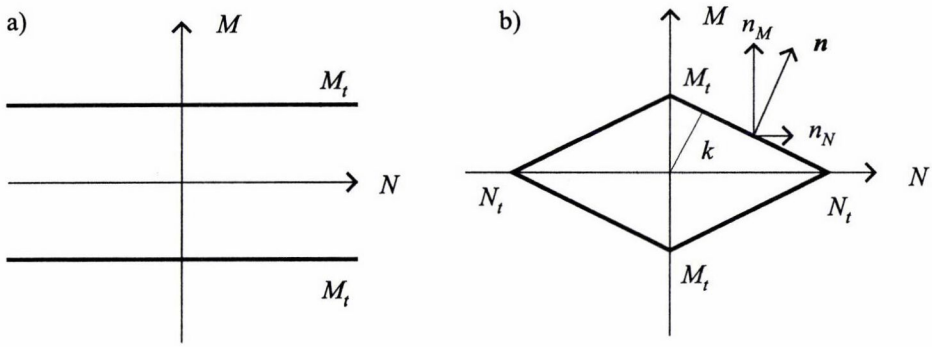
Rúdszerkezetek mechanikai állapotának jellemzésére véges szabadságfokú modellt használva a cikk bemutatja egyparaméteres terhelési folyamat vizsgálatát. Kiindul a geometriailag lineáris esetre (elsőrendű elmélet) érvényes megfogalmazásból és megadja, hogy a geometriai nemlinearitás esetén az összefüggések hogyan módosulnak. Az első esetben a feladat lineáris komplementer problémák sorozatára, a második esetben pedig teljesen nemlineáris programozási feladatok sorozatára vezet. A módszer alkalmazását a MINOS programcsomag használatával megoldott mintapélda illusztrálja.

1. Bevezetés

Ideálisan rugalmas anyagú, csak a csomópontjain terhelt keretszerkezetek állapotát diszkrét számítási modell esetén a következő változók írják le: $\mathbf{q}$ (csomóponti terhek), $\mathbf{t}$ (kinematikai terhek), $\mathbf{v}$ (csomóponti elmozdulások) és $\mathbf{s}$ (belső erők). Keretszerkezetek esetén e diszkrét modell könnyen kiterjeszthető a képlékeny tulajdonságok figyelembe vételére is további változók és két egyszerűsítő feltevés bevezetésével [1]. Az első feltevés, hogy folyás létrejöttét csak bizonyos ún. kritikus keresztmetszetekben engedjük meg; a második feltevés, amely a kezelhetőséghez szükséges pedig az, hogy ezen keresztmetszetekben a képlékenységi feltételt linearizáljuk. Ebben az esetben a képlékeny tulajdonságok két új változó bevezetésével leírhatók, melyek a következők: φ (plasztikus potenciál), λ (képlékeny szorzók). Csomópontjain terhelt keretek esetén a kritikus keresztmetszetek a rudak végkeresztmetszetei. A képlékenységi feltétel linearizálására tökéletesen képlékeny anyag feltételezésével két egyszerű esetet mutat be az 1. ábra.

Az a) eset a képlékeny csuklónak felel meg, a b) eset pedig a normálerő hatása figyelembe vételének egy lehetőségét tartalmazza. Az ábráknak megfelelő képlé-

* A szerző köszönetét fejezi ki az OTKA T069640 kutatás keretében biztosított támogatásért.



1. ábra

kenységi feltételek a következők:

$$\text{a)} \qquad \varphi = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} N \\ T \\ M \end{bmatrix} - \begin{bmatrix} M_t \\ M_t \end{bmatrix} \leq 0,$$

$$\text{b)} \qquad \varphi = \begin{bmatrix} n_N & 0 & n_M \\ -n_N & 0 & n_M \\ -n_N & 0 & -n_M \\ n_N & 0 & -n_M \end{bmatrix} \begin{bmatrix} N \\ T \\ M \end{bmatrix} - \begin{bmatrix} k \\ k \\ k \\ k \end{bmatrix} \leq 0,$$

$$n_N = \frac{M_t}{\sqrt{N_t^2 + M_t^2}}, \quad n_M = \frac{N_t}{\sqrt{N_t^2 + M_t^2}}, \quad k = \frac{N_t M_t}{\sqrt{N_t^2 + M_t^2}},$$

melyek tömör formában, az alábbi alakban írhatók: $\varphi = \mathbf{N}_K \mathbf{s}_K - \mathbf{k}_K \leq \mathbf{0}$. A K index a keresztmetszetre utal. Hasonló módon összeállítható egy tökéletesen képlékeny anyagú rúd ill. a szerkezet képlékenységi feltétele, mely az alábbi formában írható: $\varphi = \mathbf{N}\mathbf{s} - \mathbf{k} \leq \mathbf{0}$, ahol az $\mathbf{N}$ mátrix blokkdiagonál szerkezetű. A plasztikus potenciálhoz kapcsolódó folyási törvény szerkezetre vonatkozó alakja pedig a következő: $I = \{i \mid \varphi_i = 0\}$, $\dot{\mathbf{p}} = \mathbf{N}_I^T \dot{\boldsymbol{\lambda}}_I$, $\dot{\boldsymbol{\lambda}}_I \geq \mathbf{0}$, $\dot{\boldsymbol{\varphi}}_I^T \dot{\boldsymbol{\lambda}}_I = 0$, $\dot{\boldsymbol{\varphi}}_I \leq \mathbf{0}$. Itt I a φ vektor azon elemeinek az indexeiből alkotott halmazt jelöli, melyek értéke 0 (folyási helyek) és indexként használva a megfelelő részmátrixra ill. részvektorra utal, $\dot{\mathbf{p}}$ az általánosított képlékeny alakváltozássebességek vektora, a ^{T} felső index pedig a transzponálás jele. A jelen vizsgálat célja, hogy a fenti anyagtörvény esetére egyparaméteres teherrel terhelt szerkezet állapotváltozását meghatározza.

2. Elsőrendű elmélet szerinti vizsgálat

A folyási törvényből az következik, hogy az állapotváltozók közötti összefüggés a terhelési folyamat ismerete nélkül nem egyértelműen meghatározott, és csak

egy adott állapot megváltozására, azaz a változók sebességeire vonatkozóan írható fel egyértelmű összefüggés. Az alkalmazott anyagtörvény (lineárisan rugalmas — tökéletesen képlékeny, azaz nem viszkózus anyag) miatt a sebesség szó nem a változók idő szerinti, hanem a teherparaméter szerinti deriváltját jelenti. A köztük fennálló összefüggés, ha az egyensúlyi egyenleteket az eredeti geometria figyelembevételével írjuk fel és az általánosított alakváltozásokat a kis elmozdulások elmélete alapján számoljuk, (elsőrendű elmélet) az alábbi:

$$\begin{bmatrix} \mathbf{0} & \mathbf{G} & \mathbf{0} & \mathbf{0} \\ \mathbf{G}^T & \mathbf{F} & \mathbf{N}_I^T & \mathbf{0} \\ \mathbf{0} & \mathbf{N}_I & \mathbf{0} & -\mathbf{E}_I \end{bmatrix} \begin{bmatrix} \dot{\mathbf{v}} \\ \dot{\mathbf{s}} \\ \dot{\boldsymbol{\lambda}}_I \\ \dot{\boldsymbol{\varphi}}_I \end{bmatrix} + \begin{bmatrix} \dot{\mathbf{q}} \\ \dot{\mathbf{t}} \\ \mathbf{0} \end{bmatrix} = 0, \tag{2.1}$$

$$\dot{\boldsymbol{\varphi}}_I \leq \mathbf{0}, \qquad \dot{\boldsymbol{\lambda}}_I \geq \mathbf{0}, \qquad \dot{\boldsymbol{\varphi}}_I^T \dot{\boldsymbol{\lambda}}_I = 0.$$

Itt az első egyenlet-csoport a sebességekre vonatkozó egyensúlyi egyenletek, a második csoport a kompatibilitási egyenletrendszer, a harmadik pedig a plasztikus potenciál megváltozására vonatkozó egyenletek. $\mathbf{G}$ az egyensúlyi mátrix, $\mathbf{F}$ a hajlékonysági mátrix, $\mathbf{E}_I$ pedig egységmátrix. Matematikailag a rendszer egy lineáris komplementer problémát képez, mely a szimplex módszeren alapuló algoritmussal megoldható. Mindaddig, amíg újabb helyen folyás nem jön létre és a terheléssebesség $(\dot{\mathbf{q}}, \dot{\mathbf{t}})$ konstans, az összefüggést meghatározó mátrixok változatlanok és így a megoldás is konstans. Ez a tény lehetővé teszi, hogy a változók sebességeiről áttérjünk egy olyan megfogalmazásra, amely a változóknak a következő folyási helyhez tartozó értékeire vonatkozik. Mivel a geometriailag nemlineáris esetben is analóg gondolatmenetet alkalmazunk, ezt a megfogalmazást részletesen ismertetjük.

Tekintsünk egy ismert állapotot, mely kielégíti az egyensúlyi és a kompatibilitási egyenleteket, valamint a képlékenységi feltételt. Az állapotjellemzők ezen kiindulási állapothoz tartozó értékeit a ‚o' indexszel jelöljük. Egyparaméteres terhelést vizsgálva, keressük az állapotjellemzőknek a következő folyási helyhez tartozó értékeit. Ezek meghatározására az alábbi összefüggésrendszer írható fel, melyben az új jelölések közül az ‚a' index az alapteherre utal, ‚m' pedig a teherparaméter:

$$\text{a)} \quad \begin{bmatrix} \mathbf{0} & \mathbf{G} & \mathbf{0} & \mathbf{0} & \mathbf{q}_a \\ \mathbf{G}^T & \mathbf{F} & \mathbf{N}^T & \mathbf{0} & \mathbf{t}_a \\ \mathbf{0} & \mathbf{N} & \mathbf{0} & -\mathbf{E} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{v} \\ \mathbf{s} \\ \boldsymbol{\lambda} \\ \boldsymbol{\varphi} \\ m \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{k} \end{bmatrix}; \tag{2.2}$$

$$\text{b)} \quad I_o : \{i \,|\, \varphi_{oi} = 0\}, \quad \overline{I}_o : \{i \,|\, \varphi_{oi} < 0\};$$

$$\text{c)} \quad \boldsymbol{\lambda}_{I_o} \geq \boldsymbol{\lambda}_{o_{I_o}}, \quad \boldsymbol{\lambda}_{\overline{I}_o} = \boldsymbol{\lambda}_{o_{\overline{I}_o}}, \quad \boldsymbol{\varphi} \leq \mathbf{0}, \quad \boldsymbol{\varphi}^T(\boldsymbol{\lambda} - \boldsymbol{\lambda}_o) = 0;$$

$$\text{d)} \quad m = \max!$$

A (2.2) rendszer megoldható a szimplex módszeren alapuló algoritmussal. A megoldásból kapott új állapotból a vizsgálat folytatható. Ha két egymás utáni feladat ugyanazt a célfüggvény (teherparaméter) értéket adja eredményül úgy, hogy a folyási helyek sem változnak, akkor elértük a törőteherbírást.

Statikailag sokszorosan határozatlan szerkezetek esetén nagyon sok lépésre van szükség a törőteher eléréséig, ha minden folyási hely kialakulását követni kívánjuk. Ha azonban az egyes teherlépcsőkön belül a folyási helyeken a tehermentesülés lehetőségét nem vesszük figyelembe, akkor kevesebb teherlépcsővel is végigkövethető a terhelési folyamat. Az előző összefüggések a következőképp módosulnak. Az a) és d) változatlan, b) elmarad, c) pedig a következő lesz: $\lambda \geq \lambda_o$, $\varphi \leq \mathbf{0}$, $\varphi^T(\lambda - \lambda_o) = 0$, $m \leq m_o + \Delta m$. Itt Δm az előírt teherparaméter növekményt jelöli.

Végül, ha a teljes terhelési folyamat során eltekintünk a folyási helyeken a tehermentesüléstől, akkor kiindulási állapotnak a terheletlen állapotot tekinthetjük ($\lambda_o = \mathbf{0}$) és c) a következőképp alakul: $\lambda \geq \mathbf{0}$, $\varphi \leq \mathbf{0}$, $\varphi^T\lambda = 0$. Ez a harmadik, legegyszerűbb megfogalmazás tartalmazza a legtöbb közelítést, de előnye, hogy egy lépésben megadja a törőparamétert és az elmozdulások közelítő értékeit ebben az állapotban.

3. Harmadrendű elmélet szerinti vizsgálat

A valós szerkezeteken végzett vizsgálatok azt mutatják, hogy az elsőrendű elmélet szerinti számított törőteherbírás általában nem realizálható a geometriai nemlinearitás következtében még akkor sem, ha a rugalmas kritikus teher akár egy nagyságrenddel magasabb is mint a képlékeny teherbírás. A „nagy elmozdulás, de kis alakváltozás" feltevést elfogadva, ha a geometriai nemlinearitást rúdszerkezetek (vonalkontinuumok) esetén pontosan vesszük figyelembe, akkor harmadrendű elméletről beszélünk. A feladat diszkrét jellege ebben az esetben is megtartható és a szerkezet állapotát az elsőrendű elméletnél használt vektorokkal azonos méretű vektorokkal jellemezhetjük. Az elsőrendű elmélettel ellentétben viszont egy rúd lokális koordinátarendszere nem állandó, hanem a mindenkori konfiguráció függvénye. A jelen munkában a [2]-ben használt megközelítést alkalmaztuk, melynek leglényegesebb jellemzőit röviden összefoglaljuk. A lokális koordinátarendszer a rúd kisebbik sorszámú (kezdő) csomópontjával mozog együtt és kiindulási állapotban mindegyik lokális koordinátarendszer a globális koordinátarendszerrel egyállású. Az általánosított rúderők és az általánosított alakváltozások ebben a lokális koordinátarendszerben vannak értelmezve. Az általánosított rúderő a rúd kezdőkeresztmetszetének belső erőit (normálerő, nyíróerő, hajlítónyomaték), az általánosított alakváltozás pedig a rúd nagyobbik sorszámú csomópontja és a rúd végkeresztmetszete közti relatív elmozdulást jelenti. E két jellemző azért van így definiálva, hogy a rugalmas rúdszakasz alakjának meghatározását egy nemlineáris elsőrendű diffe-

renciálegyenlet rendszer és a kezdeti feltételek (ún. kezdetiérték feladat) megoldása szolgáltassa. Ez a kezdetiérték feladat általában csak numerikusan oldható meg.

Egy ismert állapotból a következő folyási helyhez tartozó állapot meghatározására vonatkozó összefüggésrendszer abban különbözik (2.2)-től, hogy a 2.2/a lineáris egyenletrendszer helyett egy nemlineáris egyenletrendszer szerepel, és azért, hogy a határpont utáni állapotokat is vizsgálni lehessen, célfüggvényként az alaptehernek az elmozdulásokon végzett munkáját vesszük:

$$\text{(3.1)} \qquad \text{a)} \quad \mathbf{e}(\mathbf{v},\mathbf{s},m)=\mathbf{0} \quad \mathbf{c}(\mathbf{v},\mathbf{s},\boldsymbol{\lambda},m)=\mathbf{0} \quad \mathbf{f}(\mathbf{v},\mathbf{s},\boldsymbol{\lambda},\boldsymbol{\varphi})=\mathbf{k}$$

$$\text{b)} \quad I_o : \{i \,|\, \varphi_{oi}=0\}, \quad \overline{I}_o : \{i \,|\, \varphi_{oi}<0\};$$

$$\text{c)} \quad \boldsymbol{\lambda}_{I_o} \geq \boldsymbol{\lambda}_{o_{I_o}}, \quad \boldsymbol{\lambda}_{\overline{I}_o} = \boldsymbol{\lambda}_{o_{\overline{I}_o}}, \quad \boldsymbol{\varphi} \leq 0, \quad \boldsymbol{\varphi}^T(\boldsymbol{\lambda}-\boldsymbol{\lambda}_o)=0;$$

$$\text{d)} \quad \mathbf{q}_a^T\mathbf{v} = \max!$$

Itt $\mathbf{e}$, $\mathbf{c}$, $\mathbf{f}$ az egyenletek bal oldalait képező nemlineáris vektor-vektor függvényeket jelöli. Az egyensúlyi egyenletek és a képlékenységi feltétel képlet formájában felírhatók és kiértékelhetők az ismeretlen változók ($\mathbf{v}$, $\mathbf{s}$, $\boldsymbol{\lambda}$, $\boldsymbol{\varphi}$, m) egy adott értékére, de a kompatibilitási egyenlet nem írható fel zárt képlet alakjában, és így csak az egyes rudakra vonatkozó kezdetiérték-feladat numerikus megoldásával értékelhető ki. Maga az összefüggésrendszer egy nemlineáris programozási problémát alkot annak minden bonyolultságával, azaz nem zárható ki az, hogy több megoldás létezik, ill. az sem, hogy nincs megoldás. Így a megoldás egyértelműségének eldöntése további vizsgálatot igényel. Ennek részleteire ebben a cikkben nem térünk ki.

4. Másodrendű elmélet szerinti vizsgálat

A harmadrendű elmélet szerinti vizsgálat igen számításigényes, mivel a kompatibilitási egyenlet kiértékelése minden rúdon egy kezdetiérték feladat numerikus megoldását igényli és a megoldás során sok kiértékelésre van szükség. Azokban az esetekben, amikor egy rúdon belül csak kis elmozdulások jönnek létre jó közelítést ad az ú.n. másodrendű elméleten alapuló közelítés is. Ez a feltétel biztosítható, ha a rudat megfelelő számú részre osztjuk. A „másodrendű elmélet" kifejezés gyűjtőfogalom, mert természetesen több, különböző közelítés létezik. A jelen munkában a [3]-ban ismertetett közelítést alkalmaztuk, mely normálerővel is terhelt gerenda differenciálegyenletének megoldásán alapszik. A megoldásból az ún. stabilitásfüggvények bevezetésével előállítható egy rúd másodrendű elmélet szerinti merevségi ill. hajlékonysági mátrixa. Így az egyensúlyi, kompatibilitási egyenlet valamint a képlékenységi feltétel mátrixegyenlet formájában felírható, de a benne szereplő mátrixok is függnek az ismeretlenektől. Ebben az esetben a következő folyási hely megkere-

sésére vonatkozó matematikai programozási feladat a következő:

$$(4.1) \qquad \text{a)} \quad \begin{bmatrix} \mathbf{0} & \mathbf{G}(\mathbf{v}) & \mathbf{0} & \mathbf{0} & \mathbf{q}_a \\ \mathbf{G}^T(\mathbf{v}) & \mathbf{F}(\mathbf{s}) & \mathbf{N}^T(\mathbf{v}) & \mathbf{0} & \mathbf{t}_a \\ \mathbf{0} & \mathbf{N}(\mathbf{v}) & \mathbf{0} & -\mathbf{E} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{v} \\ \mathbf{s} \\ \boldsymbol{\lambda} \\ \boldsymbol{\varphi} \\ m \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{k} \end{bmatrix};$$

$$\text{b)} \quad I_o : \{i \,|\, \varphi_{oi} = 0\}, \quad \overline{I}_o : \{i \,|\, \varphi_{oi} < 0\};$$

$$\text{c)} \quad \boldsymbol{\lambda}_{I_o} \geq \boldsymbol{\lambda}_{o_{I_o}}, \quad \boldsymbol{\lambda}_{\overline{I}_o} = \boldsymbol{\lambda}_{o_{\overline{I}_o}}, \quad \boldsymbol{\varphi} \leq \mathbf{0}, \quad \boldsymbol{\varphi}^T(\boldsymbol{\lambda} - \boldsymbol{\lambda}_o) = 0;$$

$$\text{d)} \quad \mathbf{q}_a^T \mathbf{v} = \max!$$

5. Numerikus megoldás és mintapélda

Sokszor fordul elő, hogy egy mérnöki probléma olyan matematikai feladatra vezet, amely megoldására még nem áll rendelkezésre precízen igazolt matematikai algoritmus. Egy példa erre a véges elemek módszere, ahol a mérnöki szemlélet alapján kimunkált megoldási módszer megelőzte a matematikai megoldást. Vannak azonban a matematikának olyan megoldási módszerei is, melyek a mérnöki gyakorlatban még csak kis mértékben terjedtek el. Ilyen például a matematikai programozás területe, mely elsősorban gazdasági problémák megoldásának igényéből fejlődött ki, és amely jelenleg már ott tart, hogy kereskedelmi forgalomban kaphatók több ezer ismeretlen és feltétel kezelésére alkalmas programcsomagok. A jelen vizsgálatban a MINOS (Modular In Core Optimization System) programcsomagot [4] alkalmaztuk a 2.–4. pontokban megfogalmazott feladatok vizsgálatára, amely a Wolfe féle redukált gradiens módszert használja a nemlineáris programozási probléma megoldására. A programcsomag nemlineáris feladat esetén a felhasználótól két szubrutin megírását kívánja meg. Az egyik a feltételek bal oldalát jelentő vektor-vektor függvény és annak Jacobi mátrixa megadását, a másik a célfüggvény és gradiense megadását kell, hogy tartalmazza. A Jacobi mátrix és a gradiens megadása nem okvetlenül szükséges. Ha elmarad, akkor a rendszer numerikus deriválással számolja. A jelen esetben kihasználtuk ezt a lehetőséget.

Illusztrációként egy, az irodalomból ismert feladat, megoldását mutatjuk be [5]. Az adatokat a 2. ábra tartalmazza. A szerkezet 14 csomópontot és 16 rudat tartalmaz. Az ismeretlen $[\mathbf{v}, \mathbf{s}, \boldsymbol{\lambda}, \boldsymbol{\varphi}, m]^T$ vektor mérete: $14 \cdot 3 + 16 \cdot (3 + 4 + 4) + 1 = 219$. A feltételek száma: $14 \cdot 3$ (egyensúly) $+16 \cdot 3$ (kompatibilitás) $+16 \cdot 4$ (képlékenységi feltétel) $+1$ (normalitás) $= 155$. A 3. ábra mutatja be a három különböző elmélettel kapott erő-elmozdulás diagrammok összehasonlítását, a 4. ábra pedig a terhelési folyamat befejeződéséhez tartozó elmozdulási és nyomatéki ábrákat tünteti fel. Elsőrendű elméletnél a terhelési folyamat a folyási határállapot eléréséig

tart. Másod- és harmadrendű elmélet esetén viszont a szerkezet ú.n. posztkritikus állapota (az erő-elmozdulás diagramm leszálló ága) is vizsgálható. Az ezen az ágon történő továbbhaladásnak korlátot szab a képlékeny csuklók alakváltozási képessége. A jelen feladatban ezt azonban nem vizsgáltuk, hanem a vizsgálatot egy bizonyos lépésszám után abbahagytuk. Ez azt jelenti, hogy a másod- és harmadrendű esetben a terhelési folyamat befejeződéséhez tartozó állapotnak nincs kitüntetett jelentése.

IRODALOM

[1] O. De Donato and G. Maier, Historical deformation analysis of elastoplastic structures as a parametric linear complementarity problem, *Meccanica*, **3** (1976).

[2] Gáspár Zs., „Rugalmas rúdszerkezetek nagy elmozdulásai", Kadidátusi értekezés, 1976.

[3] K. I. Majid, *Non-Linear Structures*, Butterworths (London, 1972).

[4] MINOS User's Guide. University of California, System Optimization Laboratory, 1987.

[5] J. A. T. De Freitas and D. Lloyd Smith, Plastic straining, unstressing and branching in large displacement perturbation analysis, *International Journal for Numerical Methods in Engineering*, **20** 1984.

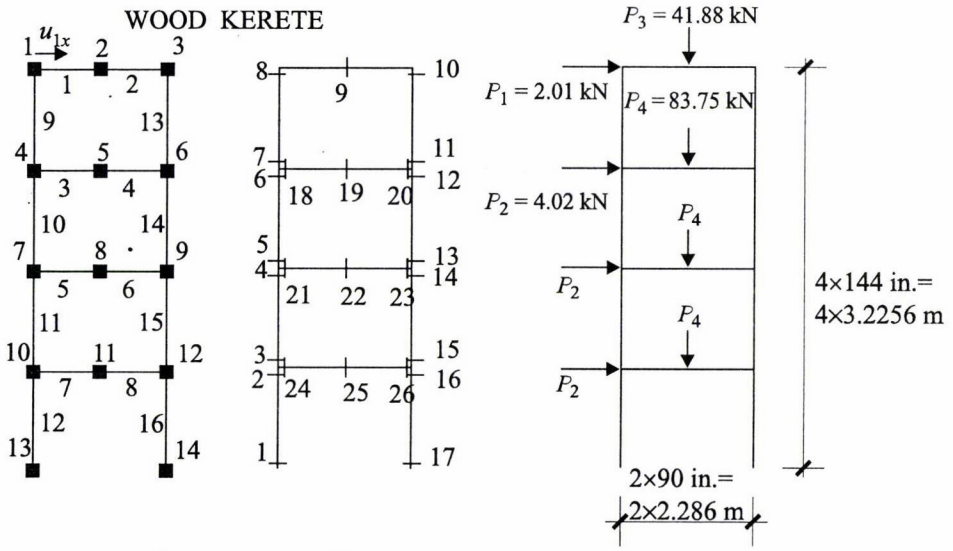
(Beérkezett: 1999. február 9.)

NÉDLI PÉTER
MTA-BME TARTÓSZERKEZETEK NUMERIKUS MECHANIKÁJA KUTATÓCSOPORT
1521 BUDAPEST
MŰEGYETEM RKP. 3.
KMF. 35.
E-mail: NEDLI@ep-mech.me.bme.hu

COMPUTATION OF ELASTIC-PLASTIC PLANE FRAMES BY MATHEMATICAL PROGRAMMING IN CASE OF FIRST, SECOND AND THIRD ORDER THEORY

PÉTER NÉDLI

Paper describes the analysis of the one parameter loading history of frames using a discrete model to characterize the mechanical behaviour of the structure. As a starting point, the geometrically linear behaviour is formulated (first order theory) and then the necessary changes in the formulation are introduced to treate the geometrically nonlinear case. The first case leads to a series of linear complementarity problems and the second case to a series of fully nonlinear programming problems. The application of the method is illustrated by an example solved using the MINOS mathematical programming package.

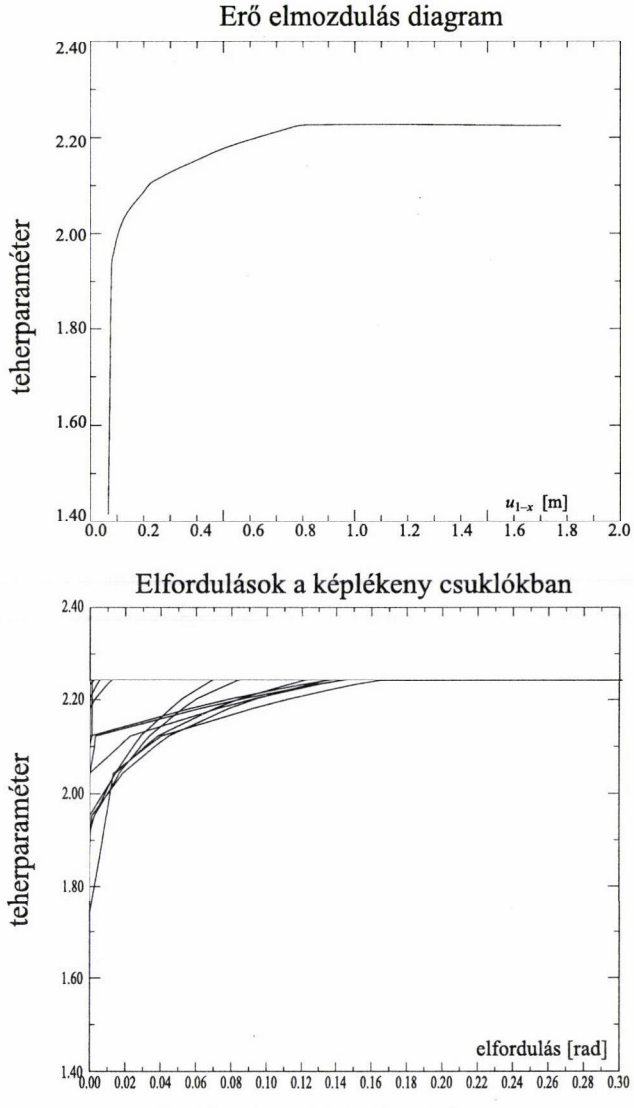


$E = 13\,548$ t/in^2 $= 21\,000$ kN/cm^2

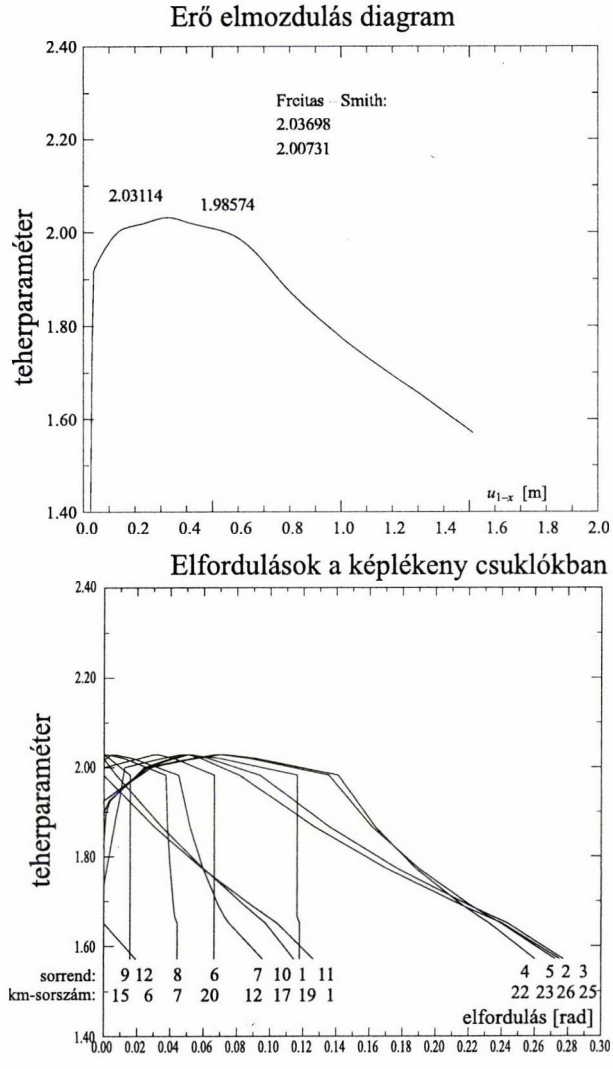
A RUDAK ATADAI:

Rúd	A		I		M_p	
	in^2	cm^2	in^4	cm^4	t.in.	kNm
1, 2	5.30	34.19	55.63	2315	244.0	61.98
3–8	7.35	47.42	122.34	5092	428.0	108.71
9, 13	5.89	38.00	34.71	1445	205.3	52.15
10, 14	7.37	47.55	43.69	1819	259.3	65.86
11, 15	8.28	53.42	86.69	3608	393.5	99.95
12, 16	10.32	65.58	115.06	4789	502.3	127.58

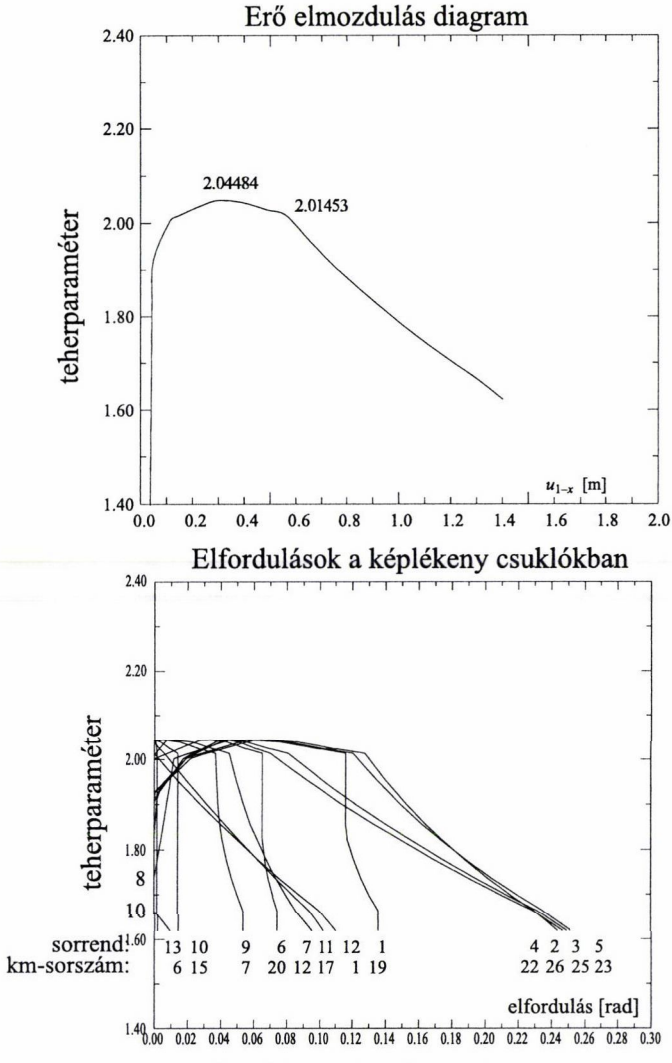
2. ábra



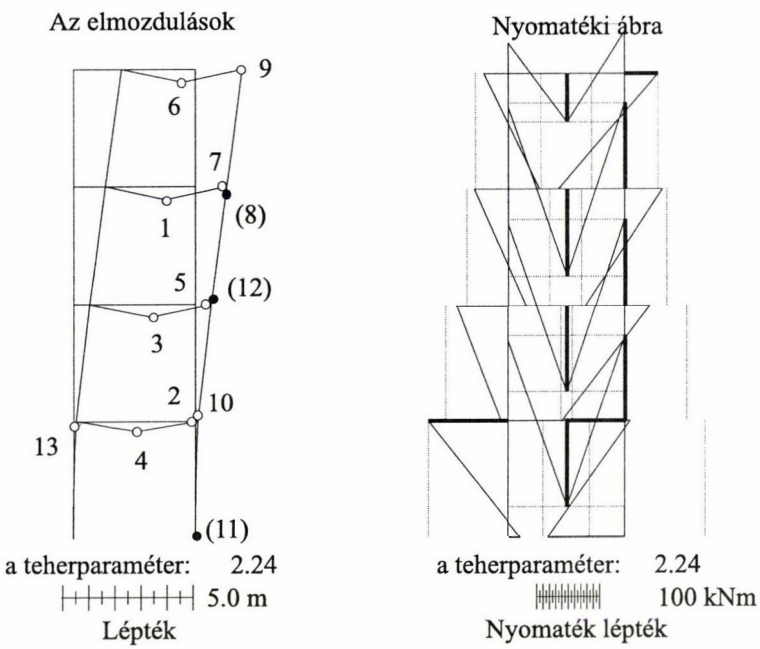
3a. ábra. I. rendű elmélet szerint



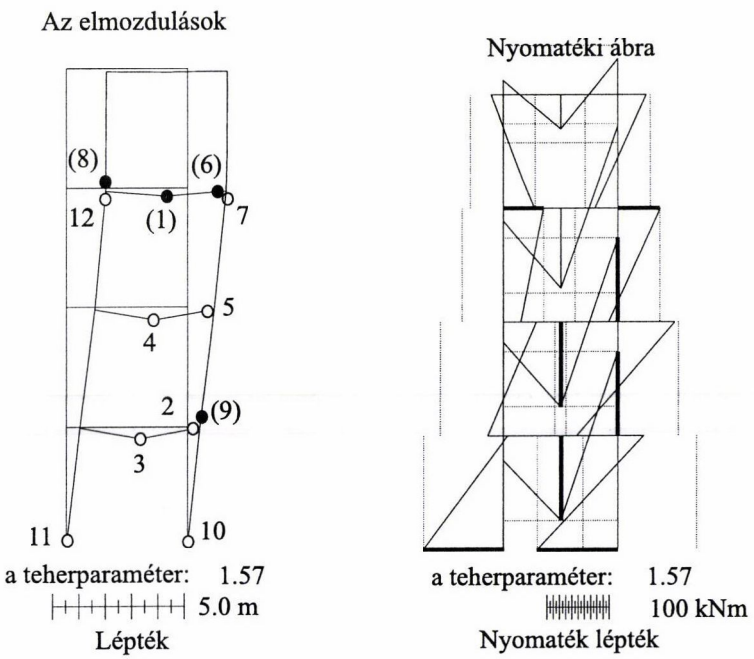
3b. ábra. II. rendű elmélet szerint



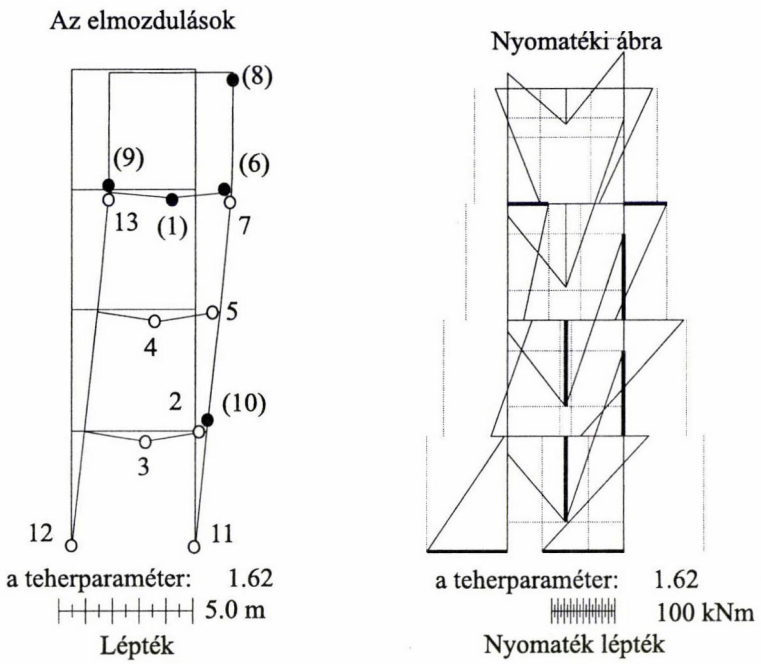
3c. ábra. III. rendű elmélet szerint



4a. ábra. I. rendű



4b. ábra. II. rendű



4c. ábra. III. rendű

Alkalmazott Matematikai Lapok **19** (1999) 199–216

AZ EXPONENCIÁLIS BARRIER PROGRAMOZÁS, MINT A LINEÁRIS PROGRAMOZÁS ANALITIKUS KÖZELÍTÉSE

KLAFSZKY EMIL ÉS MÁLYUSZ LEVENTE

Budapest

Dolgozatunkban a standard Lineáris Programozási (LP) feladatpárok megoldásával foglalkozunk. A dolgozat célja egy speciális konvex programozási feladat bemutatása, amely alkalmas az LP tetszőleges pontosságú közelítésére. Az LP egyensúlyi feltételét perturbáljuk és a kotangens hiperbolikusz függvényt használjuk egyensúlyi feltételként. Így olyan α, β paraméteres konvex programozási feladatpárt definiálunk, nevezzük Exponenciális barrier Programozási feladatnak, azaz EP(α,β) feladatnak, amelyre igaz, hogy ha $\alpha\beta \to 0$ akkor EP$(\alpha,\beta) \to$ LP. A főbb elméleti eredményekre, például az erős dualitási tételre egyszerű, elemi bizonyítást adunk.

1. Bevezetés

Nem újkeletű az a próbálkozás, hogy az LP feladatpárt konvex programozási feladat segítségével oldjuk meg. Már 1968-ban Fiacco–McCormick [5] definiált egy konvex programozási feladatot, amelyről megmutatták, hogy az alkalmas az LP tetszőleges pontosságú közelítésére. A téma kapcsolódik a belsőpontos algoritmusokhoz (lásd [6], [15]), ezért azóta sem vesztett aktualitásából, amint azt többek között Roos–Terlaky–Vial 1997-ben megjelent könyve [12] is mutatja. Cikkünkben az általunk definiált konvex programozási feladatpárra vonatkozóan elsősorban néhány e két könyvben megjelent tételekkel analóg tételeket bizonyítunk. Az első fejezetben röviden összefoglaljuk az LP feladatpárral kapcsolatos alapvető eredményeket. A következő részben definiáljuk az EP(α,β) feladatot. A harmadik fejezetben megmutatjuk, hogy az EP(α,β) az LP analitikus közelítése, mert az EP(α,β) optimális $\mathbf{x}_{(\alpha,\beta)}, \mathbf{z}_{(\alpha,\beta)}$ megoldására fennáll, hogy $\lim_{\alpha\beta\to 0} \mathbf{x}_{(\alpha,\beta)} = \mathbf{x}^*$, $\lim_{\alpha\beta\to 0} \mathbf{z}_{(\alpha,\beta)} = \mathbf{z}^*$, ahol $\mathbf{x}^*, \mathbf{z}^*$ LP optimálisok. Megmutatjuk továbbá, hogy $\mathbf{x}_{(\alpha,1)}, \mathbf{z}_{(\alpha,1)}$ α illetve $\mathbf{x}_{(1,\beta)}, \mathbf{z}_{(1,\beta)}$ β függvényében differenciálható görbe. Végül egy-egy konvex programozási feladattal definiáljuk az $\mathbf{x}^*$, $\mathbf{z}^*$ LP optimális megoldásokat.

Több alkalmazásban, például a mechanikában, kémiában és a közgazdasági feladatokban, az egyensúlyi feladatokat egyszerűbb és természetesebb interpretálni. Elég, ha csak arra, gondolunk, hogy mechanikában használt úgynevezett potenci-

álfüggvények valójában fiktív függvények. Az általunk vizsgált matematikai programozási feladat ekvivalens egy egyensúlyi feladattal (egyenlőség vagy egyenlőtlenség rendszerrel) abban az értelemben, hogy ha az egyik megoldható, akkor és csak akkor a másik is, és a programozási feladat optimális megoldása egybeesik az egyensúlyi feladat megoldásával.[1] Ezért először az egyensúlyi feladatokat fogalmazzuk meg, majd az egyensúlyi feladattal ekvivalens matematikai programozási feladatot. A következőkben röviden felelevenítjük a lineáris programozás feladatát, alaplemmáját és dualitási tételét.

Az alábbi jelöléseket kívánjuk használni: $A \in R^{m\times n}$ mátrix, $\mathbf{x},\mathbf{z} \in R^n$ ismeretlen vektorok, $\widehat{\mathbf{z}}, \widehat{\mathbf{x}} \in R^n$ adott vektorok, $\mathbf{y} \in R^m$ ismeretlen vektor. Az általánosság elvének megsértése nélkül tegyük fel, hogy az A mátrix teljes sorrangú. Tetszőleges $\mathbf{a} \in R^n$ és $\mathbf{b} \in R^n$ vektorok skaláris szorzatát $\mathbf{ab}$-vel jelöljük.

1.1. FELADAT. *Az* LP *egyensúlyi feladata*(LPE).

$$A\mathbf{x} = A\widehat{\mathbf{x}} \qquad \mathbf{z} = \widehat{\mathbf{z}} + A^T\mathbf{y} \tag{1.1}$$

$$\mathbf{x} \geq \mathbf{0} \qquad \mathbf{z} \geq \mathbf{0} \tag{1.2}$$

$$\mathbf{zx} = 0 \tag{1.3}$$

Itt (1.1) a lineáris affin feltétel, (1.2) a nemnegativitási feltétel és (1.3) a komplementaritási feltétel. Az utóbbi kettőt együtt egyensúlyi feltételnek nevezzük.

A LPE-t az alábbi ekvivalens formában — LP feladatpárként — is megfogalmazzuk.

1.2. FELADAT. *Az* LP *feladatpár* (LP).

	Primál	*Duál*
(1.4)	$A\mathbf{x} = A\widehat{\mathbf{x}}$	$\mathbf{z} = \widehat{\mathbf{z}} + A^T\mathbf{y}$
(1.5)	$\mathbf{x} \geq 0$	$\mathbf{z} \geq 0$
(1.6)	$\min \rightarrow \widehat{\mathbf{z}}\mathbf{x}$	$\min \rightarrow \mathbf{z}\widehat{\mathbf{x}}$

Az (1.6) feltétel a primál illetve a duál célfüggvényt írja le. Vezessük be a primál és a duál megengedett megoldások halmazára az alábbi jelöléseket:

$$P = \{\mathbf{x} \in \mathbf{R}^n_{\oplus} \mid A\mathbf{x} = A\widehat{\mathbf{x}}\}$$

$$D = \{\mathbf{z} \in \mathbf{R}^n_{\oplus}, \mathbf{y} \in \mathbf{R}^m \mid \mathbf{z} = \widehat{\mathbf{z}} + A^T\mathbf{y}\}$$

Elemi számolással igazolható, hogy az (1.4)-et kielégítő $\mathbf{x}$ és $\mathbf{z}$ vektorok között fennáll az alábbi egyenlőség.

$$\widehat{\mathbf{z}}\mathbf{x} + \mathbf{z}\widehat{\mathbf{x}} = \mathbf{zx} + \widehat{\mathbf{z}}\widehat{\mathbf{x}} \tag{1.7}$$

[1] A továbbiakban egy egyensúlyi feladat és egy programozási feladatpár ekvivalenciáján mindig az előző mondatban leírtakat értjük, ha ezt külön nem is hangsúlyozzuk.

A fenti két feladat, az LP és az LPE ekvivalens egymással olyan értelemben, hogy az LP akkor és csak akkor oldható meg ha az LPE is megoldható és az LP optimális megoldásainak halmaza egybeesik az LPE megoldásával. A továbbiakban összegezzük az ekvivalencia bizonyításra szolgáló lemmát, dualitási tételt és következményeiket ([4], [12] és [17]).

1.1. LEMMA (alaplemma). *Minden* $\mathbf{x} \in P$*-re és* $\mathbf{z} \in D$*-re igaz, hogy*

$$\widehat{\mathbf{z}}\mathbf{x} + \mathbf{z}\widehat{\mathbf{x}} \geq \widehat{\mathbf{z}}\widehat{\mathbf{x}} \tag{1.8}$$

és egyenlőség akkor és csak akkor áll fenn, ha $z_j x_j = 0$, $\forall j-re$*, vagyis ha teljesül a komplementaritási feltétel.*

1.2. KÖVETKEZMÉNY (gyenge egyensúly). *Ha* $\mathbf{x} \in P$ *és* $\mathbf{z} \in D$*, valamint* (1.8)*-ban az egyenlőség teljesül akkor* $\mathbf{x}$ *primál,* $\mathbf{z}$ *duál optimális megoldása az* LP *feladatpárnak.*

1.3. TÉTEL (dualitás). *Ha* P *és* D *nem üres akkor létezik olyan* $\mathbf{x} \in P$ *és* $\mathbf{z} \in D$*, amelyre a komplementaritási feltétel teljesül.*

1.4. KÖVETKEZMÉNY (erős egyensúly). *Ha* $\mathbf{x} \in P$ *és* $\mathbf{z} \in D$ *optimális megoldásai a primál és a duál feladatnak, akkor* (1.8)*-ban az egyenlőség teljesül.*

A továbbiakban az LPE feladat egyensúlyi feltételét perturbáljuk. Így olyan konvex programozási feladatot kapunk, amelyre vonatkozó alaplemma és a dualitási tétel speciális esetként tartalmazza az LP feladatra vonatkozó ezen tételeket, amennyiben a LP feladatnak van pozitív megoldása, azaz létezik olyan $\mathbf{z} \in D$, $\mathbf{x} \in P$, amelyekre $\mathbf{z}, \mathbf{x} > 0$ [7].

Először az egyensúlyi feladatot fogjuk felírni, majd definiálunk egy speciális konvex programozási feladatot és belátjuk, hogy egy paraméter (α vagy β) változtatásával az LP feladatot közelíti. Az általunk bevezetett egyensúlyi függvény a cth függvényből egyszerű transzformációval származtatható. Mivel a primál célfüggvény exponenciális függvény, ezért az így kapott speciális konvex programozási feladatot exponenciális programozási, azaz EP (α, β) feladatnak fogjuk nevezni. Az EP (α, β) feladat bemutatását és tárgyalását az LP feladattal analóg módon építjük fel.

2. Az EP (α, β) feladat

Ebben és a következő fejezetekben tegyük fel, hogy mind a primál, mind a duál feladatnak van pozitív megoldása. Ekkor feltehető, hogy $\widehat{\mathbf{x}} > \mathbf{0}, \widehat{\mathbf{z}} > \mathbf{0}$ és kielégítik (1.4)-et.

A LPE és LP feladatokat az alábbiakban definiált EPE (α, β) illetve EP (α, β) feladatattal fogjuk közelíteni. Legyen $\alpha, \beta \in R_+$, tetszőleges, de rögzített skalárok, $\widehat{\mathbf{x}}, \widehat{\mathbf{z}} \in R_+^n$ tetszőleges, de rögzített vektorok.

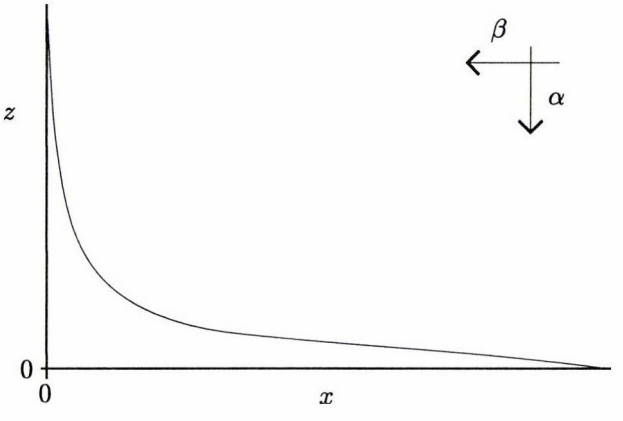
2.1. Feladat. Az EP (α, β) egyensúlyi feladata $\big(\mathrm{EPE}(\alpha, \beta)\big)$.

$$A\mathbf{x} = A\widehat{\mathbf{x}} \qquad \mathbf{z} = \widehat{\mathbf{z}} + A^T\mathbf{y}$$

$$\mathbf{x} > 0 \qquad \mathbf{z} > 0$$

Az egyensúlyi feltétel: $z_j = \dfrac{\alpha}{\exp\left(\frac{x_j}{\beta}\right) - 1} \ \forall j, \quad \Big(\text{inverze: } x_j = \beta \log \dfrac{\alpha + z_j}{z_j} \ \forall j\Big).$

Megjegyezzük, hogy az egyensúlyi feltétel alábbi, grafikus ábrája adta az ötletet a szerzőknek arra, hogy az EP (α, β) feladattal analitikusan közelítsék az LP feladatot. Vizuális típusú olvasóink számára megjegyezzük, hogy az ábrán lévő nyilak az jelzik, hogy az α illetve β paraméterrel, mely irányban tudjuk az EPE (α, β) feladat egyensúlyi függvényét, a vastag vonallal jelzett LPE egyensúlyi feltételéhez közelíteni.



1. ábra

Az LPE-hez hasonlóan az EPE (α, β) egyensúlyi feladat is megfogalmazható feladatpárként.

2.2. Feladat. Az EP (α, β) feladatpár.

Primál feladat

$$A\mathbf{x} = A\widehat{\mathbf{x}}$$

$$\mathbf{x} > \mathbf{0}$$

$$\min \rightarrow \widehat{\mathbf{z}}\mathbf{x} - \alpha\beta \sum_{j=1}^{n} \log\left(1 - \exp\left(\frac{-x_j}{\beta}\right)\right) \tag{2.1}$$

Duál feladat

$$\mathbf{z} = \hat{\mathbf{z}} + A^T\mathbf{y}$$

$$\mathbf{z} > \mathbf{0}$$

$$\min \rightarrow \mathbf{z}\hat{\mathbf{x}} + \beta \sum_{j=1}^{n} z_j \log z_j - \beta \sum_{j=1}^{n} (\alpha + z_j) \log (\alpha + z_j) \tag{2.2}$$

Vezessük be az alábbi jelöléseket:

$$P^0 = \{\mathbf{x} \in \mathbf{R}_+^n \mid A\mathbf{x} = A\hat{\mathbf{x}}\}$$

$$D^0 = \{\mathbf{z} \in \mathbf{R}_+^n, \mathbf{y} \in \mathbf{R}^m \mid \mathbf{z} = \hat{\mathbf{z}} + A^T\mathbf{y}\}$$

Azt állítjuk, hogy az EPE(α, β) feladat ekvivalens az EP(α, β) feladattal. Az ekvivalenciát és a megoldás létezését a következőkben egy alaplemmával és egy dualitási tétellel mutatjuk meg. Itt jegyezzük meg, hogy az EP(α, β) feladattal rokon, úgynevezett entrópia programozásra vonatkozó alaplemma és dualitási tétel bizonyítása megtalálható [8]-ban.

Az alaplemma és a dualitási tétel bizonyításában felhasználjuk az alábbi egyenlőtlenséget (legyen $a, b \in R_+$):

$$a \log \frac{a}{b} - a + b \geq 0, \qquad \forall a, b > 0, \tag{2.3}$$

és egyenlőség akkor és csak akkor van, ha $a = b$. Megjegyezzük, hogy (2.3) baloldala két pozitív szám (a, b) Kulbach-Leibler (K–L) eltérése néven is ismeretes [3].

2.1. LEMMA (alaplemma). *Minden* $\mathbf{x} \in \mathcal{P}^0$, $\mathbf{z} \in \mathcal{D}^0$ *és* $\alpha, \beta \in R_+$ *esetén fennáll az alábbi egyenlőtlenség:*

$$\hat{\mathbf{z}}\mathbf{x} - \alpha\beta \sum_{j=1}^{n} \log \left(1 - \exp\left(\frac{-x_j}{\beta}\right)\right) + \mathbf{z}\hat{\mathbf{x}} + \beta \sum_{j=1}^{n} z_j \log z_j - \tag{2.4}$$

$$-\beta \sum_{j=1}^{n} (\alpha + z_j) \log (\alpha + z_j) + n\alpha\beta \log \alpha \geq \hat{\mathbf{x}}\hat{\mathbf{z}}$$

és egyenlőség akkor és csak akkor áll fenn, ha

$$z_j = \frac{\alpha}{\exp\left(\frac{x_j}{\beta}\right) - 1} \qquad \forall j. \tag{2.5}$$

Bizonyítás. Legyenek

$$a_1 := \frac{z_j}{1+z_j}, \qquad b_1 := \frac{1}{\exp(x_j)} \tag{2.6}$$

$$a_2 := \frac{1}{1+z_j}, \qquad b_2 := \frac{\exp(x_j)-1}{\exp(x_j)}. \tag{2.7}$$

A (2.6) és (2.7) kifejezéseket (2.3)-ba helyettesítve az alábbi egyenlőtlenségeket kapjuk,

$$\frac{z_j}{1+z_j}\log\frac{\frac{z_j}{1+z_j}}{\frac{1}{\exp(x_j)}} - \frac{z_j}{1+z_j} + \frac{1}{\exp(x_j)} \geq 0 \qquad \forall j \tag{2.8}$$

$$\frac{1}{1+z_j}\log\frac{\frac{1}{1+z_j}}{\frac{\exp(x_j)-1}{\exp(x_j)}} - \frac{1}{1+z_j} + \frac{\exp(x_j)-1}{\exp(x_j)} \geq 0 \qquad \forall j. \tag{2.9}$$

Mindkét esetben egyenlőség akkor és csak akkor áll fenn, ha

$$a_1 = b_1 \text{ illetve } a_2 = b_2, \quad \text{azaz ha} \quad z_j = \frac{1}{\exp(x_j)-1}.$$

Összeadva és rendezve a (2.8) és (2.9) egyenlőtlenségeket, a következőket kapjuk

$$\frac{z_j}{1+z_j}\log\frac{z_j}{1+z_j} + \frac{1}{1+z_j}\log\frac{1}{1+z_j} + x_j - \frac{1}{1+z_j}\log\big(\exp(x_j)-1\big) \geq 0 \quad \forall j.$$

Szorozzuk meg az egyenlőtlenséget az $1+z_j$ pozitív számmal és elemi átalakítások után kapjuk az alábbi összefüggést:

$$z_j\log z_j - z_j\log(1+z_j) - \log(1+z_j) + z_j x_j + x_j - \log\big(\exp(x_j)-1\big) \geq 0 \quad \forall j.$$

Felhasználva, hogy $x_j = \log\big(\exp(x_j)\big)$ az utolsó két tagot vonjuk össze.

$$z_j\log z_j - z_j\log(1+z_j) - \log(1+z_j) + z_j x_j + \log\frac{\exp(x_j)}{\exp(x_j)-1} \geq 0 \quad \forall j.$$

Most x_j helyére írjunk $\frac{x_j}{\beta}$-t, illetve z_j helyére $\frac{z_j}{\alpha}$-t, majd szorozzunk $\alpha\beta$-val és egyszerű átalakítások után a következő egyenlőtlenséget kapjuk:

$$\alpha x_j\log z_j - \alpha x_j\log(\alpha+z_j) - \alpha\beta\log(\alpha+z_j) + \alpha\beta\log\alpha + z_j x_j -$$
$$-\alpha\beta\log\left(1-\exp\left(\frac{-x_j}{\beta}\right)\right) \geq 0 \quad \forall j.$$

$\forall j$–re történő szummázás után, felhasználva (1.7)-et a kívánt egyenlőtlenséget kapjuk.

$$\hat{\mathbf{z}}\mathbf{x} - \alpha\beta \sum_{j=1}^{n} \log\left(1 - \exp\left(\frac{-x_j}{\beta}\right)\right) + \mathbf{z}\hat{\mathbf{x}} + \beta \sum_{j=1}^{n} z_j \log z_j -$$

$$-\beta \sum_{j=1}^{n} (\alpha + z_j) \log(\alpha + z_j) + n\alpha\beta \log \alpha \geq \hat{\mathbf{x}}\hat{\mathbf{z}}. \qquad \square$$

Megjegyezzük, hogy bár a következő tétel bizonyítása általánosabb formában megtalálható például [1]-ben, [2]-ben és [13]-ban, egy rövid és egyszerűbb bizonyítást közlünk az alábbiakban.

2.2. KÖVETKEZMÉNY (gyenge egyensúly). *Ha* $\mathbf{x}_{(\alpha,\beta)} \in \mathcal{P}^0$ *és* $\mathbf{z}_{(\alpha,\beta)} \in \mathcal{D}^0$, *valamint (2.4)-ben az egyenlőség teljesül, akkor* $\mathbf{x}_{(\alpha,\beta)}$ *és* $\mathbf{z}_{(\alpha,\beta)}$ *optimális megoldása az* EP (α, β) *feladatpárnak.*

Bizonyítás. Elemi meggondolásból következik. $\square$

2.3. TÉTEL (dualitás). *Létezik egy és csak egy olyan* $\mathbf{x}^*_{(\alpha,\beta)} \in \mathcal{P}^0$ *és* $\mathbf{z}^*_{(\alpha,\beta)} \in \mathcal{D}^0$, *amelyre (2.4)-ben az egyenlőség teljesül.*

Bizonyítás. I. Könnyen igazolhatók az alábbi állítások:

(i) A primál célfüggvény szeparábilis, tagonként szigorúan konvex, folytonos függvény.

(ii)

$$\lim_{x_j \to 0+0} \hat{z}_j x_j - \alpha\beta \log\left(1 - \exp\left(\frac{-x_j}{\beta}\right)\right) = +\infty, \quad \forall j.$$

(iii) Mivel $\hat{z} > 0$, ezért

$$\lim_{x_j \to +\infty} \hat{z}_j x_j - \alpha\beta \log\left(1 - \exp\left(\frac{-x_j}{\beta}\right)\right) = +\infty, \quad \forall j.$$

(iv) A primál célfüggvény nívóhalmazai $\forall \alpha, \beta \in R_+$ esetén zártak és korlátosak.

$$\left\{ \mathbf{x} > 0 \middle| \hat{\mathbf{z}}\mathbf{x} - \alpha\beta \sum_{j=1}^{n} \log\left(1 - \exp\left(\frac{-x_j}{\beta}\right)\right) \leq \right.$$

$$\left. \leq \hat{\mathbf{z}}\hat{\mathbf{x}} - \alpha\beta \sum_{j=1}^{n} \log\left(1 - \exp\left(\frac{-\hat{x}_j}{\beta}\right)\right) \right\}$$

A fentiekből következik, hogy a

$$\left\{ \mathbf{x} > 0 \middle| \widehat{\mathbf{z}}\mathbf{x} - \alpha\beta \sum_{j=1}^{n} \log\left(1 - \exp\left(\frac{-x_j}{\beta}\right)\right) \leq \right.$$

$$\left. \leq \widehat{\mathbf{z}}\widehat{\mathbf{x}} - \alpha\beta \sum_{j=1}^{n} \log\left(1 - \exp\left(\frac{-\widehat{x}_j}{\beta}\right)\right), \quad \mathbf{x} = \widehat{\mathbf{x}} + A^T\mathbf{y} \right\}$$

is zárt korlátos halmaz. Felhasználva Weierstrass tételét, miszerint folytonos függvény zárt korlátos tartományon felveszi minimumát, azt kapjuk, hogy a primál célfüggvény felveszi minimumát valamely $\mathbf{x}^*_{(\alpha,\beta)} \in \mathcal{P}^0$ pontban. Hasonló módon igazolható, hogy a duál célfüggvény is felveszi minimumát valamely $\mathbf{z}^*_{(\alpha,\beta)} \in \mathcal{D}^0$-pontban.

II.a., Jelölje $a^{(i)}$ az A mátrix $i-$dik sorvektorát és $\mathbf{x}^*_{(\alpha,\beta)}$ az EP(α,β) feladat optimális megoldását. Legyen

$$z_j^* = \frac{\alpha}{\exp\left(\frac{x^*(\alpha,\beta)_j}{\beta}\right) - 1} > 0$$

$\forall j$-re. Megmutatjuk, $\mathbf{x}^*_{(\alpha,\beta)}$ és $\mathbf{z}^*$ az EPE(α,β) egyensúlyi feladat megoldása. Ha most nemlétezik olyan $\mathbf{y}$ amelyre $\mathbf{z}^* - \widehat{\mathbf{z}} = A^T\mathbf{y}$ azaz $\mathbf{z}^* - \widehat{\mathbf{z}} \notin \mathcal{L}\left(a^{(1)}, \ldots, a^{(i)}, \ldots, a^{(m)}\right)$ akkor létezik olyan $\widetilde{\mathbf{x}} \in \mathcal{L}^{\perp}\left(a^{(1)}, \ldots, a^{(i)}, \ldots, a^{(m)}\right)$ amelyre $\widetilde{\mathbf{x}}(\mathbf{z}^* - \widehat{\mathbf{z}}) \neq 0$. Tehát, ha $\widetilde{\mathbf{x}}(\mathbf{z}^* - \widehat{\mathbf{z}}) = 0$ akkor $\mathbf{z}^* \in \mathcal{D}^0$. Legyen $\mathbf{x}^{**}_{(\alpha,\beta)} = \mathbf{x}^*_{(\alpha,\beta)} + \vartheta\widetilde{\mathbf{x}}_{(\alpha,\beta)}$, ahol nyilván $\exists\vartheta > 0$ amelyre $\mathbf{x}^{**}_{(\alpha,\beta)} > 0$. Mivel feltevésünk szerint $\mathbf{x}^*_{(\alpha,\beta)}$ primál optimális megoldás, ezért

$$\frac{d}{d\vartheta} \sum_{j=1}^{n} \left(\widehat{z}_j \left(x^*_{(\alpha,\beta)j} + \vartheta\widetilde{x}_{(\alpha,\beta)j} \right) - \right.$$

$$\left. - \alpha\beta \log\left(1 - \exp\left(\frac{-\left(x^*_{(\alpha,\beta)j} + \vartheta \sum_{i=1}^{m} a_{ij}\right)}{\beta}\right)\right)\right)\Bigg|_{\vartheta=0} =$$

$$= \mathbf{x}^*_{(\alpha,\beta)}(\widehat{\mathbf{z}} - \mathbf{z}^*) = 0,$$

azaz $\mathbf{z}^* \in \mathcal{D}^0$. Tehát, ha a primál célfüggvény az $\mathbf{x}^*_{(\alpha,\beta)} \in \mathcal{P}^0$ helyen veszi fel a minimumát és

$$z_j^* = \frac{\alpha}{\exp\left(\frac{x^*(\alpha,\beta)_j}{\beta}\right) - 1}$$

$\forall j$-re, akkor $\mathbf{z}^* \in \mathcal{D}^0$.

II.b., Legyen $\mathbf{z}^*_{(\alpha,\beta)}$ az EP(α,β) feladat duál optimális megoldása és $x^*_j = \beta \log \frac{\alpha + z^*(\alpha,\beta)_j}{z^*(\alpha,\beta)_j} > 0 \quad \forall j$-re. Megmutatjuk, hogy ekkor $\mathbf{x}^*$ és $\mathbf{z}^*_{(\alpha,\beta)}$ az EPE(α,β) egyensúlyi feladat megoldása. Legyen $\mathbf{z}^{**} = \mathbf{z}^*_{(\alpha,\beta)} + A^T\mathbf{y}$, ahol nyilván $\exists \mathbf{y} > \mathbf{0}$, hogy $\mathbf{z}^{**} > \mathbf{0}$. Mivel feltevésünk szerint $\mathbf{z}^*_{(\alpha,\beta)}$ duál optimális megoldás ezért

$$\frac{d}{dy_i} \sum_{j=1}^{n} \left(\widehat{x}_j z^*_{(\alpha,\beta)j} - \beta z^{**}_{(\alpha,\beta)j} \log z^{**}_{(\alpha,\beta)j} - \beta\left(\alpha + z^{**}_{(\alpha,\beta)j}\right) \log \left(\alpha + z^{**}_{(\alpha,\beta)j}\right)\right)\Bigg|_{y_i=0} =$$

$$= \mathbf{a}^{(i)}\widehat{\mathbf{x}} - \mathbf{a}^{(i)}\mathbf{x}^* = 0,$$

$\forall i$-re, azaz $\mathbf{x}^* \in \mathcal{P}^0$. Tehát, ha a duál célfüggvény a $\mathbf{z}^*_{(\alpha,\beta)} \in \mathcal{D}^0$ helyen veszi fel a minimumát és $x^*_j = \beta \log \frac{\alpha + z^*(\alpha,\beta)_j}{z^*(\alpha,\beta)_j}$ $\forall j$-re akkor $\mathbf{x}^* \in \mathcal{P}^0$.

Ha $\mathbf{x}^*_{(\alpha,\beta)} \in \mathcal{P}^0$ és $\mathbf{z}^*_{(\alpha,\beta)} \in \mathcal{D}^0$ az EPE(α,β) egyensúlyi feladat megoldásai, akkor (2.4)-ben az egyenlőség teljesül és a 2.2. Következmény szerint $\mathbf{x}^*_{(\alpha,\beta)}$ és $\mathbf{z}^*_{(\alpha,\beta)}$ az EP(α,β) feladat optimális megoldásai. Ha a primál célfüggvény az $\mathbf{x}^*_2$ helyen is felveszi minimumát, akkor (2.4)-ben nyilván $\mathbf{x}^*_{(\alpha,\beta)}$ helyére $\mathbf{x}^*_2$-őt írva is teljesül az egyenlőség. Ekkor azonban a 2.1. Lemma szerint fennáll, hogy $x^*_{2j} = \beta \log \frac{\alpha + z^*(\alpha,\beta)_j}{z^*(\alpha,\beta)_j}$, $\forall j$-re, tehát $\mathbf{x}^*_{(\alpha,\beta)} = \mathbf{x}^*_2$, ezért az EP$(\alpha,\beta)$ feladat primál optimális megoldása $\mathbf{x}^*_{(\alpha,\beta)}$ egyértelmű. Hasonlóképpen belátható, hogy az EP(α,β) feladat duál optimális megoldása is egyértelmű. Ezzel a bizonyítást befejeztük. □

2.4. KÖVETKEZMÉNY (erős egyensúly). *Ha* $\mathbf{x}_{(\alpha,\beta)} \in P^0$, *és* $\mathbf{z}_{(\alpha,\beta)} \in D^0$ *optimális megoldásai az* EP(α,β) *feladatnak, akkor* (2.4) *egyenlőséggel teljesül.*

Bizonyítás. Elemi meggondolásokból adódik. □

3. Az LP feladat közelítése az EP(α,β) feladattal

A továbbiakban megmutatjuk, hogy az EP(α,β) feladat a LP feladat analitikus approximációjának tekinthető, mert ha $\alpha\beta \to 0$ akkor EP$(\alpha,\beta) \to$LP.

3.1. LEMMA. *Legyenek* $\alpha, \beta \in R_+$ *tetszőleges, de rögzített skalárok, valamint* $z_j := \frac{\alpha}{\exp\left(\frac{x_j}{\beta}\right) - 1}$ $\forall j$, *ahol* $\mathbf{x}, \mathbf{z} \in \mathbf{R}^n$. *Ekkor* $0 \le x_j z_j \le \alpha\beta$, $\forall j$-re.

Bizonyítás. Könnyen igazolhatók az alábbi állítások, amelyekből nyilvánvalóan következik a lemma állítása:

(i)

$$\lim_{x_j \to 0} \frac{\alpha x_j}{\exp\left(\frac{x_j}{\beta}\right) - 1} = \alpha\beta;$$

(ii)

$$\lim_{z_j \to \infty} \frac{\alpha x_j}{\exp\left(\frac{x_j}{\beta}\right) - 1} = 0;$$

(iii)

$$\frac{\alpha x_j}{\exp\left(\frac{x_j}{\beta}\right) - 1} \quad \text{monoton csökkenő } (0, \infty)\text{-en;}$$

(iv)

$$x_j z_j = \frac{\alpha x_j}{\exp\left(\frac{x_j}{\beta}\right) - 1}. \qquad \square$$

3.2. LEMMA. *Legyenek* $\mathbf{x}_{(\alpha,\beta)}$ *és* $\mathbf{z}_{(\alpha,\beta)}$ *optimális megoldásai* EP(α, β)*-nak,* $\alpha \leq \alpha^0$, $\beta \leq \beta^0$, *és* $K := \frac{\widehat{xz} + n\alpha^0\beta^0}{\min_j(\widehat{x}_j, \widehat{z}_j)} > 0$. *Ekkor*

$$x_{(\alpha,\beta)j}, \quad z_{(\alpha,\beta)j} \leq K \quad \forall j\text{-re.}$$

Bizonyítás. A 3.1. Lemmából és (1.7)-ből következik, hogy

$$\mathbf{x}_{(\alpha,\beta)}\widehat{\mathbf{z}} + \widehat{\mathbf{x}}\mathbf{z}_{(\alpha,\beta)} \leq \widehat{\mathbf{x}}\widehat{\mathbf{z}} + n\alpha\beta \leq \widehat{\mathbf{x}}\widehat{\mathbf{z}} + n\alpha^0\beta^0.$$

Mivel $\mathbf{x}_{(\alpha,\beta)}\widehat{\mathbf{z}} > 0$, $\widehat{\mathbf{x}}\mathbf{z}_{(\alpha,\beta)} > 0$, $\widehat{\mathbf{x}}, \widehat{\mathbf{z}} > 0$, következik az állítás. $\square$

3.3. KÖVETKEZMÉNY. *Legyenek* $\mathbf{x}_{(\alpha,\beta)}$ *és* $\mathbf{z}_{(\alpha,\beta)}$ *optimális megoldásai* EP(α, β)*-nak,* α^0, β^0 *adott skalár értékek és* $\alpha \leq \alpha^0$, $\beta \leq \beta^0$. *Ha* $\alpha\beta \to 0$ *akkor az* $\mathbf{x}_{(\alpha,\beta)}$ *és* $\mathbf{z}_{(\alpha,\beta)}$ *optimális megoldások sorozatából kiválasztható egy konvergens részsorozat úgy, hogy minden* $x_{(\alpha,\beta)j}$, $z_{(\alpha,\beta)j}$ *tagnak a* $(K, 0]$ *halmazon van torlódási pontja.*

Bizonyítás. Nyilvánvaló (lásd [18]). $\square$

3.4. LEMMA. *Legyenek* $\mathbf{x}_{(\alpha,\beta)}$ *és* $\mathbf{z}_{(\alpha,\beta)}$ *optimális megoldásai* EP(α, β)*-nak. Tegyük fel, hogy* $\alpha\beta \to 0$. *Jelölje* $\mathbf{x}^*$ *és* $\mathbf{z}^*$ *az* $\mathbf{x}_{(\alpha,\beta)}$ *és* $\mathbf{z}_{(\alpha,\beta)}$ *optimális megoldások sorozatából kiválasztható egyik konvergens részsorozat torlódási pontját. Ekkor igazak az alábbi állítások.*

(i) $\lim_{\alpha\beta \to 0} \mathbf{x}_{(\alpha,\beta)}\mathbf{z}_{(\alpha,\beta)} = \mathbf{x}^*\mathbf{z}^* = 0,$

(ii) $\mathbf{x}^*$ *és* $\mathbf{z}^*$ *optimális megoldásai az* LP *primál és duál feladatnak.*

Bizonyítás. (i) a 3.1. Lemma és a 3.2. Lemma következménye. Mivel $\mathbf{x}_{(\alpha,\beta)} \in P^0$ és $\mathbf{z}_{(\alpha,\beta)} \in D^0$ $\forall(\alpha, \beta) > 0$-ra ezért a torlódási pontban $\mathbf{x}^* \in P$ és $\mathbf{z}^* \in D$. Az 1.2. Következmény szerint ebből következik (ii). $\square$

3.5. KÖVETKEZMÉNY. *Legyenek* $\mathbf{x}_{(\alpha,\beta)}$ *és* $\mathbf{z}_{(\alpha,\beta)}$ *optimális megoldásai* EP(α, β)*-nak. Ekkor az alábbi egyenlőtlenségek fennállnak.*

$$\frac{\alpha K}{\exp\left(\frac{K}{\beta}\right) - 1} \leq x_{(\alpha,\beta)j} z_{(\alpha,\beta)j} \leq \alpha\beta, \quad \forall j.$$

Bizonyítás. A jobboldali egyenlőtlenség a 3.1. Lemma állítása. A baloldali egyenlőtlenség a 3.1. Lemma(iii) állításából és a 3.2. Lemmából következik. □

A továbbiakban tegyük fel, hogy az A mátrix sorai függetlenek és tekintsük az EP (α, β) feladatból származtatott EP $(\alpha, 1)$ és az EP $(1, \beta)$ speciális feladatokat.

A következőkben az LP célfüggvényekkel kapcsolatos állításokat teszünk.

3.6. LEMMA. *Jelölje* $B \in \mathbf{R}^{n\times(n-m)}$ *az* A *mátrix null tér mátrixát. Ha* $\mathbf{x} \in P$, $\mathbf{z}, \mathbf{y} \in D$ *akkor igazak az alábbi állítások.*

(i) *Az* LP *primál célfüggvény,* $\widehat{\mathbf{z}}\mathbf{x}$, *akkor és csak akkor konstans* P^0*-n, ha létezik olyan* $\mathbf{y}$ *amelyre,*

$$\widehat{\mathbf{z}} = A^T\mathbf{y}. \tag{3.1}$$

(ii) *Az* LP *duál célfüggvény,* $\widehat{\mathbf{x}}\mathbf{z}$, *akkor és csak akkor konstans* D^0*-n, ha létezik olyan* $\mathbf{t} \in \mathbf{R}^n$ *amelyre*

$$\widehat{\mathbf{x}} = \mathbf{t}B.$$

Bizonyítás. (i) Először belátjuk, hogy ha (3.1) fennáll, akkor $\widehat{\mathbf{z}}\mathbf{x}$ konstans P^0-n. Induljunk ki az alábbi egyenletrendszerből.

$$A\mathbf{x} = A\widehat{\mathbf{x}}$$

Szorozzuk meg mindkét oldalt $\mathbf{y}$-nal, és használjuk fel (3.1)-t:

$$\widehat{\mathbf{z}}\mathbf{x} = \widehat{\mathbf{z}}\widehat{\mathbf{x}}.$$

Most belátjuk, hogy ha $\widehat{\mathbf{z}}\mathbf{x}$ konstans P^0-n akkor (3.1) fennáll. Ha ugyanis $\widehat{\mathbf{z}}\mathbf{x}$ konstans P^0-n akkor fennáll, hogy

$$\widehat{\mathbf{z}}\mathbf{x} = \widehat{\mathbf{z}}\widehat{\mathbf{x}}$$

Átrendezve és felhasználva (1.4)-et kapjuk, hogy:

$$\widehat{\mathbf{z}}(\mathbf{x} - \widehat{\mathbf{x}}) = 0 \qquad \text{és} \qquad A(\mathbf{x} - \widehat{\mathbf{x}}) = 0,$$

amiből következik az állítás. Hasonlóképpen bizonyítható (ii), ezért a bizonyítást az olvasóra bízzuk. □

3.7. TÉTEL. (i) *Tekintsük az* EP $(\alpha, 1)$ *feladatot. Legyen* $\alpha^{**} < \alpha^*$ *és az* α^**-hoz tartozó optimális megoldás* $\mathbf{x}^*$ *valamint az* α^{**}*-hoz tartozó optimális megoldás* $\mathbf{x}^{**}$. *Ekkor* $\widehat{\mathbf{z}}\mathbf{x}^{**} \leq \widehat{\mathbf{z}}\mathbf{x}^*$ *és egyenlőség akkor és csak akkor van, ha* $\widehat{\mathbf{z}}\mathbf{x}$ *konstans az egész megengedett tartományon.*

(ii) *Tekintsük az* EP $(1, \beta)$ *feladatot. Legyen* $\beta^{**} < \beta^*$ *és a* β^**-hoz tartozó optimális megoldás* $\mathbf{z}^*$ *valamint a* β^{**}*-hoz tartozó optimális megoldások* $\mathbf{z}^{**}$. *Ekkor*

$\mathbf{z}^{**}\widehat{\mathbf{x}} \leq \mathbf{z}^{*}\widehat{\mathbf{x}}$ *és egyenlőség akkor és csak akkor van, ha* $\mathbf{z}\widehat{\mathbf{x}}$ *konstans az egész megengedett tartományon.*

Bizonyítás. (i) Az adott $\mathbf{x}^*, \mathbf{x}^{**}$ vektorok optimalitását felírva az alábbi egyenlőtlenségek adódnak.

$$\widehat{\mathbf{z}}\mathbf{x}^{*} - \alpha^{*}\sum_{j=1}^{n}\log\left(1-\exp\left(-x_j^{*}\right)\right) \leq \widehat{\mathbf{z}}\mathbf{x}^{**} - \alpha^{*}\sum_{j=1}^{n}\log\left(1-\exp\left(-x_j^{**}\right)\right) \tag{3.2}$$

$$\widehat{\mathbf{z}}\mathbf{x}^{**} - \alpha^{**}\sum_{j=1}^{n}\log\left(1-\exp(-x_j^{**})\right) \leq \widehat{\mathbf{z}}\mathbf{x}^{*} - \alpha^{**}\sum_{j=1}^{n}\log\left(1-\exp\left(-x_j^{*}\right)\right) \tag{3.3}$$

Mindkét esetben egyenlőség akkor és csak akkor, ha $\mathbf{x}^{**} = \mathbf{x}^{*}$. (3.2)-t szorozva $\frac{\alpha^{**}}{\alpha^{*}}$-gal és hozzáadva (3.3)-t, kapjuk: $\mathbf{x}^{**}\widehat{\mathbf{z}}\left(1-\frac{\alpha^{**}}{\alpha^{*}}\right) \leq \mathbf{x}^{*}\widehat{\mathbf{z}}\left(1-\frac{\alpha^{**}}{\alpha^{*}}\right)$, ahol $1-\frac{\alpha^{**}}{\alpha^{*}} > 0$, tehát $\widehat{\mathbf{z}}\mathbf{x}^{**} \leq \widehat{\mathbf{z}}\mathbf{x}^{*}$.

Hasonló módon bizonyítható (ii) is, ezért ezt az olvasóra bízzuk. □

A továbbiakban tegyük fel, hogy nem létezik (3.1)-et kielégítő $\mathbf{y}$ vektor. A következő tétel rávilágít arra, hogyan lehet az LP-t az EP(α,β) feladat segítségével tetszőleges pontossággal közelíteni.

3.8. TÉTEL. *Tekintsük az* EP$(\alpha,1)$ *feladatot. Tegyük fel, hogy az* A *mátrix sorai függetlenek. Jelölje* $\mathbf{x}_{(\alpha,1)}, \mathbf{y}_{(\alpha,1)}, \mathbf{z}_{(\alpha,1)}$ *az adott* α*-hoz tartozó optimális megoldást és legyen* $\alpha^0 \geq \alpha > 0$. *Ekkor* $\mathbf{x}_{(\alpha,1)}, \mathbf{y}_{(\alpha,1)}, \mathbf{z}_{(\alpha,1)}$ α *szerint folytonosan deriválható görbe.*

Bizonyítás. A bizonyítás [5] ötlete alapján a következő.

Tekintsük az EP$(\alpha,1)$ feladattal ekvivalens alábbi EPE$(\alpha,1)$ feladatot.

$$\begin{aligned} A\mathbf{x}_{(\alpha,1)} &= A\widehat{\mathbf{x}} \\ \mathbf{x}_{(\alpha,1)} &> \mathbf{0} \\ \frac{\alpha}{\exp\left(x_{(\alpha,\beta)j}\right)-1} &= \widehat{z}_j + \sum_{i=1}^{m} y_{(\alpha,1)i}a_{ij} \qquad \forall j. \end{aligned} \tag{3.4}$$

A 2.3. Tétel szerint, ennek minden α-ra egy és csak egy megoldása van. Jelöljük ezt $\left(\mathbf{x}_{(\alpha,1)}, \mathbf{y}_{(\alpha,1)}\right)$-gyel. Az implicitfüggvény-tétele (lásd [18]) értelmében α kis környezetében $\left(\mathbf{x}_{(\alpha,\beta)}, \mathbf{y}_{(\alpha,1)}\right)$ α folytonosan deriválható függvénye, ha az alábbi mátrix teljes rangú az $\left(\mathbf{x}_{(\alpha,1)}, \mathbf{y}_{(\alpha,1)}\right)$ pontban.

$$\begin{pmatrix} \operatorname{diag}\left(\alpha\frac{\partial\frac{\alpha}{\exp(x_{(\alpha,1)j})-1}}{\partial x_{(\alpha,1)j}}\right) & A^T \\ A & 0 \end{pmatrix}$$

Feltevéseinkből következik, hogy a fenti mátrix invertálható, tehát $\mathbf{x}_{(\alpha,1)}, \mathbf{y}_{(\alpha,1)}$ α szerint folytonosan deriválható görbe az $\alpha \in (0, \alpha^0]$ tartományon. Mivel $\mathbf{y}_{(\alpha,1)}$ folytonosan deriválható, következik, hogy $\mathbf{z}_{(\alpha,1)}$ szintén folytonosan deriválható görbe az $\alpha \in (0, \alpha^0]$ tartományon. □

A következőkben definiálunk egy úgynevezett exponenciális centrumot. Megmutatjuk, hogy az exponenciális centrum karakterizálja az EP$(\alpha, 1)$ feladat optimális megoldását. Célunk az, hogy az exponenciális centrum segítségével definiáljuk az $\mathbf{x}_{(\alpha,1)}$, $\alpha \to 0$-hoz $(\alpha^0 > \alpha > 0)$ határértékét, $\left(\mathbf{x}^* = \lim_{\alpha \to 0} \mathbf{x}_{(\alpha,1)}\right)$-et. Jelöljön $\mathbf{x}^*$ egy LP optimális megoldást, továbbá legyen $\mu = \widehat{\mathbf{z}}\mathbf{x}^*$. Az általánosság elvének megsértése nélkül feltehető, hogy μ pozitív.

3.9. Definíció. Jelölje $\mathbf{x}_{(\alpha,1)}$ az EP(α, β) optimális megoldását és legyen ε olyan pozitív szám, amelyre teljesül, hogy $\mathbf{x}_{(\alpha,1)}\widehat{z} = \mu(1+\varepsilon)$. Ekkor az alábbi (3.5) feladat optimális megoldását (jelöljük $\mathbf{x}_{(\varepsilon)}$-nal), az EP$(\alpha, 1)$ — primál célfüggvénye szerinti — exponenciális centrumának nevezzük.

$$
\begin{aligned}
(3.5) \qquad & A\mathbf{x} = A\widehat{\mathbf{x}} \\
& \mathbf{x}\widehat{\mathbf{z}} = \mu(1+\varepsilon) \\
& \mathbf{x} > \mathbf{0} \\
& \min \to \widehat{\mathbf{z}}\mathbf{x} - \sum_{j=1}^{n} \log\left(1 - \exp\left(-x_j\right)\right)
\end{aligned}
$$

Megjegyezzük, hogy az exponenciális centrum analóg fogalom az úgynevezett analitikus centrum (lásd [14]) fogalommal. A következő tételben — amelynek analóg megfelelője megtalálható [14]-ben — megmutatjuk a kapcsolatot az EP$(\alpha, 1)$ feladat és az exponenciális centrum között.

3.10. TÉTEL. *Az* EP$(\alpha, 1)$ *feladat optimális megoldása legyen* $\mathbf{x}_{(\alpha,1)}$. *Ha* $\mathbf{x}_{(\alpha,1)}\widehat{\mathbf{z}} = \mu(1+\varepsilon)$ *akkor* $\mathbf{x}_{(\varepsilon)} = \mathbf{x}_{(\alpha,1)}$.

Bizonyítás. Megmutatjuk, hogy (3.5) és az EPE$(\alpha, 1)$ ekvivalensek. Legyenek $\eta_{(\varepsilon)} \in R$, $y_{(\alpha,1)}, y_{(\varepsilon)} \in \mathbf{R}^m$ valamint $P_{(\varepsilon)} := \left\{\mathbf{x} \in \mathbf{R}^n_+ \mid A\mathbf{x} = A\widehat{\mathbf{x}},\ \mathbf{x}\widehat{\mathbf{z}} = \mu(1+\varepsilon)\right\}$. Tekintsük az EP$(\alpha,1)$ feladatot. Feltevésünk szerint ennek optimális $\mathbf{x}_{(\alpha,1)}$ megoldására fennáll, hogy $\mathbf{x}_{(\alpha,1)}\widehat{\mathbf{z}} = \mu(1+\varepsilon)$. Az EPE$(\alpha, 1)$ feladat ekkor az alábbi formát ölti:

$$
\begin{aligned}
(3.6) \qquad & A\mathbf{x}_{(\alpha,1)} = A\widehat{\mathbf{x}} \qquad \mathbf{z}_{(\alpha,1)} = \widehat{\mathbf{z}} + A^T\mathbf{y}_{(\alpha,1)} \\
& \mathbf{x}_{(\alpha,1)} > \mathbf{0} \qquad \mathbf{z}_{(\alpha,1)} > \mathbf{0} \\
(3.7) \qquad & z_{(\alpha,1)j} = \frac{\alpha}{\exp\left(x_{(\alpha,1)j}\right) - 1} \quad \forall j.
\end{aligned}
$$

Valamint $\mathbf{x}_{(\alpha,1)}\widehat{\mathbf{z}} = \mu(1+\varepsilon)$.

(3.7)-et behelyettesítve (3.6) jobboldali egyenletrendszerébe, kapjuk az alábbi egyenletrendszert.

$$\frac{1}{\exp\left(x_{(\alpha,1)j}\right)-1} = \frac{\widehat{z}_j}{\alpha} + \sum_{i=1}^{m} \frac{y_{(\alpha,1)i}}{\alpha} a_{ij} \qquad \forall j. \tag{3.8}$$

Most tekintsük a (3.5)-tel ekvivalens egyensúlyi feladatot. (Az ekvivalencia az A mátrix speciális megválasztásával a 2.3. Tételből következik. Ezt az olvasó könnyen ellenőrizheti.)

$$\begin{gathered} A\mathbf{x}_{(\varepsilon)} = A\widehat{\mathbf{x}} \qquad \mathbf{z}_{(\varepsilon)} = \widehat{\mathbf{z}}\left(1+\eta_{(\varepsilon)}\right) + A^T\mathbf{y}_{(\varepsilon)} \\ \mathbf{x}_{(\varepsilon)}\widehat{\mathbf{z}} = \mu(1+\varepsilon) \\ \mathbf{x}_{(\varepsilon)} > \mathbf{0} \qquad \mathbf{z}_{(\varepsilon)} > 0 \\ z_{(\varepsilon)j} = \frac{1}{\exp\left(x_{(\varepsilon)j}\right)-1} \qquad \forall j. \end{gathered} \tag{3.9}$$

$\left(z_{(\varepsilon)j} = \frac{1}{\exp(x_{(\varepsilon)j})-1} - \forall j\right)$-t behelyettesítve (3.9) jobboldali egyenletrendszerébe, kapjuk az alábbi egyenletrendszert.

$$\frac{1}{\exp\left(x_{(\varepsilon)j}\right)-1} = \widehat{z}_j\left(1+\eta_{(\varepsilon)}\right) + \sum_{i=1}^{m} y_{(\varepsilon)i}a_{ij} \qquad \forall j. \tag{3.10}$$

Tegyük fel, hogy $\mathbf{x}_{(\varepsilon)} \neq \mathbf{x}_{(\alpha,1)}$. Mivel $\mathbf{x}_{(\varepsilon)}, \mathbf{x}_{(\alpha,1)} \in P_{(\varepsilon)}$ ezért a $\eta_{(\varepsilon)} := \frac{1}{\alpha} - 1$ és $\mathbf{y}_{(\varepsilon)} := \frac{\mathbf{y}_{(\alpha,1)}}{\alpha}$ választással $\mathbf{x}_{(\varepsilon)} = \mathbf{x}_{(\alpha,1)}$ is megoldása (3.9)-nek, ami ellentmond a 2.3. Tételnek. Ezzel a bizonyítást befejeztük. □

3.11. KÖVETKEZMÉNY. *Legyen $\alpha^0 > \alpha > 0$ és jelölje az* EP$(\alpha,1)$ *feladat optimális megoldását $\mathbf{x}_{(\alpha,1)}$ továbbá legyen $\varepsilon = \frac{\mathbf{x}_{(\alpha,1)}\widehat{\mathbf{z}}-\mu}{\mu}$.*

(i) *Ekkor minden $\varepsilon^0 > \varepsilon > 0$-ra létezik optimális megoldása a 3.9. Definícióban megadott centrum feladatnak, ahol $\varepsilon^0 = \frac{\mathbf{x}_{(\alpha^0,1)}\widehat{\mathbf{z}}-\mu}{\mu}$.*

(ii) *Ha $\alpha\beta \to 0$, akkor $\varepsilon \to 0$.*

Bizonyítás. (i) $\mathbf{x}_{(\alpha,1)}\widehat{\mathbf{z}}$ P^0-n folytonos függvény az $\alpha^0 > \alpha > 0$ tartományban. Ebből következik az állítás. (ii) Nyilvánvaló. □

A továbbiakban ε legyen olyan pozitív szám, amelyre teljesül, hogy $\varepsilon = \frac{\mathbf{x}_{(\alpha,1)}\widehat{\mathbf{z}}-\mu}{\mu}$.

A következő tételben felhasználjuk az alábbi ún. Kronecker tételt (lásd [10], [11], [16]).

KRONECKER-TÉTEL. *Az alábbi két rendszer közül az egyik és csak az egyik megoldható.*

I. *Létezik olyan* $\mathbf{x}$ *vektor, amelyre*

$$A\mathbf{x} = \mathbf{b}.$$

II. *Létezik olyan* $\mathbf{y}$ *vektor, amelyre*

$$A^T\mathbf{y} = 0,$$
$$\mathbf{y}\mathbf{b} \neq 0.$$

Vezessük be az alábbi jelöléseket $N = \{j \mid x_j = 0\}$, $B = \{j \mid x_j > 0\}$ és tegyük fel, hogy az $A\mathbf{x} = A\widehat{\mathbf{x}}$, $\mathbf{x}\widehat{\mathbf{z}} = \mu$, $\mathbf{x} \geq 0$ egyenletrendszer legbővebb pozitív megoldása az alábbi módon áll elő:

$$x_j = 0, \qquad \forall j \in N,$$
$$x_j > 0, \qquad \forall j \in B.$$

A következő 3.12. Lemmában és 3.13. Tételben definiáljuk $\lim\limits_{\alpha\to 0} \mathbf{x}_{(\alpha,1)} = \mathbf{x}^*$-ot.

3.12. LEMMA. *Tekintsük az* EP$(\alpha, 1)$ *feladatot. Jelölje* $\mathbf{x}_{(\alpha,1)}, \mathbf{z}_{(\alpha,1)}$ *az adott* α*-hoz tartozó optimális megoldást és legyen* $\overline{\mathbf{x}}, \overline{\mathbf{z}}$ *az LP feladat egyik legbővebb pozitív megoldáspárja valamint* $\delta := \min\left\{\overline{x}_j \mid j \in B \text{ és } \overline{z}_j \mid j \in N\right\}$. *Ekkor* $\lim\limits_{\alpha\to 0} \mathbf{x}_{(\alpha,1)} > 0$ $\forall j \in B$ *és* $\lim\limits_{\alpha\to 0} \mathbf{z}_{(\alpha,1)} > 0$ $\forall j \in N$.

Bizonyítás. Az 1.1. Lemma és a 3.5. Következmény felhasználásával kapjuk a következő becslést.

$$\sum_{j\in B} \overline{x}_j z_{(\alpha,1)j} + \sum_{j\in N} \overline{z}_j x_{(\alpha,1)j} = \sum_{j=1}^{n} x_{(\alpha,1)j} z_{(\alpha,1)j} \leq n\alpha.$$

Innen az alábbi összefüggéseket kapjuk.

Ha $j \in B$ akkor $\delta z_{(\alpha,1)j} \leq n\alpha$, $z_{(\alpha,1)j} \leq \frac{n\alpha}{\delta}$.

$\frac{1}{\exp(x_{(\alpha,1)j})-1} \leq \frac{n}{\delta}$ $\qquad x_{(\alpha,1)j} \geq \log\frac{n+\delta}{n} > 0$ $\forall j \in B$.

Ha $j \in N$ akkor $\delta x_{(\alpha,1)j} \leq n\alpha$, $x_{(\alpha,1)j} \leq \frac{n\alpha}{\delta}$.

Felhasználva, hogy $\lim\limits_{\alpha\to 0} \frac{n\alpha}{\delta} \frac{1}{\exp\left(\frac{n\alpha}{\delta}\right)-1} = 1$, kapjuk, hogy

$$z_{(\alpha,1)j} \geq \frac{n\alpha}{\delta} \frac{1}{\exp\left(\frac{n\alpha}{\delta}\right) - 1} \frac{\delta}{n} \geq \frac{\delta}{n} > 0 \quad \forall j \in N. \qquad \square$$

3.13. TÉTEL. *Tekintsük az* EP$(\alpha, 1)$ *feladatot. Jelölje* $\mathbf{x}_{(\alpha,1)}$ *az adott* α*-hoz tartozó optimális megoldást és legyen* $\lim\limits_{\alpha\to 0} \mathbf{x}_{(\alpha,1)} = \mathbf{x}^*$, *valamint* $\mu = x^*\widehat{z}$. *Ekkor* $\mathbf{x}^*$ *a következő feladat optimális megoldása.*

$$A\mathbf{x} = A\widehat{\mathbf{x}}$$

$$\mathbf{x}\widehat{\mathbf{z}} = \mu$$

$$\mathbf{x} \geq \mathbf{0}$$

$$\min \to \sum_{j \in B} x_j \widehat{z}_j - \log\left(1 - \exp\left(-x_j\right)\right)$$

Bizonyítás. Legyen $\varepsilon = \frac{x(\alpha,1)\widehat{z}-\mu}{\mu}$. Ekkor a 3.10. Tétel szerint az $\mathbf{x}_{(\alpha,1)}$ karakterizálható az exponenciális centrummal azaz (3.5)-tel vagy (3.9)-cel is. Jelöljük (3.9) megoldását $\mathbf{x}_{(\varepsilon)}$-nal. Mivel $\mathbf{x}_{(\alpha,1)} = \mathbf{x}_{(\varepsilon)}$ ezért nyilván $\lim\limits_{\alpha \to 0} \mathbf{x}_{(\alpha,1)} = \lim\limits_{\varepsilon \to 0} \mathbf{x}_{(\varepsilon)} = \mathbf{x}^*$. Tudjuk továbbá, hogy minden olyan ε-ra, amelyre $\varepsilon^0 > \varepsilon > 0$ létezik egy és csak egy $\mathbf{x}_{(\varepsilon)}$, $\mathbf{z}_{(\varepsilon)}$ illetve $\eta_{(\varepsilon)}$ $\mathbf{y}_{(\varepsilon)}$, megoldás. Tehát $\exists \mathbf{x}_{(\varepsilon)} \in P_{(\varepsilon)}$ amelyre (3.11) megoldható minden $\varepsilon^0 > \varepsilon > 0$ esetén.

$$\frac{1}{\exp\left(x_{(\varepsilon)j}\right) - 1} = \widehat{z}_j\left(1 + \eta_{(\varepsilon)}\right) + \mathbf{a}_j^T \mathbf{y}_{(\varepsilon)} \qquad \forall j. \tag{3.11}$$

A 3.12. Lemma alapján tegyük fel, hogy $x_j^* = \lim\limits_{\varepsilon \to 0} x_{(\varepsilon)j} = 0$ $\forall j \in N$ és $x_j^* = \lim\limits_{\varepsilon \to 0} x_{(\varepsilon)j} > 0$, $\forall j \in B$. Így (3.11)-t az alábbi két részre bonthatjuk.

$$\frac{1}{\exp\left(x_j^*\right) - 1} = \widehat{z}_j\left(1 + \eta_{(0)}\right) + \mathbf{a}_j^T \mathbf{y}_{(0)} \qquad \forall j \in B, \tag{3.12}$$

$$x_j^* = 0 \qquad \forall j \in N. \tag{3.13}$$

Mivel $\mathbf{x}^* \in P$ elég azt belátnunk, hogy (3.12) megoldható a $\eta_{(0)}$, $\mathbf{y}_{(0)}$, ismeretlenekre. Tekintsük (3.12)-t a Kronecker tétel I. rendszerének és tegyük fel, hogy nem oldható meg. Írjuk fel Kronecker altenatíváját, amely most feltevésünk szerint megoldható tehát létezik olyan $\mathbf{r} \in \mathbf{R}^n$ vektor, amelyre

$$\sum_{\forall j \in B} r_j a_{ij} = 0 \qquad \forall i \quad (i = 1 \ldots m),$$

$$\sum_{\forall j \in B} r_j \widehat{z}_j = 0,$$

$$\sum_{\forall j \in B} r_j \frac{1}{\exp\left(x_j^*\right) - 1} \neq 0. \tag{3.14}$$

Mivel $\frac{1}{\exp\left(x_j^*\right)-1}$, $\forall j \in B$-re pozitív véges érték és a $\frac{1}{\exp\left(x_{(\varepsilon)j}\right)-1}$ folytonos függvénye ε-nak ezért létezik olyan kis $\mathbf{d} \in \mathbf{R}^n$ vektor hogy (3.14) megoldható minden olyan $\mathbf{x} \in \mathbf{R}^n$ vektorra, amelyre $\mathbf{x}^* - \mathbf{d} \leq \mathbf{x} \leq \mathbf{x}^* + \mathbf{d}$. De ez azt jelenti, hogy Kronecker alternatívája (3.12) nem oldható meg ilyen $\mathbf{x}$ vektorra tehát $\mathbf{x}^*$ nem lehet

torlódási pont, ami ellentmond feltevésünknek. Tehát ha $\mathbf{x}^*$ torlódási pont, akkor szükségképpen (3.12) megoldható a $\eta_{(0)}$, $\mathbf{y}_{(0)}$, ismeretlenekre. Ezzel a bizonyítást befejeztük. □

3.14. TÉTEL. *Tekintsük az* EP $(\alpha, 1)$ *feladat exponenciális centrum feladatát azaz (3.5)-öt. Tegyük fel, hogy az LP primál célfüggvénye nem konstans* P^0*-n. Jelölje* $\mathbf{x}_{(\varepsilon)}$ *az adott* ε*-hoz tartozó optimális megoldást és legyen* $\varepsilon^0 \geq \varepsilon > 0$*. Ekkor* $\mathbf{x}_{(\varepsilon)}$, ε *szerint folytonosan deriválható görbe.*

Bizonyítás. A bizonyítás analóg a 3.8. Tétel bizonyításával, ezért ettől eltekintünk. □

Az EP $(1, \beta)$ feladatra vonatkozóan, a 3.8.–3.14. tételekhez, lemmákhoz, következményhez és definícióhoz hasonló állításokat lehet tenni. Ezt az olvasóra bízzuk.

IRODALOM

[1] M. S. Bazarra, H. D. Sherali, C. M. Shetty, *Nonlinear Programming: Theory and Algorithms,* John Wiley & Sons, Inc. (1993).

[2] A. Charnes, W. W. Cooper and K. O. Kortanek, A Duality Theory for Convex Programmings with Convex Constraints, *Bull. American Mathematical Society,* **68** (1962), 605–660.

[3] Csiszár I., Entrópiamaximalizálás és rokon módszerek: Axiomatika és Algoritmusok, *Szigma,* **3–4.** 24. évfolyam (1993), 111—137.

[4] G. B. Dantzig, *Linear Programming and Extension,* Princeton University Press (Princeton, NJ, 1963).

[5] A. V. Fiacco, G. P. McCormick, *Nonlinear Programming,* Sequential Uncontsrained Minimization Techniques, John Wiley & Sons (Inc. New York, 1968).

[6] B. Jansen, C. Roos, T. Terlaky, J. Ph. Vial, "Primal-Dual Target Following Algorithmus for Linear Programming", Delft University of Technology Report, Nr. 93–107, 1993.

[7] B. Jansen, C. Roos, T. Terlaky, J. Ph. Vial, Duality Theory of LP: Skew Symmetric Self Dual Problems, *Optimization.*

[8] P. Kas, E. Klafszky, On the Duality of the Mixed Entropy Programming, *Optimization,* **27** (1993), 253–258.

[9] J. N. Kapur, *Maximum Entropy Models in Science and Engineering,* John Wiley & Sons (Inc., 1989).

[10] Klafszky E., „Geometriai programozás és néhány alkalmazása", kandidátusi értekezés, Budapest, 1973.

[11] Klafszky E., Terlaky T., A pivottechnika szerepe a lineáris algebra néhány alapvető tételének bizonyításában, *Alkalmazott Matematikai Lapok,* **14** (1989).

[12] C. Roos, T. Terlaky, J. Ph. Vial, *Interior Point Approach to Linear Optimization: Theory and Algorithms,* John Wiley & Sons (Inc., 1997).

[13] R. T. Rockafeller, *Convex Analysis,* Princeton University Press (Princeton, N.J., 1970).

[14] Gy. Sonnevend, "An Analytical Centre for Polyhedrons and New Classes of Global Algorithms for Linear (Smooth, Convex) Programming", in: *System Modelling and Optimization,* Ed. A. Prékopa, J. Szrelezsán, B. Strazicky, Proceedings of the 12th IFIP-Conference held in Budapest, Hungary, September (1985), volume 84 of Lecture Notes in Control and Information Sciences, pages 866–876. Springer Verlag (Berlin, West-Germany, 1986).

[15] Terlaky T., A Karmarkar típusú algoritmusokról, *Alkalmazott Matematikai Lapok,* **15** (1990–91), 133–162.

[16] A. W. Tucker, "Dual Systems of Homogeneous Linear Relations", in: *Linear Inequalities and Related Systems, Annals of Mathematical Studies,* Ed. H. W. Kuhn and A. W. Tucker, No. 38, Princeton University Press, Princeton, New Jersey (1956), pp. 3–18.

[17] Prékopa A., *Lineáris Programozás,* Bolyai János Matematikai Társulat (1968).

[18] W. Rudin, *A Matematikai Análízis Alapjai,* Műszaki Könyvkiadó (Budapest, 1978).

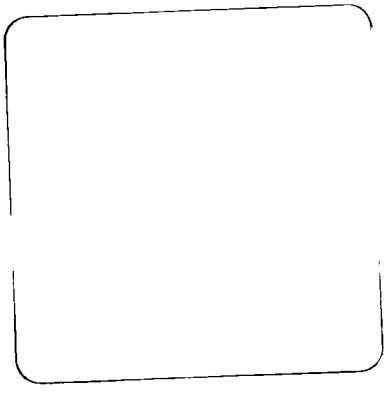
(Beérkezett: 1999. január 5.)

KLAFSZKY EMIL ÉS MÁLYUSZ LEVENTE
BUDAPESTI MŰSZAKI EGYETEM
ÉPÍTÉSZMÉRNÖKI KAR
ÉPÍTÉSKIVITELEZÉSI TANSZÉK
1111 BUDAPEST
MŰEGYETEM RKP. 3.

THE EXPONENCIAL BARRIER PROGRAMMMING, AS AN ANALITICAL APPROXIMATION OF THE LINEAR PROGRAMMING

EMIL KLAFSZKY AND LEVENTE MÁLYUSZ

We presented a special convex programming problem as an analitical approximation of the standard linear programming problem with arbitrary accuracy. The compatibility condition of linear programming problem is perturbed and the cotangent hyperbolic function is used as an equilibrium function. It leads to a convex programming problem with parameters α, β, called Exponential Barrier Programming problem $\left(\mathrm{EP}\left(alpha, \beta\right)\right)$. The following statement is proved: if $\alpha\beta \to 0$ then $\mathrm{EP}\left(\alpha, \beta\right) \to \mathrm{LP}$, moreover a simple proof is given on the duality theory of $\mathrm{EP}\left(\alpha, \beta\right)$.



A kiadásért felelős a BJMT főtitkára
Nyomta az MSZH Nyomda és Kiadó Kft., Budapest, 2000.002
Felelős vezető: Nagy László
Megjelent 9,3 (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 után, a kézirat befejezéseképpen fel kell tüntetni a szerző teljes nevét és a munkahelye (illetve lakása) pontos 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 kel 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 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-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 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* (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.

TARTALOMJEGYZÉK

INDEX