Alkalmazott matematikai lapok

1979/1-2

A MAGYAR TUDOMÁNYOS AKADÉMIA
MATEMATIKAI ÉS FIZIKAI TUDOMÁNYOK
OSZTÁLYÁNAK KÖZLEMÉNYEI

5.

KÖTET

AKADÉMIAI KIADÓ, BUDAPEST

A MAGYAR TUDOMÁNYOS AKADÉMIA
MATEMATIKAI ÉS FIZIKAI TUDOMÁNYOK OSZTÁLYÁNAK
ALKALMAZOTT MATEMATIKAI LAPJA

V. kötet 1—2. szám

Szerkesztőség: 1502 Budapest XI., Kende u. 13—17.

Kiadóhivatal: 1055 Budapest V., Alkotmány u. 21.

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

Kéziratok a következő címre küldendők:

Prékopa András, főszerkesztő
1502 Budapest XI., Kende u. 13—17.

Ugyanerre a címre küldendő minden szerkesztőségi levelezés.

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 84 forint. Belföldi megrendelések az Akadémiai Kiadó, 1055 Budapest V., Alkotmány u. 21. címen (pénzforgalmi jelzőszám 215—11 488), külföldi megrendelések a Kultúra Külkereskedelmi Vállalat, H-1389 Budapest, Pf. 149. címen (pénzforgalmi jelzőszám 218—10 990) lehetségesek.

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

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

Alkalmazott Matematikai Lapok 5 (1979) 1—48

NEM-AUTONÓM DIFFERENCIÁLEGYENLET-RENDSZEREK MEGOLDÁSAINAK STABILITÁSA ÉS PARCIÁLIS STABILITÁSA

HATVANI LÁSZLÓ

Szeged

A dolgozat a *Ljapunov-féle direkt módszerről* ad áttekintést. Bebizonyítjuk a *Ljapunov—Csetajev-féle alaptételek* parciális stabilitásra vonatkozó alakjait. Ezek alkalmazásaként a konzervatív mechanikai rendszerek egyensúlyi helyzetére vonatkozó *Lagrange—Dirichlet-tételt* és megfordíthatóságát tárgyaljuk. Ismertetjük az első közelítés (linearizálás) alapján végzett stabilitásvizsgálat alaptételeit, külön foglalkozva a konstans együtthatós és periodikus együtthatós lineáris rendszerek esetével. A direkt módszer továbbfejlesztésével foglalkozó modern kutatások eredményeiből olyan tételeket tárgyalunk, amelyek aszimptotikus stabilitást állapítanak meg szemidefinit deriválttal rendelkező, illetve felülről nem-korlátos *Ljapunov-függvény* segítségével. Foglalkozunk mechanikai rendszerek különböző típusú kiegészítő erőkkel való stabilizálásával. Alkalmazásként általában disszipatív és giroszkopikus erők hatása alatt álló mechanikai rendszerek (pl. súrlódó közegben mozgó, változó fonalhosszúságú inga) stabilitási viszonyait vizsgáljuk.

1. Bevezetés

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

$$\dot{x} = X(t, x) \quad (t \geqq 0,\ x \in R^n) \tag{1.1}$$

közönséges differenciálegyenlet-rendszer; t az időt jelöli, x pedig a rendszert jellemző állapothatározókból álló vektor. Ilyen például egy anyagi pontokból álló rendszer vagy egy merev test a klasszikus mechanikában — az állapothatározók a pontok helyét megadó koordináták és a sebességvektor komponensei — de ilyen modell írja le például különböző, egymásba átalakuló anyagok kémiai reakcióját is, ahol az állapothatározók az egymásra ható anyagok mennyisége egy adott időpillanatban. Tegyük fel, hogy a rendszer determinált a következő értelemben: az a tény, hogy a rendszer a t_0 időpillanatban az x_0 állapotban van, egyértelműen meghatározza a rendszer további állapotait, vagyis az (1.1) differenciálegyenlet-rendszernek egyetlen olyan $x(t; t_0, x_0)$ megoldása létezik, amely a t_0-ban az x_0 értéket veszi fel: $x(t_0; t_0, x_0) = x_0$. Az x_0 kezdeti állapot meghatározása méréssel történik, tehát nem abszolút pontos. Ha a valódi x_0 állapotra méréssel a ξ közelítés adódott, akkor modellünk alapján arra a következtetésre jutunk, hogy a rendszer a t időpillanatban az $x(t; t_0, \xi)$ állapotban lesz, pedig $x(t; t_0, x_0)$ lesz a valódi állapota. A modell csak akkor használható a gyakorlatban a rendszer leírására, ha ez a két állapot kicsit tér el egymástól, feltéve, hogy ξ elég jól közelíti x_0-t. Vagyis, ha megadjuk az állapothatározók meghatározásának pontosságát ($\varepsilon > 0$), akkor létezik egy olyan $\delta > 0$ tűrés az x_0 kezdeti állapot meghatározására, hogy annak betartá-

sával $x(t;\ t_0, \xi)$ az adott ε pontossággal közelíti a valódi $x(t;\ t_0, x_0)$ állapotot egy adott intervallumon.

A differenciálegyenletek elméletéből ismeretes [27], hogy ha a t egy $[t_0, T]$ korlátos intervallumon változik, akkor bármely $\varepsilon>0$-hoz van ilyen $\delta>0$ (a megoldások a kezdeti értékektől folytonosan függnek, ha (1.1) jobb oldala elég szabályos). A problémát az okozza, hogy akármilyen szabályos differenciálegyenlet esetében is előfordulhat az, hogy δ a T növelésével minden határon túl csökken, tehát a rendszert az $x(t;\ t_0, \xi)$ megoldás hamisan írja le, ha a megfigyelés hosszú ideig tart, bármilyen pontosan is adtuk meg a ξ kezdeti állapotot. Ezért fontos feladat megadni az $X(t, x)$ függvényre olyan feltételeket, amelyek biztosítják T-től független $\delta>0$ létezését. Ha ilyen δ létezik, akkor az $x(t;\ t_0, x_0)$ *megoldást stabilis*nak nevezzük, ami matematikai terminológiával azt jelenti, hogy $x(t;\ t_0, \xi) \rightrightarrows x(t;\ t_0, x_0)$ a $t\in[t_0, \infty)$ intervallumon egyenletesen, ha $\xi \to x_0$ (a különböző stabilitási fogalmak pontos definícióit l. a 2. pontban).

A megoldások stabilitásának problémájához jutunk a következő jelenség feldolgozásánál is. Tegyük fel, hogy az (1.1) által leírt rendszert a t_0 időpillanatban valami váratlan, ellenőrizhetetlen, ismeretlen szerkezetű, de kicsiny nagyságrendű zavaró hatás, perturbáció éri, ami egy rövid, h hosszúságú időintervallumon hat. Tehát a rendszert a t_0-ig, illetve t_0+h-tól ismét (1.1) írja le, de a $[t_0, t_0+h]$ intervallumon nem. Tegyük fel, hogy rendszerünk a t_0 időpillanatban az x_0 állapotban volt. Ha a perturbáció nem hatott volna, a rendszer jövőjét az $x(t;\ t_0, x_0)$ megoldás adná, de a perturbáció miatt a $[t_0, t_0+h]$ rövid időszakaszon a rendszer „átugrott" egy másik $x(t;\ t_0, \xi)$ „pályára". Elvárjuk, hogy ha a zavaró hatás elég kicsiny, akkor $x(t;\ t_0, \xi)$ tetszőleges pontossággal közelítse $x(t;\ t_0, x_0)$-t, vagyis ismét a stabilitás problémájához jutottunk.

A stabilitás problémája már igen régen felmerült mechanikai rendszerek egyensúlyi helyzetének jellemzésével kapcsolatban (a legegyszerűbb példaként az inga alsó és felső egyensúlyi helyzetének összehasonlítása szolgálhat). Az első monográfiák E. G. ROUTH-tól származnak az 1877—1884-es évekből. Nagy lökést adott az elmélet kifejlődéséhez a *Watt-féle centrifugális regulátorral* kapcsolatban felmerült kérdések megoldása (l. például [27]). Az elmélet igazi matematikai megalapozását H. POINCARÉ és A. M. LJAPUNOV kezdte el. A híres orosz tudós, A. M. LJAPUNOV 1892-ben közölte doktori disszertációját [42], amely mérföldkövet jelent az elmélet történetében. Ebben megadta a stabilitási fogalmak egzakt definícióját, és adott egy olyan módszert a problémák megoldására, amely ma is a legalapvetőbb és legsikeresebb az elméletnek még olyan ágaiban is, amelyek azóta születtek (pl. automatikus irányítások, optimális folyamatok stabilitásának elmélete [10, 34]. A *Szovjetunióban* az 1930-as évektől, nyugaton az 1950-es évektől kezdve napjainkig a stabilitáselméletben nagy fellendülés tapasztalható, amely N. G. CSETAJEV, I. G. MALKIN, illetve S. LEFSCHETZ, J. P. LASALLE és R. BELLMAN munkásságával kezdődött. Ez a fellendülés köszönhető a már említett új tudományok kialakulásának, de főleg a számítógépek megjelenésének. Az eddigiekből kitűnik ugyanis, hogy egy differenciálegyenlet numerikus megoldását is stabilitásvizsgálatnak kell megelőzni, hiszen csak akkor várhatunk megfelelő közelítést, ha a kiszemelt megoldás stabilis. Ennek az időszaknak egy fontos eredménye a parciális stabilitás fogalmának bevezetése és vizsgálata, amely V. V. RUMJANCEV [49, 51] szovjet tudós nevéhez fűződik. Az $x(t;\ t_0, x_0)$ megoldás parciális stabilitása azt jelenti, hogy az $x(t;\ t_0, \xi)$ megoldás *bizonyos komponensei* kicsit térnek el az $x(t;\ t_0, x_0)$ megoldáséitól, ha ξ az x_0-hoz közel van.

A stabilitáselméletben külön foglalkoznak az autonóm és nem-autonóm differenciálegyenlet-rendszerek megoldásaival. Az (1.1) rendszert *autonómnak* nevezzük, ha a jobb oldalán álló X függvény nem függ az időtől. Az elnevezés onnan ered, hogy az önálló, más rendszerek által nem befolyásolt jelenségek fejlődéstörvénye általában független az időtől. A két eset különválasztásának az az oka, hogy autonóm rendszerekre nagyon sok olyan ismerettel rendelkezünk, amelyek a stabilitáselméletben fontosak (l. dinamikus rendszerek elmélete [48]), de nem-autonóm esetben sokszor még megfelelőjük sincs. Ennek az az oka, hogy az idő a stabilitáselméletben mindig egy nem-kompakt halmazon változik, ami lényegesen megkülönbözteti a többi állapothatározótól.

Jelen dolgozat áttekintést ad nem-autonóm differenciálegyenletek stabilitási, parciális stabilitási tulajdonságainak *Ljapunov második módszerével* való vizsgálatáról. Célunk az, hogy a klasszikus eredményekből kiindulva megismertessük az olvasót az elmélet alapvető tételeivel, rámutassunk az elmélet ma is művelt, aktuális problémáira, főleg olyanokra, amelyeken keresztül — a szerző megítélése szerint — sikeresen be lehet kapcsolódni a modern eredmények alkotó alkalmazásába, és így a modern stabilitáselméleti kutatásokba.

Meg kell jegyeznünk, hogy a nem-autonóm rendszerek kvalitatív vizsgálatának egy nagyon fontos eszközével, a közepelés módszerével [15, 35] a dolgozat terjedelmére vonatkozó megkötés miatt nem foglalkozunk. Ugyanezen okból nem szerepel a stabilitáselmélet néhány fontos és érdekes területe, mint például az állandóan ható perturbációk melletti stabilitás, ún. totális stabilitás elmélete [14, 21, 39, 43].

A tételekre folyamatosan alkalmazásokat mutatunk. Hogy megkönnyítsük az egyes tételek összevetését, alkalmazásaink legtöbbször ugyanarra a konkrét, vagy egy bizonyos típusú mechanikai rendszerre vonatkoznak. További, más tudományágakkal kapcsolatos alkalmazásokat az olvasó többek között a [10, 28, 46] monográfiákban találhat.

2. Jelölések és alapvető definíciók

Az $x=(x_1, x_2, \ldots, x_n)$ valós szám$-n$-esek euklideszi terét R^n-nel jelöljük. Megállapodunk abban, hogy $x\in R^n$ egyaránt jelölhet sor- vagy oszlopvektort; erre vonatkozó egyedüli megszorítás az, hogy a felírt műveleteknek legyen értelmük. A valós számok R^1 terét egyszerűen R-rel jelöljük, továbbá legyen R_+ a nemnegatív, R_- a nem-pozitív valós számok halmaza. Egy $x\in R^n$ elem normáját $|x|$-kel jelöljük: $|x|:=(x_1^2+x_2^2+\ldots+x_n^2)^{1/2}$.

Ha $a\in R$, legyen $[a]_+:=\max\{a, 0\}$, $[a]_-:=\max\{-a, 0\}$; $x\in R^n$ esetén $[x]_+:=$ $:=([x_1]_+, [x_2]_+, \ldots, [x_n]_+)$; $[x]_-$ pedig hasonlóan van definiálva.

Ha $x, y\in R^n$, akkor $(x, y):=x_1y_1+x_2y_2+\ldots+x_ny_n$ a két vektor *skaláris szorzatát* jelöli. Ha $H\subset R^n$, akkor H^c a H *halmaz komplementere*. Az R^n tér két részhalmazának $(H, K\subset R^n)$ távolságát a

$$\varrho(H, K):=\inf\{|x-y|:\ x\in H, y\in K\}$$

képlettel értelmezzük. Használni fogjuk még a

$$G^n(a, \varepsilon):=\{x\in R^n:\ |x-a|<\varepsilon\},\quad S(H, \varepsilon):=\{x\in R^n:\ \varrho(x, H)<\varepsilon\}$$

jelöléseket $(a\in R^n, H\subset R^n, \varepsilon>0)$.

Gyakran szükségünk lesz az $x\in R^n$ vektor egy $x=(y, z)$ *particionálás*ára, ahol $y\in R^p$, $z\in R^q$ $(0<p\leqq n,\ 0\leqq q,\ p+q=n)$, ezért az x, y, z, n, p, q betűket következetesen ebben az összefüggésben használjuk, valahányszor általános differenciálegyenlet-rendszert tárgyalunk.

Legyen $U: R^p\times R^q\to R$ az első p változója szerint differenciálható függvény. Ekkor a $(\partial U(y, z)/\partial y_1, \ldots, \partial U(y, z)/\partial y_p)$ vektort röviden $\partial U(y, z)/\partial y$-nal is fogjuk jelölni.

Jelölje $\mathscr{K}$ az $a: R_+\to R_+$ folytonos, szigorúan monoton növekvő, $a(0)=0$ tulajdonságú *függvények osztályát*.

A továbbiakban Γ az R^n tér egy olyan tartományát jelöli, amely az $x=0$ pontot tartalmazza. Egy $V: R_+\times\Gamma\to R$ függvényt *y-ban pozitív definitnek* nevezünk, ha $V(t, 0)\equiv 0$, és van az $y=0$ pontnak olyan $U_1\subset R^p$ környezete és olyan $W: U_1\to R$ függvény, hogy $W(y)>0$, ha $y\neq 0$ és $V(t, x)\geqq W(y)$ $(x=(y, z)\in\Gamma$, $y\in U_1)$. Könnyű megmutatni [49], hogy ez ekvivalens a következővel: $V(t, 0)\equiv 0$, és van az $y=0$ pontnak olyan $U_1\subset R^p$ környezete és olyan $a\in\mathscr{K}$, hogy $V(t, x)\geqq$ $\geqq a(|y|)$ $(x\in\Gamma, y\in U_1)$. Jól ismert (pl. l. [3, 53]), hogy egy (y, Ay) kvadratikus alak $(A: p\times p$ típusú konstans mátrix), akkor és csakis akkor pozitív definit, ha determinánsának minden diagonális főminora pozitív. A V függvényt *y-ban negatív definit*nek nevezzük, ha $-V$ y-ban pozitív definit.

Tekintsük most az

$$\dot{x} = X(t, x) \tag{2.1}$$

differenciálegyenlet-rendszert, ahol $X: R_+\times\Omega\to R^n$ folytonos függvény, $\Omega\subset R^n$ tartomány, továbbá bármely $t_0\in R_+$, $x_0\in\Omega$ párra lokálisan létezik a (2.1) egyenlet egyetlen olyan $x(t; t_0, x_0)$ megoldása, amely kielégíti az $x(t_0; t_0, x_0)=x_0$ kezdeti feltételt. Tegyük fel, hogy bármely $t_0\in R_+$-hoz van olyan $\varrho(t_0)>0$, hogy ha $|x_0-\xi|<\varrho(t_0)$, akkor $x(t; t_0, \xi)$ értelmezve van a $[t_0, \infty)$ intervallumon. A különböző stabilitási fogalmak azt kívánják meg a megoldásoktól, hogy az $|x(t; t_0, \xi)-$ $-x(t; t_0, x_0)|$ eltérés valamilyen értelemben kicsiny legyen, ha $|\xi-x_0|$ elég kicsi. Hogy a definíciók a lehető legegyszerűbbek legyenek, hajtsuk végre az $u:=x-$ $-x(t; t_0, x_0)$ transzformációt. A (2.1) rendszer az új változóval felírva

$$\dot{u} = X(t, u+x(t; t_0, x_0))-X(t, x(t; t_0, x_0)) =: U(t, u)$$

alakú, ahol $U: R_+\times\Gamma\to R^n$ folytonos, $0\in\Gamma$, $\Gamma\subset R^n$ tartomány, az $x=x(t; t_0, x_0)$ megoldás az $u=0$ megoldásnak felel meg (tehát $U(t, 0)\equiv 0$), és a vizsgálandó eltérés $|u(t; t_0, u_0)|$. Visszatérve a szokásos jelölésre, a

$$\dot{x} = X(t, x) \quad (X(t, 0)\equiv 0) \tag{2.2}$$

differenciálegyenlet-rendszert vizsgáljuk, ahol $X: R_+\times\Gamma\to R^n$ folytonos, és bármely t_0-hoz van olyan $\varrho(t_0)>0$, hogy ha $|x_0|<\varrho(t_0)$, akkor létezik a (2.2) egyetlen $x(t; t_0, x_0)$ megoldása a $[t_0, \infty)$ intervallumon.

Tekintsük ismét az $x\in R^n$ vektornak a már bevezetett $x=(y, z)$ particionálását. A (2.2) 0-megoldását

a) *y-stabilisnak* nevezzük, ha bármely $\varepsilon>0$, $t_0\in R_+$-hoz létezik olyan $\delta(\varepsilon, t_0)>0$, hogy ha $|x_0|<\delta$, akkor $|y(t; t_0, x_0)|<\varepsilon$ a $[t_0, \infty)$ intervallumon;

b) *egyenletesen y-stabilisnak* nevezzük, ha az a) definícióban $\delta(\varepsilon, t_0)$ csak ε-tól függ, t_0-tól nem;

c) *y-attraktívnak* nevezzük, ha bármely $t_0 \in R_+$-hoz létezik olyan $\sigma(t_0)>0$, hogy ha $|x_0|<\sigma$, akkor $|y(t; t_0, x_0)| \to 0$, ha $t \to \infty$;

d) *aszimptotikusan y-stabilisnak* nevezzük, ha y-stabilis és y-attraktív;

e) *egyenletesen aszimptotikusan y-stabilisnak* nevezzük, ha egyenletesen y-stabilis, és van olyan $\sigma>0$, hogy bármely $\eta>0$-hoz létezik olyan $T(\eta)$, hogy ha $|x_0|<\sigma$, $t_0 \in R_+$, akkor $|y(t; t_0, x_0)|<\eta$ a $[t_0+T(\eta), \infty)$ intervallumon.

Megállapodunk abban, hogy ha a 0-megoldást egyszerűen stabilisnak, attraktívnak stb. mondjuk, akkor az x-stabilitást, x-attraktivitást stb. jelent, amelyek egybeesnek a LJAPUNOV által [42]-ben bevezetett stabilitási fogalmakkal.

A dolgozatban központi szerepet játszanak a $V: R_+ \times \Gamma \to R$ differenciálható függvények, amelyeket *Ljapunov-függvényeknek* fogunk nevezni. *A V Ljapunov-függvényeknek a* (2.2) *rendszerre vonatkozó deriváltján* a

$$\dot{V}(t, x) = \sum_{i=1}^{n} \frac{\partial V(t, x)}{\partial x_i} X_i(t, x) + \frac{\partial V(t, x)}{\partial t} = \left(\frac{\partial V(t, x)}{\partial x}, X(t, x) \right) + \frac{\partial V(t, x)}{\partial t}$$

függvényt értjük. (Ha több rendszer is szerepel a tárgyalásban, akkor megkülönböztetésül a rendszer azonosítóját is kiírjuk a derivált jelölésében: $\dot{V}_{(2.2)}(t, x)$). Ezt a jelölést az alábbi tény indokolja:

2.1. LEMMA. Legyen V Ljapunov-függvény, és $x(t)$ (2.2)-nek tetszőleges megoldása. Ekkor

$$\frac{d}{dt} V(t, x(t)) = \dot{V}(t, x(t))$$

az $x(t)$ megoldás teljes létezési intervallumán.

Bizonyítás. A láncszabály szerint

$$\frac{d}{dt} V(t, x(t)) = \left[\frac{\partial V(t, x)}{\partial t} + \left(\frac{\partial V(t, x)}{\partial x}, \dot{x}(t) \right) \right]_{x=x(t)} = \dot{V}(t, x(t)),$$

hiszen $\dot{x}(t) = X(t, x(t))$.

A parciális stabilitás tanulmányozásához a (2.2) rendszert néha hasznos az

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

alakban írni, ahol $Y: R_+ \times \Gamma \to R^p$, $Z: R_+ \times \Gamma \to R^q$.

3. A Ljapunov-féle direkt módszer alaptételei

Mint ahogyan a bevezetésben már említettük, a stabilitáselmélet leghatékonyabb módszerének, az úgynevezett direkt módszernek az alapjait A. M. LJAPUNOV orosz tudós fektette le 1892-ben megjelent híres doktori disszertációjában. Ezek a tételek ma is a stabilitási vizsgálatok alapvető és általánosan használt eszközei. Méltán szokás a direkt módszer alaptételei közé sorolni még N. G. CSETAJEV instabilitási tételét, amelyet konzervatív mechanikai rendszerek stabilitására vonatkozó, az 1930—40-es években folytatott vizsgálatai eredményeként kapott [53]. A módszer

alkalmas módosításokkal a parciális stabilitás tanulmányozására is használható, amit V. V. RUMJANCEV bizonyított be 1957-ben [49, 51]. Itt mi az alaptételek általa kimondott változatait közöljük, mivel a klasszikus tételek ezek nyilvánvaló speciális esetei, amelyek az $x=x$ triviális particionálással adódnak.

Legyen adva az

$$\dot{x} = X(t, x) \quad (X\colon R_+ \times \Gamma \to R^n;\ X(t, 0) \equiv 0) \tag{3.1}$$

differenciálegyenlet-rendszer, amelyet az $x=(y, z)$ particionálással a 2. pontban bevezetett jelölésekkel az

$$\dot{y} = Y(t, y, z), \quad \dot{z} = Z(t, y, z) \tag{3.2}$$

alakban is írunk.

3.1. TÉTEL. Ha a (3.2) rendszerhez létezik olyan $V\colon R_+ \times U \to R$ y-ban pozitív definit függvény, amelyre $\dot{V} \leqq 0$, akkor (3.2) 0-megoldása y-stabilis.

Bizonyítás. Mivel V y-ban pozitív definit, létezik olyan $a \in \mathscr{K}$ függvény, hogy $V(t, x) \geqq a(|y|)$ $(t \in R_+, y \in U_1)$. Legyen ε és t_0 adott. V folytonossága és $V(t_0, 0)=0$ miatt létezik olyan $\delta(\varepsilon, t_0)>0$, hogy ha $|x|<\delta$, akkor $V(t_0, x)<a(\varepsilon)$. Legyen $|x_0|<\delta$, és tekintsük az $x(t)=x(t;\ t_0, x_0)$ megoldást. A 2.1. lemma szerint

$$a(|y(t)|) \leqq V(t, x(t)) \leqq V(t_0, x_0) < a(\varepsilon),$$

vagyis $|y(t)|<\varepsilon$, ha $t \geqq t_0$. Ez azt jelenti, hogy a 0-megoldás y-stabilis.

3.1. Megjegyzés. Ha a tétel feltételeinek teljesülése mellett létezik olyan $b \in \mathscr{K}$ függvény, és 0-nak olyan $U \subset R^n$ környezete, hogy $V(t, x) \leqq b(|x|)$ $(t \in R_+, x \in U)$, akkor a 0-megoldás egyenletesen y-stabilis. Ugyanis ebben az esetben a bizonyításban szereplő δ választható $b^{-1}(a(\varepsilon))$-nak, ami nem függ t_0-tól, csak ε-tól.

3.2. TÉTEL. Ha a (3.2) rendszerhez létezik olyan $V\colon R_+ \times U \to R$ y-ban pozitív definit függvény, továbbá $b, c \in \mathscr{K}$ úgy, hogy

$$V(t, x) \leqq b(|u|), \quad \dot{V}(t, x) \leqq -c(|u|) \quad (|u| < \gamma), \tag{3.3}$$

ahol $u=(x_1, \ldots, x_p, x_{p+1}, \ldots, x_{p+k}) \in R^{p+k}$, $0 \leqq k \leqq q$, $\gamma>0$, akkor (3.2) 0-megoldása aszimptotikusan y-stabilis.

Bizonyítás. Előző tételünk szerint a 0-megoldás y-stabilis. Legyen $\varepsilon_0>0$ olyan, hogy ha $|x|<\varepsilon_0$, akkor $x \in U$, és tekintsük a $\sigma(t_0) := \delta(\varepsilon_0, t_0)>0$ számot, ahol $\delta(\varepsilon, t_0)$ az y-stabilitás definíciójában ε, t_0-hoz tartozó, az előző bizonyításban megadott szám. Be fogjuk bizonyítani, hogy ha $|x_0|<\sigma(t_0)$, akkor

$$v(t) := V(t;\ x(t;\ t_0, x_0)) \to 0,$$

amiből V y-ban pozitív definitsége miatt $y(t;\ t_0, x_0) \to 0$, ha $t \to \infty$.

Tegyük fel, hogy $\lim_{t\to\infty} v(t)=v^*>0$. Ekkor (3.3) miatt $|u(t;\ t_0, x_0)| \geqq b^{-1}(v^*)$ és

$$\dot{v}(t) \leqq -c(|u(t;\ t_0, x_0)|) \leqq -c(b^{-1}(v^*)) < 0,$$

tehát $v(t) \to -\infty$, ha $t \to \infty$, ami lehetetlen. Tehát $v^*=0$.

3.2. Megjegyzés. A 3.2. tétel feltételeiből következik a 0-megoldás *egyenletes* aszimptotikus stabilitása is. Valóban, legyen ugyanis $\sigma := b^{-1}(a(\varepsilon_0))$. Megmutatjuk, hogy bármely $\eta > 0$-hoz létezik olyan $T(\eta)$ szám, hogy ha $|x_0| < \sigma$, akkor $|y(t; t_0, x_0)| < \eta$ a $[t_0 + T, \infty)$ intervallumon. Ehhez nyilván elegendő megmutatni, hogy $|y(t_0 + T; t_0, x_0)| < \delta(\eta) = b^{-1}(a(\eta))$. A 3.2. tétel bizonyításának végén található okoskodás mutatja, hogy a $T(\eta) = a(\varepsilon_0)/c(b^{-1}(a(\eta)))$ választás megfelelő.

3.3. TÉTEL. Tegyük fel, hogy valamely $\varepsilon_0 > 0$, $t_0 \in R_+$ számokhoz létezik olyan $V: D = R_+ \times G^p(0, \varepsilon_0) \times R^q \to R$ függvény, amely rendelkezik a következő tulajdonságokkal:

(1) V korlátos a $D_+ := \{(t, x) \in D: V(t, x) > 0\}$ halmazon;
(2) a $(t_0, 0)$ torlódási pontja D_+-nak;
(3) létezik olyan $c \in \mathcal{K}$, hogy

$$\dot{V}(t, x) \geqq c(V(t, x)) \quad ((t, x) \in D_+).$$

Ekkor (3.2) 0-megoldása y-instabilis.

Bizonyítás. A (2) feltétel miatt bármely $\delta > 0$-hoz létezik olyan x_0, hogy $|x_0| < \delta$, $(t_0, x_0) \in D_+$. Tekintsük az $x(t) = x(t; t_0, x_0)$ megoldást. Ha a $t \in [t_0, T]$ intervallumon $(t, x(t)) \in D_+$, akkor a 2.1. lemma szerint

$$v(t) \geqq v(t_0) + c(V(t_0, x_0))(T - t_0) \to \infty \quad (T \to \infty),$$

tehát (1) miatt van olyan T, hogy $|y(T; t_0, x_0)| = \varepsilon_0$, vagyis a 0-megoldás y-instabilis.

3.3. Megjegyzés. Nem nehéz bebizonyítani, hogy ha a (3.1) rendszer autonóm és V nem függ t-től, akkor a tétel (3) feltétele helyettesíthető a következővel:

$$\dot{V}(x) > 0 \quad ((t, x) \in D_+).$$

Az alaptételek közvetlen alkalmazhatóságának illusztrálására vizsgáljuk olyan mechanikai rendszerek egyensúlyi helyzetének stabilitását, amelyekben a kényszerfeltételek az időtől függetlenek és a helykoordináták segítségével kifejezhetők (szkleronom, holonom rendszerek, l. [4]). Az általánosított koordinátákat jelölje $q = (q_1, q_2, \ldots, q_n) \in R^n$, a kinetikai energiát $T(q, \dot{q}) = (\dot{q}, A(q)\dot{q})/2$, ahol A szimmetrikus, pozitív definit mátrix. Tegyük fel először, hogy a rendszerre csak konzervatív erők hatnak; a rendszer potenciális energiáját jelölje $P(q)$. A mozgást a

$$\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{q}_i}\right) - \frac{\partial T}{\partial q_i} = -\frac{\partial P}{\partial q_i} \quad (i = 1, 2, \ldots, n).$$

Lagrange-féle másodfajú mozgásegyenletek írják le. Ilyen rendszerek egyensúlyi helyzetének stabilitásáról szól a klasszikus *Lagrange—Dirichlet-tétel.* A tételt LAGRANGE mondta ki 1788-ban, de csak arra az esetre tudta bizonyítani, amikor P kvadratikus alak. Az általános esetre DIRICHLET adott egy rendkívül elegáns bizonyítást, amely A. M. LJAPUNOV direkt módszerének is kiindulópontja volt.

3.4. TÉTEL. Konzervatív szkleronom mechanikai rendszer egyensúlyi helyzete stabilis, ha ott a potenciális energiának szigorú helyi minimuma van.

Bizonyítás. Mindenek előtt térjünk át a $p_i := \partial T/\partial \dot{q}_i$ képletekkel a q, p *Hamilton-féle változókra;* a teljes energiát jelölje $H(q, p)$ $(:= (A^{-1}(q)p, p)/2 + P(q))$. A mozgásegyenletek az új változókkal a

$$\dot{q}_i = \frac{\partial H}{\partial p_i}; \quad \dot{p}_i = -\frac{\partial H}{\partial q_i} \quad (i = 1, 2, \ldots, n) \tag{3.4}$$

ún. *Hamilton-féle kanonikus alakba* írhatók, amelyre az alaptételek már közvetlenül alkalmazhatók.

Az általánosság megszorítása nélkül feltehetjük, hogy az egyensúlyi helyzet a $q=0$ pont, és $P(0)=0$. Ekkor $H(q, p)$ pozitív definit. Másrészt

$$\dot{H}(q, p) = \sum_{i=1}^{n} \left[\frac{\partial H}{\partial q_i} \frac{\partial H}{\partial p_i} + \frac{\partial H}{\partial p_i} \left(-\frac{\partial H}{\partial q_i} \right) \right] \equiv 0,$$

ami azt jelenti, hogy tetszőleges megoldás mentén H konstans értéket vesz fel (a mechanikai energia megmaradásának tétele), azaz H első integrálja (3.4)-nek. A 3.1 tételt alkalmazva kapjuk, hogy a (3.4) rendszer $q=p=0$ megoldása stabilis, ami ugyanazt jelenti, hogy az egyensúlyi helyzet stabilis.

Már a tétel megszületésekor felmerült a kérdés, hogy a tétel megfordítható-e. Annak ellenére, hogy nagyon sokan foglalkoztak vele, az általános esetre vonatkozó kérdés ma is megválaszolatlan. A probléma történetéből (részletesen l. [28, Chapter III]) idézzünk fel néhány eredményt.

A. WINTNER vette észre 1941-ben azt — a tétel megfordíthatósága szempontjából lényeges — tényt, hogy a potenciális energiára vonatkozó feltétel helyett elegendő megkövetelni a következőt: A $q=0$ pont bármely kicsiny környezetében van olyan, a 0-t tartalmazó nyitott halmaz, amelynek határán P pozitív értékeket vesz fel. Könnyű belátni, hogy az így kapott tétel $n=1$ esetben már megfordítható; az energiamegmaradás tételén alapuló bizonyítást az olvasóra bízzuk. Sajnos, a módszer $n \geqq 2$-re már csak nagyon speciális esetben vihető át. Most ismertetjük N. G. CSETAJEV 1952-ből származó tételét, amely a kérdéskör egyik alapvető eredménye.

3.5. TÉTEL. Ha valamely $\varepsilon > 0$ számra

(1) a $\Theta := \{q: |q| < \varepsilon, P(q) < 0\}$ halmaz nem üres;
(2) a $q=0$ torlódási pontja Θ-nak;
(3) $(\operatorname{grad} P(q), q) < 0$ $(q \in \Theta)$, akkor a $q=p=0$ egyensúlyi helyzet instabilis.

Bizonyítás. Konstruálunk egy D halmazt és egy függvényt, amelyek teljesítik a 3.3. tétel feltételeit. Legyen

$$\Delta := \{(q, p): q \in \Theta, \ |p| < \varepsilon, \ H(q, p) < 0, \ (q, p) > 0\},$$

és $V(q, p) := -(q, p)H(q, p)$. Könnyen látható, hogy $(0, 0)$ torlódási pontja Δ-nak, V pozitív, korlátos Δ-n, továbbá

$$\dot{V}(q, p) = -[-(q, \operatorname{grad}_q H) + (\operatorname{grad}_p H, p)]H =$$

$$= -[2T - (\operatorname{grad}_q T, q) - (\operatorname{grad} P, q)]H.$$

A (3) feltétel és T pozitív definitása miatt ε választható olyan kicsinyre, hogy $\dot{V}(q,p)>0$ teljesüljön a Δ halmazon, ami azt jelenti, hogy a 3.3. tétel feltételei teljesülnek (l. 3.3. megjegyzés).

3.1. KÖVETKEZMÉNY. Ha teljesül a 3.5. tétel (1), (2) feltétele, továbbá P analitikus a $q=0$ egy környezetében:

$P(q)=\sum_{i=2}^{\infty} P_i(q)$ (P_i: i-edfokú homogén forma), és valamely $k\geqq 2$-re $P_i(q)\geqq 0$, ha $i<k$, illetve $P_i(q)\leqq 0$, ha $i>k$, akkor a $q=p=0$ egyensúlyi helyzet instabilis.

Bizonyítás. A

$$(\operatorname{grad} P, q) = \sum_{i=2}^{\infty} (\operatorname{grad} P_i, q) = kP - \sum_{i=2}^{k-1} (k-i) P_i + \sum_{i=k+1}^{\infty} (i-k) P_i$$

formula mutatja, hogy a 3.5. tétel feltételei teljesülnek.

Az elméleti mechanikusok azt sejtik, hogy a *Wintner-féle feltétel analitikus* P esetében $n\geqq 2$-re szükséges. A sejtést kettő szabadsági fokú mechanikai rendszerre ($n=2$) V. P. PALAMODOV [50] bizonyította be 1977-ben a funkcionálanalízis módszereivel. Az $n>2$ esetekre a probléma ma is nyitott.

Közelítsünk most a kérdéshez a másik oldalról: milyen stabilitási tulajdonságokat tudunk mondani a $q=p=0$ egyensúlyi helyzetről a $P(q)\geqq 0$ esetben? Induljunk ki az eddiginél valamivel általánosabb rendszerből, nevezetesen tegyük fel, hogy a rendszerre nem csak konzervatív erők, hanem disszipatív erők (súrlódás, közegellenállás) és giroszkopikus erők is hatnak. Vagyis, tekintsük a

$$\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{q}_i}\right) - \frac{\partial T}{\partial q_i} = -\frac{\partial P}{\partial q_i} + \sum_{j=1}^{n} g_{ij}\dot{q}_j - \frac{\partial R}{\partial \dot{q}_i} \tag{3.5}$$

rendszert [28, 45—47], ahol $R=R(t,q,\dot{q})=(\dot{q}, B(t,q)\dot{q})/2\geqq 0$ a disszipatív erő megadására szolgáló *Rayleigh-féle függvény;* $B(t,q)$ szimmetrikus, $G(t,q)=[g_{ij}(t,q)]$ ferdén szimmetrikus ($g_{ij}=-g_{ji}$), $n\times n$-es mátrix (ezek elemei a „súrlódási", illetve giroszkopikus együtthatók, amelyek a helynek és időnek folytonos függvényei). Könnyen kiszámítható, hogy a $H(q,p)$ teljes energiának a (3.5) rendszer szerinti deriváltja: $\dot{H}(t,q,p)=-2R\leqq 0$. A 3.1. és 3.2. tétel alkalmazásával adódik a

3.6. TÉTEL. a) Ha létezik olyan k ($0\leqq k\leqq n$) és $a\in\mathcal{K}$, hogy

$$P(q) \geqq a\left(\sum_{i=1}^{k} q_i^2\right), \quad (P(0)=0)$$

a $q=0$ egy környezetében, akkor a $q=\dot{q}=0$ egyensúlyi helyzet ($q_1, q_2, \ldots, q_k, \dot{q}_1, \dot{q}_2, \ldots, \dot{q}_n$)-stabilis.

b) Ha a disszipáció teljes, vagyis van olyan $\beta>0$, hogy $R(t,q,\dot{q})\geqq\beta|\dot{q}|^2$, és potenciális erő a rendszerre nem hat ($P(q)\equiv 0$), akkor a $q=\dot{q}=0$ egyensúlyi helyzet aszimptotikusan $\dot{q}$-stabilis.

Végezetül fel kell hívnunk a figyelmet a *Ljapunov-féle direkt módszer* egy nagy előnyére, amely a gyakorlati alkalmazások szempontjából rendkívül lényeges. A tételekben megadott tulajdonságú V függvény birtokában nem csak a stabilitás,

illetve az aszimptotikus stabilitás tényét tudjuk megállapítani. A gyakorlatban sokszor ezen tulajdonságok megléte még nem elegendő a biztonságos működtetéshez, mert azt is kell tudni, hogy adott $\varepsilon>0$ tűréshez és $t_0 \geqq 0$ időpillanathoz a kezdeti eltérésnek *mekkora* $\delta(\varepsilon, t_0)>0$ eltérést engedélyezhetünk, illetve adott t_0 kezdeti időpillanathoz *mekkora* az a $\sigma(t_0)>0$ szám, amelyet a kezdeti perturbáció nem haladhat meg ahhoz, hogy a kívánt állapottól való eltérés az idő növekedésével „eltűnjön". A 3.1., illetve 3.2. tétel bizonyításából látható, hogy ha

$$\max\{V(t_0, x): |x| \leqq \lambda\} < a(\varepsilon),$$

akkor a $\delta(\varepsilon, t_0) := \lambda$, illetve a $\sigma(t_0) := \delta(\varepsilon_0, t_0)$ választás megfelelő, és ezek a számok a V függvény birtokában, *a* (3.1) *megoldásainak ismerete nélkül* közvetlenül meghatározhatók.

4. Lineáris rendszerek stabilitása. Stabilitásvizsgálat az első közelítés alapján

Tekintsük az

$$\dot{x} = X(t, x) \quad (X: R_+ \times \Gamma \to R^n) \tag{4.1}$$

rendszert, ahol X folytonosan differenciálható függvény, és legyen $x=\varphi(t)$ ennek egy adott megoldása, amelynek stabilitási tulajdonságait vizsgáljuk. A stabilitásvizsgálatnak kialakulása óta legalapvetőbb módszere a következő. Arra vagyunk kíváncsiak, hogy a megoldások mennyire térnek el a kiszemelt $y=\varphi(t)$ megoldástól, ezért vezessük be az $u=x-\varphi(t)$ változót. Mivel X differenciálható, a (4.1) egyenlet a

$$\dot{u} = A(t)u + R(t, u) \tag{4.2}$$

alakba írható, ahol $A(t)$ $n\times n$-es folytonos mátrixfüggvény, amelynek elemei $a_{ij}(t) := [\partial X_i(t, x)/\partial x_j]_{x=\varphi(t)}$, továbbá $R(t, 0) \equiv 0$, $[\partial R_i(t, u)/\partial u_j]_{u=0} \equiv 0$ $(i, j = 1, 2, \ldots, n)$. A stabilitásvizsgálatnál a φ-hez közeli megoldásokat vizsgáljuk, vagyis az $u=0$ egy kis környezetét. De $R(t, u)/|u| \to 0$ $(u \to 0)$, ezért természetesnek látszik, hogy (4.2)-ből R „elhanyagolható" és elegendő a (4.1) „linearizálásával" kapott

$$\dot{v} = A(t)v \tag{4.3}$$

egyenlet 0-megoldásának stabilitását vizsgálni; éspedig ha (4.3) 0-megoldása stabilis, akkor az $x=\varphi(t)$ megoldás is stabilis. A. M. LJAPUNOV mutatott rá, hogy ez nem mindig van így. (Ezt az olvasó is könnyen ellenőrizheti a következő egyszerű példán. Tekintsük az

$$\dot{x}_1 = -x_2 + ax_1^3, \quad \dot{x}_2 = x_1 + ax_2^3 \quad (a = \text{konstans})$$

rendszert. A $V(x_1, x_2) = x_1^2 + x_2^2$ *Ljapunov-függvény* segítségével ellenőrizhető, hogy az $x_1 = x_2 = 0$ megoldás $a>0$ esetben instabilis, $a<0$ esetben aszimptotikusan stabilis. Ugyanakkor a linearizált egyenlet 0-megoldása stabilis.) Így felmerül a következő kérdés: az $A(t)$ mátrixfüggvény milyen tulajdonságai biztosítják a fenti következtetés helyességét? Már maga LJAPUNOV [42] kidolgozott egy elméletet a mód-

szer egzakt megalapozására, de a témakör ma is a stabilitáselmélet egyik központi kutatási területe.

Ebben a fejezetben először (4.3) 0-megoldásának stabilitási tulajdonságait vizsgáljuk, különös tekintettel az állandó, illetve periodikus együtthatók esetére. Ezután ismertetünk néhány, az első közelítés (linearizálás) alapján végzett stabilitásvizsgálat megalapozását szolgáló klasszikus eredményt. Alkalmazásként a parametrikus rezonancia jelenségét tárgyaljuk a változó hosszúságú ingára. Nevezetesen azt vizsgáljuk, hogy az inga hosszának mely frekvenciával való változtatása teszi az alsó egyensúlyi helyzetet instabilissá. Bebizonyítjuk továbbá, hogy ha egy inga felfüggesztési pontja elég nagy frekvenciával függőlegesen rezeg, akkor az inga felső egyensúlyi helyzete stabilissá válik.

Visszatérve a szokásos jelölésre, vizsgáljuk a

$$\dot{x} = A(t)x \quad (t \in R_+,\ x \in \Gamma;\ 0 \in \Gamma) \tag{4.4}$$

lineáris, homogén rendszer 0-megoldásának stabilitását, ahol A az R_+-on értelmezett, $n \times n$-es mátrix-értékű folytonos függvény.

Ebben a fejezetben nem beszélünk parciális stabilitásról, aminek az az oka, hogy az első közelítés alapján való parciálisstabilitás-vizsgálat elmélete még nincs teljesen kidolgozva. A klasszikus eredmények átvitele néhány (érdekes és nehéz) problémába ütközik, amelyek megoldásában még csak az első lépések történtek meg (l. [49]).

4.1. TÉTEL. *A* (4.4) egyenlet 0-megoldása akkor és csak akkor

a) stabilis, ha (4.4) minden megoldása korlátos,

b) aszimptotikusan stabilis, ha (4.4) minden megoldásának 0 a határértéke, ha $t \to \infty$.

Bizonyítás. a) Tegyük fel, hogy (4.4) minden megoldása korlátos, és jelöljük $\Phi(t)$-vel (4.4) egy alapmátrixát (l. [27]). Ekkor van olyan K állandó, hogy

$$|x(t; t_0, x_0)| = |\Phi(t)\,\Phi^{-1}(t_0)\,x_0| \leqq K\,|x_0|,$$

hiszen $\Phi(t)$ oszlopvektorai (4.4) megoldásai. A kapott egyenlőtlenségből a 0-megoldás stabilitása azonnal adódik.

Tegyük fel, hogy a 0-megoldás stabilis. Ha (4.4)-nek van nem-korlátos megoldása, akkor annak alkalmas valós számmal való megszorzásával olyan megoldás is kapható, amely a t_0-ban tetszőlegesen kicsiny normájú és nem korlátos. Ez pedig ellentmond annak, hogy a 0-megoldás stabilis.

b) Tegyük fel, hogy (4.4) minden megoldása 0-hoz tart, ha $t \to \infty$. A 0-megoldás aszimptotikus stabilitáshoz már csak a 0-megoldás stabilitását kell bizonyítani. Mivel minden megoldás korlátos, ez következik *a)*-ból.

Legyen most a 0-megoldás aszimptotikusan stabilis. Ha (4.4)-nek van olyan megoldása, amelyre nem igaz, hogy $t \to \infty$ mellett 0 a határértéke, akkor van olyan is, amely rendelkezik ezzel a tulajdonsággal, de t_0-ban a normája tetszőlegesen kicsiny, ami ellentmond az aszimptotikus stabilitásnak.

Ha a (4.1) egyenlet autonóm, és a perturbálatlan $x = \varphi(t)$ mozgás egyensúlyi helyzet ($\varphi(t) =$ állandó), illetve periodikus mozgás, akkor a (4.3) első közelítés állandó együtthatós, illetve periodikus együtthatós rendszer. Fogalmazzuk meg a fenti tétel következményeit erre a két speciális esetre.

Ismeretes [27], hogy az

$$\dot{x} = Ax \quad (A\ n\times n\text{-es konstans mátrix}) \tag{4.5}$$

differenciálegyenlet-rendszernek van olyan alaprendszere, amelyben az A mátrix k-szoros λ sajátértékéhez k darab $x^{(i)}=p_{i-1}(t)e^{\lambda t}$ $(i=1, 2, \ldots, k)$ alakú megoldás tartozik, ahol $p_j(t)$ *legfeljebb* j-edfokú polinomokból álló vektor. Speciálisan, p_j akkor és csakis akkor 0-adfokú minden j-re, ha az A-nak λ-hoz k lineárisan független sajátvektora tartozik. Ebből azonnal adódik a

4.1. KÖVETKEZMÉNY. A (4.5) 0-megoldása akkor és csakis akkor

a) stabilis, ha A minden sajátértékének valós része nem pozitív, továbbá minden 0 valós részű sajátértékhez annyi lineárisan független sajátvektor tartozik, amennyi a sajátérték multiplicitása a $\det(A-\lambda E)=0$ karakterisztikus egyenletben.

b) aszimptotikusan stabilis, ha A minden sajátértékének valós része negatív.

Vizsgáljuk most az

$$\dot{x} = A(t)x \quad (A(t+\tau) = A(t);\ \tau > 0) \tag{4.6}$$

periodikus együtthatós rendszert. *Floquet tétele* [15] szerint a (4.6) egyenlet $\Phi(t)$ $(\Phi(0)=E)$ alapmátrixához van olyan $P(t)$ τ-periodikus, nem elfajuló, folytonos mátrixfüggvény, hogy az $x=P(t)u$ transzformáció a (4.6) rendszert az $\dot{u}=Bu$ konstans együtthatós rendszerbe viszi át, ahol B eleget tesz a $\Phi(\tau)=e^{B\tau}$ feltételnek. A $\Phi(\tau)$ monodromiamátrix sajátértékeit *karakterisztikus tényezőknek*, B sajátértékeit *karakterisztikus kitevőknek* nevezzük. A 4.1. következményből adódik a

4.2. KÖVETKEZMÉNY. A (4.6) 0-megoldása akkor és csakis akkor

a) stabilis, ha minden karakterisztikus tényező abszolút értéke nem nagyobb 1-nél, és minden γ $(|\gamma|=1)$ karakterisztikus tényezőhöz (mint sajátértékhez) a monodromiamátrixnak annyi lineárisan független sajátvektora tartozik, mint amennyi γ multiplicitása,

b) aszimptotikusan stabilis, ha minden karakterisztikus tényező abszolút értéke kisebb 1-nél.

Meg kell jegyezni, hogy a 4.1. és 4.2. következmény közötti hasonlóság ellenére a periodikus rendszer vizsgálata sokkal bonyolultabb, hiszen amíg a konstans együtthatós rendszernél a feltételek *közvetlenül* az A birtokában ellenőrizhetők, a periodikus esetben előbb meg kell határozni az n darab lineárisan független *megoldásból* álló alapmátrix τ helyen vett értékét, ami igen nehéz feladat. Most vázlatosan ismertetünk egy közelítő módszert a karakterisztikus tényezők meghatározására. Osszuk be a $[0, \tau]$ intervallumot m egyenlő részre:

$$0 = t_0 < t_1 < \ldots < t_{m-1} < t_m = \tau, \quad t_k - t_{k-1} =: h \quad (k = 1, 2, \ldots, m),$$

és vegyük a szakaszonként állandó

$$A_h(t) := \frac{1}{h}\int_{t_k}^{t_{k+1}} A(s)\,ds, \quad \text{ha} \quad t_k < t < t_{k+1} \quad (k = 0, 1, \ldots, m-1)$$

mátrixfüggvényt. Jelölje $\Phi_h(t)$ azt a folytonos mátrixfüggvényt, amely minden $t\neq t_k$ pontban kielégíti a $\dot{\Phi}_h=A_h\Phi_h$ mátrixegyenletet, és $\Phi_h(0)=E$. Mivel A_h szakaszonként állandó, a konstans együtthatós differenciálegyenlet-rendszerek meg-

oldási szabályát az egyes szakaszokra alkalmazva $\Phi_h(t)$ elemi függvények segítségével felírható. Be lehet bizonyítani [36], hogy $\lim_{h\to 0} \Phi_h(\tau)=\Phi(\tau)$. Mivel mátrix sajátértékei a mátrix elemeitől folytonosan függnek, $\Phi_h(\tau)$ sajátértékei tetszőleges pontossággal közelítik $\Phi(\tau)$ sajátértékeit, vagyis (4.6) karakterisztikus tényezőit, ha h elég kicsiny.

Most rátérünk az első közelítés alapján való stabilitásvizsgálatot megalapozó tételek ismertetésére. Ezek mindegyike azon alapszik, hogy ha van olyan V *Ljapunov-függvény,* amely az első közelítésre vonatkozóan kielégíti a 3.2. vagy 3.3. tétel feltételeit, akkor *ugyanaz* a V kielégíti ugyanannak a tételnek a feltételeit az eredeti egyenletre vonatkozóan is. Így felmerül az a fontos kérdés, hogy a (4.3) egyenlet 0-megoldásának aszimptotikus stabilitása (illetve instabilitása) biztosítja-e a 3.2. (illetve 3.3.) tétel feltételeit teljesítő V függvény létezését.

A 3.1—3.3. tételek megfordíthatóságának kérdése a nem-lineáris (3.1) egyenletre is felmerült [25, 39, 43]. Ez nemcsak elméletileg érdekes probléma, hanem az ún. totális stabilitás elegendő feltételének megadása szempontjából is fontos [43].

Tekintsük először az $A(t)$ = konstans esetet. Vizsgáljuk meg, hogy van-e olyan $V(x)=(x, Bx)$ kvadratikus alak, amelynek a (4.5) egyenletre vonatkozó deriváltja egy előre adott $W=(x, Cx)$ kvadratikus alakkal egyenlő (B, C szimmetrikus, konstans mátrix). Mivel $\dot{V}=(x, [A^*B+BA]x)$, ahol A^* A transzponáltja, a kérdés a

$$A^*B+BA = C \tag{4.7}$$

B-re való megoldhatóságával ekvivalens.

4.1. LEMMA. Ha az A mátrix bármely két sajátértékének összege 0-tól különböző, akkor bármely $W(x)=(x, Cx)$ kvadratikus alakhoz pontosan egy olyan $V(x)=(x, Bx)$ tartozik, amelynek (4.5)-re vonatkozó deriváltjára $\dot{V}(x)=W(x)$ teljesül.

Bizonyítás. Tekintsük az $F(B):=A^*B+BA$ operátort. Azt kell megmutatni, hogy F invertálható, ami ekvivalens azzal, hogy F minden sajátértéke 0-tól különböző.

Tegyük fel, hogy γ sajátértéke F-nek, vagyis van olyan $B\neq 0$ mátrix, hogy $F(B)=\gamma B$, azaz $A^*B+BA=\gamma B$, ahonnan $(A^*-\gamma E)B=-BA$ is következik.

Megmutatjuk, hogy az $A^*-\gamma E$ és a $-A$ mátrixoknak van közös sajátértékük. Ha nincs, akkor ezen mátrixok karakterisztikus polinomjait rendre $g(\lambda)$-val és $f(\lambda)$-val jelölve, $g(\lambda)$-nak és $f(\lambda)$-nak nincs legalább elsőfokú közös osztója. Ekkor létezik olyan $g_1(\lambda)$ és $f_1(\lambda)$ polinom, hogy $g(\lambda)g_1(\lambda)+f(\lambda)f_1(\lambda)=1$. Legyen $h(\lambda):=g_1(\lambda)g(\lambda)$. A *Cayley—Hamilton-tétel* [36] szerint minden mátrix kielégíti karakterisztikus egyenletét, tehát $h(A^*-\gamma E)=0$ és $h(-A)=E$. Másrészt, $h(A^*-\gamma E)B=Bh(-A)$, így $B=0$, ami ellentmondás.

Tehát, van A-nak olyan λ és μ sajátértéke, hogy $\lambda-\gamma=-\mu$. A lemma feltétele szerint $\lambda+\mu=\gamma\neq 0$, és ezt akartuk bizonyítani.

4.2. LEMMA. *a)* Ha (4.5) 0-megoldása aszimptotikusan stabilis, akkor bármely negatív definit W kvadratikus alakhoz pontosan egy olyan V kvadratikus alak tartozik, amelynek a (4.5) egyenletre vonatkozó deriváltja W, és ez a kvadratikus alak pozitív definit.

b) Ha a (4.5) egyenlet 0-megoldása úgy instabilis, hogy létezik pozitív valós részű sajátértéke A-nak, akkor bármely pozitív definit W kvadratikus alakhoz létezik olyan α pozitív szám és V pozitív értéket is felvevő kvadratikus alak, hogy V (4.5)-re vonatkozó deriváltjára $\dot{V} \geqq \alpha V + W$ teljesül.

Bizonyítás. a) Tegyük fel, hogy (4.5) 0-megoldása aszimptotikusan stabilis, ami a 4.1. következmény szerint azt jelenti, hogy A minden sajátértéke negatív valós részű. Ebből következik, hogy teljesül a 4.1. lemma feltétele, tehát tetszőleges W-hez pontosan egy olyan V létezik, hogy $\dot{V} = W$. Azt kell csak bebizonyítani, hogy V pozitív definit. Az világos, hogy V sehol nem vehet fel negatív értéket, mert ellenkező esetben az $x=0$ tetszőleges környezetében is venne fel negatív értéket, amiből W negatív definitása miatt a 3.3. tételből a 0-megoldás instabilitása következne. De akkor $V(x_0)=0$ $(x_0 \neq 0)$ sem teljesülhet, hiszen W negatív definit, ezért $V(x(t; 0, x_0))$ szigorúan csökkenő, így V felvesz negatív értéket is, amiről az előbb mutattuk meg, hogy lehetetlen.

b) Tekintsük az

$$\dot{u} = \left(A - \frac{\alpha}{2} E\right) u \quad (\alpha > 0) \tag{4.8}$$

rendszert. Egyszerű meggondolás mutatja, hogy α megválasztható úgy, hogy az $A-(\alpha/2)E$ mátrixnak van pozitív valós részű sajátértéke, és bármely két saját értékének összege 0-tól különböző. A 4.1. lemma szerint pontosan egy olyan $V(u) = (u, Bu)$ létezik, amelynek (4.8)-ra vonatkozó deriváltjára $\dot{V}_{(4.8)}(u) = W(u) = (u, Cu)$ teljesül. Mivel

$$\dot{V}_{(4.8)}(u) = \left(\left[A - \frac{\alpha}{2} E\right] u, \operatorname{grad} V(u)\right) = 2(Au, Bu) - \alpha(u, Bu) =$$

$$= \dot{V}_{(4.5)}(u) - \alpha V(u) = W(u),$$

a $V(x)$-nek a (4.5) rendszerre vonatkozó deriváltjára a $\dot{V}_{(4.5)}(x) = \alpha V(x) + W(x)$ teljesül. Azt kell még bebizonyítani, hogy V-nek van pozitív értéke. Ha V negatív definit, akkor $-V$ pozitív definit, és $(-V)'_{(4.8)}(u) = -W(u)$ negatív, így a 3.1.tétel szerint (4.8) 0-megoldása stabilis, ami lehetetlen, hiszen együttható-mátrixának van pozitív valós részű sajátértéke. Következésképpen van olyan $u_0 \neq 0$, hogy $V(u_0)=0$. De $\dot{V}_{(4.8)}(u_0) = W(u_0) > 0$, így a $V(u(t; 0, u_0))$ függvénynek a $t=0$ szigorú növekedési pontja, tehát V vesz fel pozitív értéket.

Ezzel a lemma bizonyítva van.

A 4.2. lemmának nem csak elméleti jelentősége van. Sevítségével levezethetünk egy olyan becslést, ami igen jól használható a gyakorlatban a (4.5) rendszer megoldásainak tanulmányozásánál.

Tegyük fel, hogy az A mátrix minden sajátértékének valós része negatív. A 4.2. lemma szerint van olyan $V(x)=(x, Bx)$ pozitív definit kvadratikus alak, amelyre $\dot{V}(x) = -|x|^2$. A feltételes szélsőérték meghatározásának *Lagrange-féle módszerével* könnyen megállapítható, hogy

$$\min\{V(x): |x| = r\} = \lambda r^2, \quad \max\{V(x): |x| = r\} = \Lambda r^2,$$

ahol λ, Λ a B szimmetrikus mátrix legkisebb, illetve legnagyobb sajátértéke. Ezért

$$V(x)/\Lambda \leqq |x|^2 \leqq V(x)/\lambda, \tag{4.9}$$

amelynek felhasználásával a $-V/\lambda \leqq \dot{V} \leqq -V/\Lambda$ becslés adódik.

Tekintsük most a (4.5) egyenlet $x(t;\ t_0, x_0)$ megoldását és definiáljuk a $v(t) := V(x(t;\ t_0, x_0))$ függvényt. Az eddigiek szerint erre igaz a $-1/\lambda \leqq \dot{v}(t)/v(t) \leqq -1/\Lambda$ becslés, amelynek integrálásával, a (4.9) becslést is felhasználva, az

$$\frac{V(x_0)}{\Lambda} \exp\left[-\frac{t-t_0}{\lambda}\right] \leqq |x(t;\ t_0, x_0)|^2 \leqq \frac{V(x_0)}{\lambda} \exp\left[-\frac{t-t_0}{\Lambda}\right] \tag{4.10}$$

egyenlőtlenség adódik ($t \geqq t_0$). Ennek segítségével meg lehet becsülni többek között annak az időtartamnak a nagyságát, amelynek el kell telnie ahhoz, hogy a megoldás az ε sugarú, origó középpontú gömbbe jusson:

$$\lambda \ln \frac{V(x_0)}{\Lambda \varepsilon^2} \leqq T - t_0 \leqq \Lambda \ln \frac{V(x_0)}{\lambda \varepsilon^2}.$$

4.3. LEMMA. Legyen W definit, V tetszőleges kvadratikus alak, $R: R_+ \times \times \Gamma \to R^n$ pedig olyan függvény, hogy $R(t, x) = o(|x|)$ t-ben egyenletesen a $[0, \infty)$ intervallumlon, azaz bármely $\varepsilon > 0$-hoz van olyan $\delta(\varepsilon) > 0$, hogy ha $|x| < \delta$, akkor $|R(t, x)| \leqq \varepsilon |x|$. Ekkor $W + (\operatorname{grad} V, R)$ ugyanolyan értelemben definit az $x = 0$ egy kis környezetében, mint W.

Bizonyítás. Tegyük fel a határozottság kedvéért, hogy W negatív definit. (4.9) szerint van olyan $\Lambda < 0$ és $\Theta > 0$ szám, hogy

$$W(x) \leqq \Lambda |x|^2, \quad |\operatorname{grad} V(x)|^2 \leqq \Theta |x|^2.$$

A *Cauchy-féle egyenlőtlenség* szerint

$$W(x) + (\operatorname{grad} V(x),\ R(t, x)) \leqq W(x) + |\operatorname{grad} V(x)|\, |R(t, x)| \leqq$$

$$\leqq (\Lambda + \varepsilon \sqrt{\Theta})\, |x|^2 \quad (|x| < \delta(\varepsilon)),$$

tehát $W + (\operatorname{grad} V, R)$ is negatív definit, ha $|x|$ elég kicsi.

4.2. TÉTEL. Tekintsük az

$$\dot{x} = Ax + R(t, x) \quad (A = \textit{konstans}) \tag{4.11}$$

differenciálegyenlet-rendszert, és tegyük fel, hogy $R(t, x) = o(|x|)$ t-ben egyenletesen a $[0, \infty)$ intervallumon.

a) Ha A minden sajátértékének valós része negatív, akkor (4.11) 0-megoldása aszimptotikusan stabilis, sőt exponenciálisan stabilis. Ez utóbbi azt jelenti, hogy létezik olyan a, σ, γ pozitív szám, amelyekkel (4.11) megoldásaira

$$|x(t;\ t_0, x_0)| \leqq a |x_0| e^{-\gamma(t-t_0)} \quad (t \geqq t_0,\ t_0 \in R_+,\ |x_0| < \sigma) \tag{4.12}$$

teljesül.

b) Ha A-nak van pozitív valós részű sajátértéke, akkor (4.1) 0-megoldása instabilis.

Bizonyítás. a) A 4.2. lemma biztosítja olyan V pozitív definit kvadratikus alak létezését, amelyre $\dot{V}_{(4.5)}(x) = -|x|^2$. Mivel

$$\dot{V}_{(4.11)}(t, x) = -|x|^2 + (\operatorname{grad} V(x),\ R(t, x)),$$

a 4.3. lemma szerint $\dot{V}_{(4.11)}$ negatív definit az $x=0$ egy kis környezetében, ami azt jelenti, hogy a 3.2. tétel minden feltétele teljesül, tehát (4.11) 0-megoldása aszimptotikusan stabilis.

(4.12) ugyanúgy vezethető le, mint (4.10) a (4.5) egyenlet megoldásaira.

b) A 4.2. lemma biztosítja olyan V, pozitív értékeket is felvevő kvadratikus alak létezését, amelyre $\dot{V}_{(4.5)}(x) = |x|^2 + \alpha V(x)$ $(\alpha > 0)$. A 4.3. lemma szerint van olyan $c > 0$, hogy

$$\dot{V}_{(4.11)}(t, x) = \alpha V(x) + \left(|x|^2 + (\operatorname{grad} V(x),\ R(t, x))\right) \geqq \alpha V(x) + c\,|x|^2.$$

Felhasználva, hogy $|x|^2 \geqq V/\Lambda$, ahol a Λ pozitív szám a V kvadratikus alak mátrixának legnagyobb sajátértéke, a $\dot{V}_{(4.11)} \geqq (\alpha + c/\Lambda)V$ becslést kapjuk. A tétel állítása ezek után a 3.3. tétel következménye.

A 4.2. következmény levezetésénél láttuk, hogy *Floquet tételének* felhasználásával a periodikus lineáris egyenlet tanulmányozása a konstans együtthatós egyenletre visszavezethető. Ezzel a módszerrel adódik a most bizonyított tételből a

4.3. TÉTEL. Tekintsük a

$$\dot{x} = A(t)x + R(t, x) \quad \left(A(t+\tau) = A(t),\ \tau > 0\right) \tag{4.13}$$

differenciálegyenlet-rendszert, és tegyük fel, hogy $R(t, x) = o(|x|)$ t-ben egyenletesen a $[0, \infty)$ intervallumon.

a) Ha (4.6) minden karakterisztikus tényezőjének abszolút értéke kisebb 1-nél, akkor (4.13) 0-megoldása aszimptotikusan stabilis, sőt exponenciálisan stabilis (l. (4.12)).

b) Ha (4.6)-nak van 1-nél nagyobb abszolút értékű karakterisztikus tényezője, akkor (4.13) 0-megoldása instabilis.

Térjünk most rá az általános eset vizsgálatára: (4.3) mely tulajdonságából tudunk következtetni (4.2) 0-megoldásának stabilitására, ha $A(t)$ folytonos mátrixfüggvény, amelyről periodicitás nincs feltételezve? Az $A(t) = A = \text{konstans}$ esetben azt kaptuk, hogy ha az első közelítés 0-megoldása aszimptotikusan stabilis, akkor az eredeti rendszer $\varphi(t)$ megoldása is aszimptotikusan stabilis (l. 4.2. tétel, *a)* állítás). Az általános esetben ez a kijelentés nem igaz. Ez azon múlik, hogy az $A(t) = \text{konstans}$ esetben a lineáris rendszer 0-megoldásának aszimptotikus stabilitása *implikál* egy jóval erősebb tulajdonságot, nevezetesen, hogy a megoldások normája *exponenciálisan* tart 0-hoz, ha $t \to \infty$-hez (l. (4.10)). Változó együtthatók mellett ez nem igaz (például az $\dot{x} = -(1+t)/x$ skaláris egyenlet általános megoldása $x = x_0/(1+t)$). Így felmerül a kérdés: ha az első közelítés megoldásairól *feltesszük,* hogy exponenciálisan 0-hoz tartanak $(t \to \infty)$, akkor az eredeti rendszer $\varphi(t)$ megoldása már stabilis-e? A válasz igenlő.

4.4. TÉTEL. Tekintsük az

$$\dot{x} = A(t)x + R(t, x) \tag{4.14}$$

differenciálegyenlet-rendszert, ahol $R(t, x) = o(|x|)$ t-ben egyenletesen a $[0, \infty)$-intervallumon. Ha a (4.4) rendszer megoldásai eleget tesznek a (4.12) becslésnek, akkor (4.14) 0-megoldása aszimptotikusan stabilis.

Bizonyítás. A módszer ugyanaz, mint eddig: a feltétel segítségével konstruálunk egy *Ljapunov-függvényt,* és azt a (4.14) rendszerre alkalmazzuk. A bizonyításnak [43, 376. o.] itt csak egy egész rövid vázlatát adjuk, aminek a helyhiányon kívül az is oka, hogy — a konstansegyütthatós esettől eltérően — a konstrukció nem direkt, mivel a (4.4) rendszer megoldásainak ismeretét tételezi fel, és így az alkalmazásokban nem használható.

Jelölje $\Phi(t)$ a (4.4) egyenlet azon alapmátrixát, amelyre $\Phi(0)=E$. Legyen $W(t, x)$ az x vektor komponenseinek pozitív definit kvadratikus alakja, amelynek együtthatófüggvényei folytonosak és korlátosak a $[0, \infty)$ intervallumon. Be lehet bizonyítani [43, 322. o.], hogy a

$$V(t, x) := \int_t^\infty W(s, \Phi(s)\Phi^{-1}(t)x)ds$$

függvény eleget tesz a 3.2. tétel minden feltételének ($y=x$), és így a (4.14) 0-megoldása aszimptotikusan stabilis.

A fenti tételeknek számos érdekes alkalmazása van. A *Watt-féle centrifugális regulátor* problémáját például a 4.2. tétel segítségével sikerült megoldani [27], de ez csak egyike az automatikus szabályozások elméletébe tartozó feladatoknak, amelyeknek megoldásában a *Ljapunov-módszer* alkalmazható [10, 24, 46]. Mi most két példát mutatunk be annak illusztrálására, hogy a módszer hogyan használható a rezgések elméletében.

4.1. PÉLDA. Tekintsük a hintázás problémáját. Ez abban áll, hogy egy fizikai inga stabilis egyensúlyi helyzetét valamilyen módon instabilissá kell tenni. Ennek egyik módja külső erő alkalmazása. Jól ismert tény, hogy ha a külső periodikus erő frekvenciája megegyezik a hinta saját frekvenciájával, akkor rezonancia jelensége lép fel, az egyensúlyi helyzet stabilitása megszűnik. Bonyolultabb azonban a kérdés, ha külső erő nem hat, hanem a hintázó gyermeknek bizonyos frekvenciával periodikusan változtatnia kell súlypontjának magasságát, ami a redukált matematikai inga hosszának periodikus megváltoztatását jelenti. Mint a tapasztalat mutatja, alkalmas frekvenciánál az instabilitás létrejön. Ezt a jelenséget *parametrikus rezonanciának* nevezzük. Adjuk meg ennek matematikai tárgyalását a hintára vonatkozóan.

A hintának megfelelő matematikai inga helyzetét jellemezzük a fonalnak a gravitáció irányába mutató tengelytől az óramutató járásával ellenkező irányban mért x szöggel. Ekkor a mozgásegyenlet $\ddot{x} = -\omega^2 \sin x$ ($\omega^2=g/l$, ahol g a nehézségi gyorsulás, l az inga hossza). Ha az inga hossza v frekvenciával változik, és szakaszonként állandó, akkor a mozgásegyenlet a következő alakban írható:

$$\ddot{x} = -f^2(t)\sin x, \quad f(t) = \begin{cases} \omega+\varepsilon, & 0 \leqq t < \dfrac{\pi}{v} \\ \omega-\varepsilon, & \dfrac{\pi}{v} \leqq t < \dfrac{2\pi}{v} \end{cases} \quad (\varepsilon \ll 1) \tag{4.15}$$

$$f\left(t+\frac{2\pi}{v}\right) = f(t).$$

Kérdés: az ε, ν paraméterek mely értékeinél lesz az $x=0$, $\dot{x}=0$ egyensúlyi helyzet instabilis? Oldjuk meg a kérdést először a (4.15) rendszer $\ddot{x}=-f^2(t)x$ első közelítésére, amely az $x_1:=x$, $x_2:=\dot{x}$ változókkal az

$$\dot{x}_1 = x_2, \quad \dot{x}_2 = -f^2(t)x_1 \tag{4.16}$$

alakba írható.[1] Határozzuk meg (4.16) monodrómiamátrixát. Ez $\Phi(2\pi/\nu) = \Phi_2(\pi/\nu)\Phi_1(\pi/\nu)$ alakban áll elő, ahol $\Phi_1(t)$, illetve $\Phi_2(t)$ a (4.16) alapmátrixa a $[0, \pi/\nu)$, illetve a $[\pi/\nu, 2\pi/\nu)$ intervallumon. Nyilván

$$\Phi_k(\pi/\nu) = \begin{pmatrix} c_k & s_k/\omega_k \\ -\omega_k s_k & c_k \end{pmatrix} \begin{matrix} \omega_1 := \omega+\varepsilon & \omega_2 := \omega-\varepsilon \\ c_k := \cos\omega_k\dfrac{\pi}{\nu} & s_k := \sin\omega_k\dfrac{\pi}{\nu} \end{matrix} \quad (k=1,2).$$

A karakterisztikus tényezők eleget tesznek a

$$\lambda_1+\lambda_2 = \operatorname{tr}\Phi(2\pi/\nu) = 2c_1c_2-\left(\frac{\omega_1}{\omega_2}+\frac{\omega_2}{\omega_1}\right)s_1s_2 =: M$$

$$\lambda_1\lambda_2 = \det\Phi(2\pi/\nu) = 1$$

rendszernek.[2] Tehát ha $|M|<2$, akkor két, egységnyi abszolút értékű karakterisztikus tényező van, és a 4.2. következmény szerint (4.16) 0-megoldása stabilis. Ha $|M|>2$, akkor van 1-nél nagyobb abszolút értékű karakterisztikus tényező, és a 0-megoldás instabilis. Az ε, ν paramétersíkon a stablilis és instabilis rendszert meghatározó paraméterpárokat tehát az $|M|=2$ görbe választja el. Határozzuk meg ezt a görbét.

Egyszerű számolás mutatja, hogy

$$\frac{\omega_1}{\omega_2}+\frac{\omega_2}{\omega_1} = 2\left(1+\frac{2\varepsilon^2}{\omega^2-\varepsilon^2}\right) = 2(1+\Delta)$$

$$\Delta = \frac{2\varepsilon^2}{\omega^2}\left(1+\frac{\varepsilon^2}{\omega^2}+\frac{\varepsilon^4}{\omega^4}+\ldots\right) = \frac{2\varepsilon^2}{\omega^2}+O(\varepsilon^4)$$

$$2c_1c_2 = \cos\frac{2\pi}{\nu}\varepsilon+\cos\frac{2\pi}{\nu}\omega, \quad 2s_1s_2 = \cos\frac{2\pi}{\nu}\varepsilon-\cos\frac{2\pi}{\nu}\omega,$$

tehát az $|M|=2$ egyenlet $|-\Delta\cos 2\pi\varepsilon/\nu+(2+\Delta)\cos 2\pi\omega/\nu|=2$ alakú, vagyis két egyenletre esik szét:

$$\cos 2\pi\omega/\nu = \frac{2+\Delta\cos 2\pi\varepsilon/\nu}{2+\Delta}, \quad \cos 2\pi\omega/\nu = \frac{-2+\Delta\cos 2\pi\varepsilon/\nu}{2+\Delta}. \tag{4.17}$$

[1] Az eddigiekben mindenütt feltettük az egyszerűség kedvéért, hogy a rendszerek jobb oldalán álló függvények folytonosak. Ez itt nem teljesül, mivel f csak szakaszonként folytonos. Egyszerű azonban meggondolni, hogy ha a megoldástól folytonosságot és f ugráshelyeitől különböző pontokban differenciálhatóságot követelünk meg, akkor a tételek érvényben maradnak.

[2] Ha B négyzetes mátrix, akkor $\operatorname{tr} B$ a B mátrix nyomát jelöli, ami a főátlóban álló elemek összege.

Az elsőt vizsgálva látjuk, hogy a jobb oldal közel 1, tehát érdemes bevezetni az $\omega/\nu =: k+a$ (k nemnegatív egész) jelölést. Ekkor $\cos 2\pi\omega/\nu = \cos 2\pi a = 1-2\pi^2a^2+ +O(a^4)$, és az egyenlet

$$\cos\frac{2\pi}{\nu}\omega = 1-\frac{\Delta}{2+\Delta}\left(1-\cos\frac{2\pi}{\nu}\varepsilon\right),$$

vagyis $2\pi^2a^2+O(a^4)=\Delta(\pi^2\varepsilon^2+O(\varepsilon^4))$. Figyelembe véve, hogy $\Delta=2\varepsilon^2/\omega^2+O(\varepsilon^4)$, az $|a|=\varepsilon^2/\omega^2+o(\varepsilon^2)$ adódik, vagyis

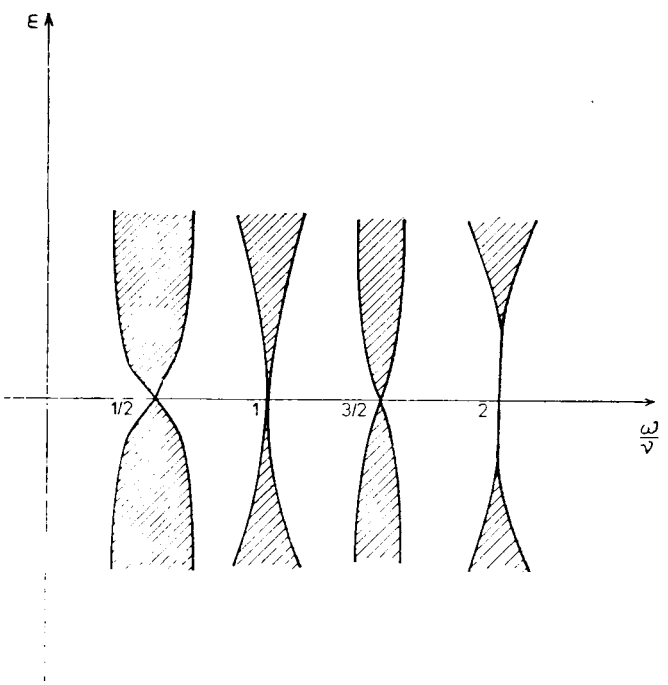
$$\frac{\omega}{\nu} = k\pm\frac{\varepsilon^2}{k^2}+o(\varepsilon^2) \quad (k=0,1,2,\ldots).$$

A (4.17) második egyenletéből hasonló módon

$$\frac{\omega}{\nu} = k+\frac{1}{2}\pm\frac{\varepsilon}{\pi\left(k+\frac{1}{2}\right)}+o(\varepsilon) \quad (k=0,1,2,\ldots)$$

következik.

A besatírozott tartományhoz tartozó paraméterpárokra a (4.16) 0-megoldása instabilis. Mivel az első közelítésnek van 1-nél nagyobb karakterisztikus tényezője, a 4.3. tétel szerint az eredeti egyenlet 0-megoldása is instabilis.



1. ábra

Tehát a tapasztalattal megegyezően azt kaptuk, hogy a destabilizálás akkor a leghatásosabb, amikor az inga hosszának a változtatása az inga saját frekvenciájának kétszeresével történik.

4.2. PÉLDA. Legyen egy tömegpont egy rúd végére erősítve, amely egy vízszintes tengely körül foroghat, és az alátámasztási pont képes függőleges irányú rezgéseket végezni. P. L. KAPICA szovjet fizikus, a *fizikai Nobel-díj* 1978. évi nyertese fedezte fel 1951-ben azt a jelenséget, hogy ha az alátámasztási pontot elég nagy frekvenciával függőleges irányban rezegtetjük, akkor az inga felső egyensúlyi helyzete, amely fix alátámasztási pont esetén instabilis, stabilissá válik.

Jelölje az inga hosszát l, és tegyük fel, hogy az alátámasztási pont rezgésének amplitúdója a ($a \ll l$), periódusa 2τ, miközben az egyes félperiódusokban a gyorsulás állandó $\pm c$ (ekkor nyilván $c = 8a/\tau^2$). A mozgásokat az

$$\ddot{x} = f(t)\sin x, \quad f(t) = \begin{cases} \omega^2+\alpha^2, & 0 \leqq t < \tau \\ \omega^2-\alpha^2, & \tau \leqq t < 2\tau \end{cases}$$

egyenlet írja le, ahol x az inga felső egyensúlyi helyzetéhez viszonyított szögkitérést jelöli, $\omega^2 = g/l$, $\alpha^2 = c/l$. Vizsgáljuk az $x = \dot{x} = 0$ stabilitását első közelítésben. Hasonlóan az előző feladathoz, a monodrómia-mátrixot itt is $\Phi(2\tau) = \Phi_2(\tau)\Phi_1(\tau)$ alakban kapjuk, ahol

$$\Phi_1(\tau) = \begin{pmatrix} \operatorname{ch} k\tau & \frac{1}{k}\operatorname{sh} k\tau \\ k \operatorname{sh} k\tau & \operatorname{ch} k\tau \end{pmatrix}, \quad \Phi_2(\tau) = \begin{pmatrix} \cos \Omega\tau & \frac{1}{\Omega}\sin \Omega\tau \\ -\Omega \sin \Omega\tau & \cos \Omega\tau \end{pmatrix}$$

$$k^2 = \alpha^2+\omega^2 \qquad \Omega^2 = \alpha^2-\omega^2$$

A stabilitás feltétele itt is $|\operatorname{tr} \Phi(2\tau)| < 2$, vagyis

$$\left| 2 \operatorname{ch} k\tau \cos \Omega\tau + \left(\frac{k}{\Omega} - \frac{\Omega}{k}\right) \operatorname{sh} k\tau \sin \Omega\tau \right| < 2. \tag{4.18}$$

Vezessük be az $\varepsilon^2 := a/l$, $\mu^2 := g/c$ jelöléseket. Felhasználva a $c = 8a/\tau^2$ összefüggést írjuk fel ezt a feltételt az új változókkal. A benne szereplő mennyiségekre a

$$k\tau = 2\sqrt{2}\,\varepsilon\sqrt{1+\mu^2}, \quad \Omega\tau = 2\sqrt{2}\,\varepsilon\sqrt{1-\mu^2},$$

$$\frac{k}{\Omega} - \frac{\Omega}{k} = \sqrt{\frac{1+\mu^2}{1-\mu^2}} - \sqrt{\frac{1-\mu^2}{1+\mu^2}} = 2\mu^2 + O(\mu^4)$$

formulák adódnak, amelyeknek felhasználásával az alábbi sorfejtések adódnak:

$$\operatorname{ch} k\tau = 1 + 4\varepsilon^2(1+\mu^2) + \frac{8}{3}\varepsilon^4 + \ldots$$

$$\cos \Omega\tau = 1 - 4\varepsilon^2(1-\mu^2) + \frac{8}{3}\varepsilon^4 + \ldots$$

$$\left(\frac{k}{\Omega} - \frac{\Omega}{k}\right) \operatorname{sh} k\tau \sin \Omega\tau = 16\varepsilon^2\mu^2 + \ldots;$$

a ki nem írt tagok $O(\varepsilon^4+\mu^4)$ nagyságrendűek. Tehát (4.18) a

$$2\left(1-16\varepsilon^4+\frac{16}{3}\varepsilon^4+8\varepsilon^2\mu^2+\ldots\right)+16\varepsilon^2\mu^2<2$$

alakba írható. Így elég kicsiny ε-ra és μ-re a stabilitás feltétele: $\mu^2<2\varepsilon^2/3$, vagyis $g/c<2a/(3l)$, ez pedig elérhető, ha az alátámasztási pont rezgésének frekvenciáját elég nagyra választjuk, mivel akkor c elég nagy.

Tehát az első közelítés egyensúlyi helyzete stabilissá tehető, amit úgy is szoktak mondani, hogy az inga egyensúlyi helyzete kis rezgésekre stabilis. Az ún. közepelési módszerrel [15] megmutatható, hogy a stabilitás nemcsak a kis rezgésekre, hanem az eredeti mozgásegyenletre vonatkozóan is megvalósul.

5. Az aszimptotikus stabilitás vizsgálata felülről nem-korlátos Ljapunov-függvényekkel

A nem-autonóm rendszerek vizsgálatában gyakori, hogy olyan *Ljapunov-függvényt* tudunk csak konstruálni, amely az időtől is explicit módon függ. Ha ilyen $V(t, x)$ függvény alapján aszimptotikus stabilitást szeretnénk megállapítani, akkor a 3.2. tételt kell használnunk. Gyakran nagy nehézséget okoz a (3.3) feltételek közül az elsőnek a kielégítése. Először V. P. MARACSKOV [44] kezdett foglalkozni azzal a kérdéssel, hogy ez a feltétel hogyan gyengíthető. Példát mutatott rá, hogy a feltétel általában nem hagyható el. Bebizonyította azonban, hogy helyette elegendő megkövetelni $|Y(t, y, z)|$ korlátosságát egy $R^+\times G^p(0, \varrho)\times R^q$ $(\varrho>0)$ halmazon. További alkalmas feltételek keresésével azóta is hatékonyan foglalkoznak [1, 12, 22, 23, 28, 49].

A kérdéskör illusztrálására a változó fonalhosszúságú, súrlódó közegben lengő matematikai inga alsó egyensúlyi helyzetének stabilitását fogjuk vizsgálni. Ha a pont tömege m, a fonalnak a függőlegesen lefelé mutató iránytól mért szögeltérése x, és a fonal hossza egy, *a rendszer mozgásától függetlenül adott* $l=l(t)$ törvény szerint változik, akkor a rendszer kinetikai energiája és a rendszerre ható erő:

$$T=\frac{1}{2}m\left[l^2(t)\dot{x}^2+(\dot{l}(t))^2\right], \quad Q=-mgl(t)\sin x-h(t)\dot{x},$$

ahol $h(t)>0$ (az időtől is függő) súrlódási együttható. A mozgásegyenlet:

$$\ddot{x}+\left[2\frac{\dot{l}(t)}{l(t)}+h(t)\right]\dot{x}+\frac{g}{l(t)}\sin x=0 \quad \left(-\frac{\pi}{2}<x<\frac{\pi}{2}\right), \tag{5.1}$$

amelyet az $a(t):=2\dot{l}(t)/l(t)+h(t)$, $b(t):=g/l(t)$; $x_1:=x$, $x_2:=\dot{x}$ jelölésekkel az

$$\dot{x}_1=x_2, \quad \dot{x}_2=-b(t)\sin x_1-a(t)x_2 \tag{5.2}$$

alakban is írhatunk. Ha az inga hossza és a súrlódási együttható állandó $(l(t)=l_0$, $h(t)=h_0>0$; $a_0:=h_0$, $b_0:=g/l_0)$, akkor tapasztalat szerint a $\varphi=\dot{\varphi}=0$ egyensúlyi helyzet aszimptotikusan stabilis. Valóban, a

$$V=\frac{1}{a_0b_0}\left[(a_0^2+b_0+b_0^2)(1-\cos x_1)+a_0x_1x_2+\frac{1+b_0}{2}x_2^2\right]$$

Ljapunov-függvény pozitív definit, $\dot{V} = -(x_1 \sin x_1 + x_2^2)$, tehát a 3.2. tétel alkalmazható. V úgy adódik, hogy (5.2) első közelítéséhez megoldjuk a (4.7) egyenletet $C = -E$ mellett, majd az (x, Bx) kvadratikus alakban $x_1^2/2$ helyére $(1-\cos x_1)$-et írunk. Érdekes, hogy az $x_2^2/2 + b_0(1-\cos x_1)$ teljes energia nem felel meg célunknak, mivel deriváltja $-a_0 x_2^2$, és ez nem negatív definit, helyette megfelel V, ami egy elvont matematikai konstrukció terméke, és semmilyen fizikai jelentése nincs!

Tegyük most fel, hogy az inga hossza állandó, de a súrlódási együttható változik, éspedig nő. Sejtjük, hogy ha ez a növekedés nagyon gyors, akkor előfordulhat, hogy a súrlódás hatására az inga „beragad", a szögkitérés nem tart 0-hoz, ha $t \to \infty$. Erre J. K. HALE [15] adott példát, amennyiben megmutatta, hogy az $\ddot{x} + (2 + e^t)\dot{x} + \sin x = 0$ egyenlet 0-megoldása nem aszimptotikusan stabilis. Vajon, ha $a(t) \to \infty$ $(t \to \infty)$, akkor soha nem lehet aszimptotikus stabilitás? Ha $a(t)$ akármilyen kicsiny is lehet bizonyos időpillanatokban, akkor az aszimptotikus stabilitás megszűnik? Az inga hosszának változása hogyan befolyásolja a jelenséget? Ezen kérdések megválaszolásához a 3.2. tétel egy olyan továbbfejlesztését adjuk meg, amelyekben (3.3) első feltétele nem szerepel.

5.1. DEFINÍCIÓ. Azt mondjuk, hogy a $\varphi: R_+ \to R_+ (R_+ \to R_-)$ folytonos függvény *integrálisan pozitív (negatív)*, ha bármely

$$S = \bigcup_{k=1}^{\infty} [a_k, b_k] \quad (a_k < b_k < a_{k+1},\ b_k - a_k \geqq \delta > 0\ (k = 1, 2, \ldots))$$

halmazra $\int_S \varphi(t)\, dt = \infty\ (= -\infty)$.

5.2. DEFINÍCIÓ. Egy $U: R_+ \times G^p(0, K) \times G^q(0, L) \to R$ folytonos függvényt *y-ban integrálisan pozitív definitnek* nevezünk, ha bármely $\alpha\ (0 < \alpha < K)$ számra a

$$\min \{U(t, y, z):\ \alpha \leqq |y| \leqq K,\ |z| \leqq L\}$$

függvény integrálisan pozitív. Az U függvény *y-ban integrálisan negatív definit*, ha $-U$, pozitív definit.

Tegyük fel az egyszerűség kedvéért, hogy van olyan $H, L > 0$ szám, amelyre áll: ha $|x_0| < H$, akkor $|z(t; t_0, x_0)| < L$ a $[t_0, \infty)$ intervallumon, vagyis minden megoldás z-korlátos. Legyen továbbá $x = (y', z')$ $(y' \in R^{p'}, z' \in R^{q'}; p' \geqq p)$ az x vektornak egy további particionálása.

5.1. TÉTEL. Tegyük fel, hogy a (3.1) egyenlet 0-megoldása y'-stabilis, továbbá létezik egy $V: R_+ \times \Gamma \to R$ és egy $W: R_+ \times \Gamma \to R^k$ folytonos parciális differenciálhányadosokkal bíró függvény, amelyek az $R_+ \times G^p(0, H) \times G^q(0, L)$ halmazon rendelkeznek az alábbi tulajdonságokkal:

(1) $V(t, x)$ alulról korlátos;
(2) $\dot{V}(t, x)$ y'-ben integrálisan negatív definit;
(3) létezik olyan $d, e \in \mathcal{K}$, hogy

$$d(|y|) \leqq |W(t, x)| \leqq e(|y'|);$$

(4) minden l, α $(1 \leqq l \leqq k;\ 0 < \alpha < H)$ számra

$$\int_0^t \max \{[\dot{W}_l(s, x)]_{+(-)}:\ \alpha \leqq |y'| \leqq H,\ |z| \leqq L\} ds$$

egyenletesen folytonos a $[0, \infty)$ intervallumon (itt a $[\cdot]_{+(-)}$ szimbólum azt jelenti, hogy vagy minden (t, x)-re $[\cdot]_+$ vagy minden (t, x)-re $[\cdot]_-$ veendő).

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

A tétel bizonyításához szükségünk van egy egyszerű függvénytani segédtételre:

SEGÉDTÉTEL. Legyen $\psi: R_+ \to R$ folytonos, 0-helyeitől eltekintve folytonosan differenciálható függvény. Tegyük fel, hogy $\lambda = \limsup_{t\to\infty} \psi(t) > 0$, továbbá létezik egy ϱ $(0 < \varrho < \lambda)$ szám, hogy ha $J(\varrho) := \{t > 0: \psi(t) > \varrho\} = \bigcup_{i=1}^{\infty} (a_i, b_i)$ (az (a_i, b_i) intervallumok $(i = 1, 2, \ldots)$ egymástól idegenek), akkor $\lim_{i\to\infty} (b_i - a_i) = 0$.

Ekkor a

$$\varphi_1(t) \equiv [\dot{\psi}(t)]_+, \quad \varphi_2(t) \equiv [\dot{\psi}(t)]_- \quad (t \in J(\varrho))$$

egyenlőtlenségekből következik, hogy az $\int_0^t \varphi_1$, $\int_0^t \varphi_2$ függvények nem egyenletesen folytonosak.

A segédtétel bizonyítása. A

$$J := \left\{t \in J(\varrho): \frac{\lambda + 2\varrho}{3} < \psi(t) < \frac{2\lambda + \varrho}{3}\right\}$$

halmaz nem üres, nem korlátos és nyitott, így léteznek olyan $\{a_i'\}$, $\{b_i'\}$, $\{a_i''\}$, $\{b_i''\}$ sorozatok, hogy

$$a_i' < b_i' < a_i'' < b_i''; \quad \lim_{i\to\infty} (b_i'' - a_i'') = \lim_{i\to\infty} (b_i' - a_i') = 0,$$

$$\int_{a_i'}^{b_i'} \dot{\psi} = \int_{a_i''}^{b_i''} (-\dot{\psi}) = \frac{\lambda - \varrho}{3} > 0 \quad (i = 1, 2, \ldots).$$

Ezek a tulajdonságok biztosítják, hogy φ_1 és φ_2 nem egyenletesen folytonos.

A tétel bizonyítása. A (3.1) rendszer 0-megoldása y'-stabilis, tehát bármely $\varepsilon > 0$, $t_0 \in R_+$-hoz van olyan $\delta(\varepsilon, t_0) > 0$, hogy ha $|x_0| < \delta$, akkor $|y'(t; t_0, x_0)| < \varepsilon$ a $[t_0, \infty)$-en. Legyen $\sigma(t_0) := \delta(H, t_0)$, és tegyük fel, hogy $|x_0| < \sigma(t_0)$. Bebizonyítjuk, hogy $|y(t; t_0, x_0)| \to 0$, ha $t \to \infty$.

Tegyük fel, hogy ez nem igaz. Akkor létezik olyan l természetes szám $(1 \leqq l \leqq k)$ és egy $\psi: R_+ \to R$ függvény úgy, hogy $\lambda := \limsup_{t\to\infty} \psi(t) > 0$ és $\psi(t)$ azonosan megegyezik a $\pm W_l(t, x(t; t_0, x_0))$ függvények valamelyikével. Tekintsük a $J(\lambda/2) := \{t: \psi(t) > \lambda/2\}$ halmazt és a

$$\varphi(t) := \max \{\dot{W}_l(t, y', z'): e^{-1}(\lambda/2) \leqq |y'| \leqq H, \ |z'| \leqq L\}$$

függvényt. A (3) feltétel szerint $|y'(t; t_0, x_0)| \geqq e^{-1}(\lambda/2)$ $(t \in J(\lambda/2))$, ezért $\varphi(t)$ majorálja a $[\dot{\psi}(t)]_+$, $[\dot{\psi}(t)]_-$ függvények valamelyikét a $J(\lambda/2)$ halmazon. Másrészt, a (2) feltétel miatt a

$$\gamma(t) = \max \{\dot{V}(t, y', z'): e^{-1}(\lambda/2) \leqq |y'| \leqq H, \ |z| \leqq L\}$$

integrálisan negatív, és $\dot{V}(t, x(t; t_0, x_0)) \leqq \gamma(t)$ a $J(\lambda/2)$ halmazon, de $V(t, x(t; t_0, x_0))$ alulról korlátos, ezért a segédtétel feltételei teljesülnek, ami ellentmond (4)-nek.

A $W(t, x) = |y|$, illetve a $W(t, x) = |y|^2$ választással adódik az

5.1. KÖVETKEZMÉNY. Tegyük fel, hogy létezik egy y-ban pozitív definit $V: R_+ \times \Gamma \to R$ függvény, amelynek (3.2)-re vonatkozó deriváltja y-ban integrálisan negatív definit. Ha minden α $(0 < \alpha < H)$ számra vagy

$$\int_0^t \max \{[Y_l(s, y, z)]_{+(-)}: \alpha \leqq |y| \leqq H, \ |z| \leqq L\} ds$$

egyenletesen folytonos $(l = 1, 2, \ldots, k)$, vagy

$$\int_0^t \max \{[(y, Y(s, y, z))]_{+(-)}: \alpha \leqq |y| \leqq H, \ |z| \leqq L\} ds$$

egyenletesen folytonos a $[0, \infty)$ intervallumon, akkor (3.2) 0-megoldása aszimptotikusan y-stabilis.

Ez a következmény *Maracskov tételét* speciális esetként tartalmazza, de annak egy lényeges továbbfejlesztését is adja. $Y(t, y, z)$ a mozgás sebességének az R^p-re való vetülete a t időpillanatban, ha a rendszer ekkor éppen az (y, z) állapotban van. Bontsuk fel ezt a vektort az y-nal párhuzamos Y_r radiális, és egy arra merőleges Y_f összetevőre. Nyilvánvaló sejtés, hogy az aszimptotikus y-stabilitás szempontjából az Y_r-nek a szerepe a döntő, hiszen ez adja az R^p-ben az origóhoz való közeledés sebességét, míg Y_f az origó körüli „forgás" sebességét mutatja. *Maracskov tételében* a két összetevő mégis ugyanúgy van korlátozva, az 5.1. következményben viszont már sejtésünknek megfelelően Y_f nagyságára vonatkozóan nincs megszorítás.

Alkalmazzuk most az 5.1. tételt a változó fonalhosszúságú, változó súrlódási együtthatóval bíró közegben mozgó ingával kapcsolatban felvetett kérdések megválaszolására.

Keressünk *Ljapunov-függvényt* (5.2)-höz

$$V = 2[Cb(t) + Ba(t)](1 - \cos x_1) + 2Bx_1x_2 + Cx_2^2 \quad (0 < B, \ C = \text{áll.})$$

alakban. Ekkor

$$\dot{V} = [2Bb + B\dot{a} + C\dot{b}]x_1^2 + 2(B - Ca)x_2^2 + o(|x|^2).$$

Legyen továbbá

$$W = \frac{x_2^2}{b(t)} + 1 - \cos x_1; \quad \dot{W} = -\frac{x_2^2}{b}\left[2a + \frac{\dot{b}}{b}\right].$$

Bizonyítsuk be, hogy B és C megválaszthatók úgy, hogy az 5.1. tétel feltételei az $y = x$ esetre teljesülnek. Ha megköveteljük, hogy

$$0 < b_0 \leqq b(t) \leqq B_0, \quad 2a + \dot{b}(t)/b(t) \geqq 0 \quad (t \geqq 0), \tag{5.3}$$

akkor a 3.1. tétel szerint (5.2) 0-megoldása stabilis, és az 5.1. tétel (3), (4) feltétele is teljesül. A (2) feltétel teljesüléséhez elegendő olyan $m > 0$ szám és $\varphi(t)$ integ-

rálisan pozitív függvény létezése, hogy

$$B(\dot{a}-2b)+C\dot{b} \leqq -\varphi, \quad B-Ca \leqq -m,$$

ez pedig teljesül alkalmas B, C konstansokkal, ha

$$a(t) \geqq a_0 > 0 \quad (t \geqq 0), \quad \dot{b}-\frac{a_0}{2}(2b-\dot{a}) \leqq -\varphi_1 \tag{5.4}$$

(φ_1 integrálisan pozitív), és ez egyben (1) teljesülését is maga után vonja. Ha az (5.3), (5.4) feltételekben az $a(t)=2\dot{l}(t)/l(t)+h(t)$, $b(t)=g/l(t)$ helyettesítéseket végrehajtjuk, akkor a következő eredményt kapjuk:

Ha

(1) $l_0 \leqq l(t) \leqq L_0$;

(2) $2\dot{l}(t)/l(t)+h(t) \geqq a_0$;

(3) $f(t) := 2\dfrac{g-\ddot{l}(t)}{l(t)}+2\left(\dfrac{\dot{l}(t)}{l(t)}\right)^2-\dot{h}(t) > 0 \quad (t \geqq 0)$,

(4) $g\dfrac{\dot{l}(t)}{l^2(t)}+\dfrac{a_0}{2}f(t)$ *integrálisan negatív,*

akkor az (5.1) *inga* $x=\dot{x}=0$ *egyensúlyi helyzete aszimptotikusan stabilis. Speciálisan, ha az inga hossza nem változik* ($l(t)=l_0$), *akkor a feltételek:* $h(t) \geqq a_0 > 0$, $\dot{h} \leqq b_0 < 2g/l_0$ ($t \in R_+$).

Az (5.2) egyenlet 0-megoldásának stabilitásával számos dolgozat foglalkozik (l. [2, 6] irodalomjegyzékét). Ha kihasználjuk az egyenlet speciális tulajdonságát és a megoldások oszcillációs tulajdonságait, sok érdekes feltétel kapható, amely az általános rendszerekre vonatkozó tételekből közvetlenül nem adódik. Ezekkel a vizsgálatokkal itt nem foglalkozunk, de illusztrálásképpen megfogalmazzuk a [16] dolgozat eredményeinek az (5.1) ingára vonatkozó két következményét:

Ha $l_0 \leqq l(t) \leqq L_0$, h *korlátos,* $\int_0^\infty [2h+3\dot{l}/l]_- < \infty$ *és* $[2h+3\dot{l}/l]_+$ *integrálisan pozitív, akkor az* (5.1) *inga* $x=\dot{x}=0$ *egyensúlyi helyzete aszimptotikusan stabilis.*

Ha $l(t) \to 0$ $(t \to \infty)$, *továbbá* $0 < k \leqq 3\dot{l}/\sqrt{l}+h\sqrt{l} < K$, *akkor az* (5.1) *inga* $x=\dot{x}=0$ *egyensúlyi helyzete aszimptotikusan x-stabilis.*

A következő fejezetben feltételt fogunk adni arra is, hogy $\dot{x}(t) \to 0$ $(t \to \infty)$.

Alkalmazzuk most az 5.1. tételt a disszipatív és giroszkopikus erőket is tartalmazó (3.5) mechanikai rendszer tanulmányozására. Mint ismeretes [45—47], ilyen egyenlet írja le a mozgó alapra helyezett giroszkópokat tartalmazó rendszerek mozgását, amelyeket széles körben alkalmaznak a műszaki gyakorlatban. A 3.6. tételben megadtuk a $q=\dot{q}=0$ egyensúlyi helyzet aszimptotikus $\dot{q}$-stabilitásának feltételét abban az esetben, amikor a rendszerre potenciális erő nem hat.

A (3.5) rendszert írjuk a q, p *Hamilton-féle változókkal* a

$$\dot{q} = \frac{\partial H}{\partial p}, \quad \dot{p} = -\frac{\partial H}{\partial q}+[G(t, q)-B(t, q)]\frac{\partial H}{\partial p} \tag{5.5}$$

kanonikus alakba, ahol $H\colon (p, A^{-1}(q)p)/2+P(q)$, $\partial H/\partial q:=(\partial H/\partial q_1, \ldots, \partial H/\partial q_n)$. Alkalmazzuk az 5.1. tételt az $x=(p, q)\in R^{2n}$, $y'=y=p\in R^n$ particionálással, miközben a segédfüggvények $V(p, q)=H(p, q)$, $W(p)=|p|^2$, amelyeknek (5.5)-re vonatkozó deriváltjai:

$$\dot{V}=-2R, \quad \dot{W}=-2\left(\frac{\partial H}{\partial q}, p\right)+([(G-B)A^{-1}-A^{-1}(G+B)]p, p).$$

Tegyük fel, hogy a $q=p=0$ egy környezetéből induló mozgások során $|q(t)|\leqq L$ ($t\geqq 0$) teljesül valamely L számmal (ez a helyzet pl. akkor, ha q lehetséges értékeinek halmaza R^n-ben nem korlátos ugyan, de $P(q)\to\infty$ ($|q|\to\infty$). Ha *a disszipáció integrálisan teljes,* vagyis van olyan $\beta\colon R_+\to R_+$ integrálisan pozitív függvény, hogy $R(t, q, \dot{q})\geqq\beta(t)|\dot{q}|^2$, akkor $\dot{V}=-\beta(t)|A^{-1}p|^2\leqq -c\beta(t)|p|^2$ ($0<c=$áll.), vagyis $\dot{V}$ p-ben integrálisan negatív definit. Jelölje $\lambda(t, q)$, ill. $\Lambda(t, q)$ a $(G-B)A^{-1}-A^{-1}(G+B)$ szimmetrikus mátrix legkisebb, illetve legnagyobb sajátértékét. Az 5.1. tétel alkalmazásával a következő eredményt kapjuk:

5.2. TÉTEL. Tegyük fel, hogy $P(q)\geqq 0$ és létezik olyan $\delta>0$, L szám, hogy ha $|(p_0, q_0)|<\delta$, akkor $|q(t; 0, q_0, p_0)|<L$.

Ha a disszipáció integrálisan teljes, és az

$$\int_0^t \max\{[\lambda(s, q)]_-\colon |q|\leqq L\}ds, \quad \int_0^t \max\{[\Lambda(s, q)]_+\colon |q|\leqq L\}ds \tag{5.6}$$

függvények valamelyike egyenletesen folytonos a $[0, \infty)$ intervallumon, akkor a (3.5) rendszer $q=\dot{q}=0$ egyensúlyi helyzete aszimptotikusan $\dot{q}$-stabilis.

Ha A felcserélhető B-vel és G-vel, akkor $\lambda(t, q)$ egybeesik $B(t, q)$ legkisebb sajátértékével, amely nem-negatív, hiszen $(B(t, q)\dot{q}, \dot{q})\geqq 0$. Tehát ekkor $[\lambda(t, q)]_-\equiv 0$, így a tétel utolsó feltétele nyilvánvalóan teljesül.

6. Attraktivitási tételek

Az aszimptotikus stabilitást biztosító 3.2. tétel feltételeit kielégítő *V Ljapunov-függvényt* konstruálni sokszor igen nehéz, amit a mechanikai rendszerek tanulmányozásánál már tapasztaltunk. Ezért fontos kérdés az, hogy a tétel feltételeit hogyan lehet gyengíteni, illetve ha csak olyan *Ljapunov-függvény* van a birtokunkban, amely nem teljesít minden kívánt feltételt, akkor mit tudunk állítani a megoldásokról. Ebben a fejezetben elsősorban azt vizsgáljuk, hogy $\dot{V}$ negatív definit voltát hogyan lehet gyengíteni, esetleg valamilyen más feltétellel helyettesíteni, illetve mit lehet a megoldásokról állítani negatív szemidefinit deriválttal bíró *Ljapunov-függvény* birtokában

Az ilyen irányú vizsgálatokat két csoportba lehet osztani. Az elsőbe tartoznak azok az eredmények, amelyekben a (3.1) rendszer jobb oldalára tett megszorításokkal pótolják V hiányzó tulajdonságait, hasonlóan az előző fejezet tételeihez [22, 23]. A második csoportba tartozó eredményeket az jellemzi, hogy $\dot{V}$ negatív definit volta helyett $\dot{V}$ és (3.1) jobb oldalának normája közötti egyenlőtlenséget tételeznek fel [5, 12]. Mindkét csoportot jellemzi, hogy aszimptotikus stabilitás helyett az R^n állapottér egy bizonyos *zárt részhalmaz*ának attraktivitását vizsgálja. A $H\subset R^n$ zárt halmazt (3.1)-re vonatkozóan *attraktívnak* nevezzük, ha H egy környezetéből

induló megoldások H-hoz tartanak, ha $t\to\infty$. Ahhoz, hogy ezeket az eredményeket áttekinthessük, szükségünk van a megoldások limeszhalmazának fogalmára és legegyszerűbb tulajdonságaira.

Tekintsük (3.1)-nek egy $x(t;\ t_0, x_0)$ megoldását. A $p\in\bar{\Gamma}$ pontot az x megoldás *pozitív limeszpontjának* nevezzük, ha létezik olyan $\{t_k\}$ sorozat, amelyre $t_k\to\infty$ és $x(t_k)\to p$ $(k\to\infty)$. Az x megoldás *pozitív limeszhalmazának* az x pozitív limeszpontjaiból álló halmazt nevezzük, és Ω-val jelöljük. A $\gamma(x):=\bigcup_{t\geqq t_0}x(t)$ halmazt az x megoldáshoz tartozó *trajektóriának* nevezzük.

6.1. LEMMA. *a)* Ω zárt halmaz.

b) Ha $\gamma(x)$ korlátos, akkor Ω nem-üres, kompakt, összefüggő halmaz.

c) Ha Ω korlátos, akkor $x(t)\to\Omega$ $(t\to\infty)$.

d) $x(t)\to\Omega_\infty(t\to\infty)$, ami azt jelenti, hogy bármely $\varepsilon_1, \varepsilon_2>0$ számhoz van olyan $T(\varepsilon_1, \varepsilon_2)>0$, hogy ha $t>T$, akkor $x(t)\in S(\Omega, \varepsilon_1)\cup(\overline{G(0, 1/\varepsilon_2)})^c$.

Bizonyítás. a) Egyszerű látni, hogy $p\in\bar{\Gamma}$ akkor és csakis akkor pozitív limeszpontja x-nek, ha bármely $\varepsilon>0, T\in R_+$ számokhoz létezik olyan $t\geqq T$, hogy $|x(t)-p|<\varepsilon$, amiből az állítás közvetlenül adódik.

b) Ha $\gamma(x)$ korlátos, akkor $\overline{\gamma(x)}$ kompakt, és így Ω nem üres. De Ω egy kompakt halmaz zárt részhalmaza, így kompakt. Ha nem lenne összefüggő, akkor felbomlana két, egymástól idegen kompakt halmaz egyesítésére. Legyen ezek távolsága $\delta>0$. Ekkor létezik olyan $\{t_k^*\}$ sorozat, amelyre $t_k^*\to\infty$, $x(t_k^*)$ mindkét összetevőtől $\delta/4$-nél távolabb van, és $x(t_k^*)$ konvergens, ha $k\to\infty$. De ez lehetetlen, hiszen a határpontnak hozzá kellene tartoznia valamelyik összetevőhöz, ugyanakkor mindkettőtől legalább $\delta/4$ távolságra van.

c) Ha az állítás nem igaz, akkor van olyan $\varepsilon>0$ és $t_k\to\infty$, hogy $\varrho(\Omega, x(t_k))\geqq\varepsilon$. De ekkor létezik x-nek olyan pozitív limeszpontja, amely nem tartozik hozzá Ω-hoz, és ez ellentmondás.

d) Az előző *c)* állításhoz hasonlóan bizonyítható.

6.2. LEMMA. Ha (3.1) autonóm, akkor $\Omega\cap\Gamma$ invariáns, ami azt jelenti, hogy bármely $p\in\Omega\cap\Gamma$ pontból kiinduló megoldás trajektóriáját tartalmazza.

Bizonyítás. Legyen $t_k\to\infty$ olyan sorozat, hogy $x(t_k)\to p$ $(k\to\infty)$. Mivel (3.1) autonóm, $x(t+(t_k-t_0))$ megoldása (3.1)-nek. Tetszőleges $T\in R_+$ esetén

$$x(T+(t_k-t_0)) = x(T;\ t_0, x(t_k))\to x(T;\ t_0, p)\quad (k\to\infty),$$

mivel a megoldások a kezdeti értékektől folytonosan függnek. De ez azt jelenti, hogy $x(T;\ t_0, p)\in\Omega$.

A megoldások limeszhalmazának ez az invarianciája az alapja E. A. BARBASIN és N. N. KRASZOVSZKIJ [34, 39] híres, sokat alkalmazott tételének, amely *autonóm* differenciálegyenletre vonatkozik és szemidefinit deriválttal rendelkező *Ljapunov-függvény* segítségével aszimptotikus stabilitást állapít meg.

6.1. TÉTEL. Ha (3.1) autonóm, és létezik olyan pozitív definit V függvény, amelyre $\dot{V}(x)\leqq 0$, és az $F:=\{x\in\Gamma:\dot{V}(x)=0\}$ halmaz nem tartalmaz egyetlen trajektóriát sem az $x=0$ ponton kívül, akkor (3.1) 0-megoldása aszimptotikusan stabilis.

Bizonyítás. A stabilitás nyilvánvalóan következik a 3.1. tételből. Meg fogjuk mutatni, hogy a $\{0\}$ halmaz attraktív is.

$\dot{V}(x) \leqq 0$, így tetszőleges $x(t)$ megoldásra $V(x(t)) \searrow m \geqq 0$ $(t \to \infty)$. Ha $|x(0)|$ elég kicsi, akkor Ω nem üres és nyilván $\Omega \subset N(m) := \{x \in \bar{\Gamma}: V(x) = m\}$. De a 6.2. lemma miatt Ω tetszőleges pontjából induló megoldás trajektóriája $N(m)$-ben van, tehát V a trajektóriák mentén állandó, azaz $\Omega \subset F$. Mivel F nem tartalmaz teljes trajektóriát az $x=0$ ponton kívül, $\Omega = \{0\}$, így a 6.1. lemma *c)* állítása szerint $x(t) \to 0$ $(t \to \infty)$.

Alkalmazzuk ezt a tételt a (3.5) mechanikai rendszerre a stacionárius esetben, tehát amikor R és G nem függnek az időtől.

6.2. TÉTEL. Ha a P potenciális energiának a $q=0$ pontban minimuma van, a $q=0$ egyensúlyi helyzet izolált, és a disszipáció teljes $(R(q, \dot{q}) \geqq \beta |\dot{q}|^2, \beta > 0)$, akkor a (3.5) rendszer $q = \dot{q} = 0$ egyensúlyi helyzete aszimptotikusan stabilis, feltéve, hogy a rendszer stacionárius.

Bizonyítás. Tekintsük a (3.5) rendszerrel ekvivalens (5.5) rendszert, és legyen $V(q, p) := H(q, p)$ a teljes energia. Mivel $\dot{V}(q, p) = -2R$, és a disszipáció teljes, $F = \{(q, p) \in R^{2n}: p = 0\}$. Mivel a $q=0$ egyensúlyi helyzet izolált, ezért az F halmaz nem tartalmaz a $q=0$ ponton kívül trajektóriát, hiszen az csak $q = q_0$ típusú lehetne. Másrészt, V pozitív definit, hiszen a T kinetikai energia pozitív p-definit, és P-nek a $q=0$ pontban izolált minimuma van. Tehát a 6.1. tétel szerint a $q = p = 0$ egyensúlyi helyzet aszimptotikusan stabilis.

Ennek a tételnek nem-stacionárius esetre való általánosítása nehéz probléma. Ugyanis a 6.1. tétel nem-autonóm esetben nem igaz, amint a következő egyszerű példa mutatja. Tekintsük a skaláris $\dot{x} = -a(t)x$ egyenletet $\left(a(t) > 0, \int_0^\infty a < \infty\right)$ és a $V(x) = x^2/2$ függvényt. $\dot{V}(t, x) = -a(t)x^2$, az F halmaznak tehát a $\{0\}$ halmaz felel meg, az $x=0$ mégsem aszimptotikusan stabilis. Hasonló jellegű feltételre azonban nagy szükségünk lenne nem-autonóm rendszerekre is. A tétel bizonyítása azt sugallja, hogy először az $\Omega \subset F$ tartalmazást kellene biztosítani valamilyen alkalmas $F \subset \bar{\Gamma}$ halmazzal, vagyis lokalizálni a megoldások limeszhalmazát, majd valamilyen, az F halmazon működő feltétellel biztosítani, hogy $\Omega = \{0\}$ is teljesüljön.

Alábbi tételeink megfogalmazásához szükségünk lesz az R^n egy új környezetrendszerére, amelyet egy adott $W: R^n \to R^k$ folytonos függvény határoz meg a következőképpen:

$$\varrho^*(x, y) := \varrho(W(x), W(y)) \quad (x, y \in R^n),$$

$$\varrho^*(H, K) := \inf\{\varrho^*(x, y): x \in H, y \in K\} \quad (H, K \subset R^n),$$

$$S^*(H, \varepsilon) := \{x \in R^n: \varrho^*(x, H) < \varepsilon\} \quad (\varepsilon > 0).$$

Az egyszerűség kedvéért tegyük fel a fejezet további részében, hogy $\Gamma = R^n$ (az általános esetre vonatkozó eredményeket l. [18]-ban).

6.3. LEMMA. Legyen $x(t)$ (3.1)-nek megoldása, és legyen $M \subset R^n$ olyan halmaz, amely $x(t)$ trajektóriáját tartalmazza. Legyenek $V: R_+ \times R^n \to R$, $W: R^n \to R^k$ folytonosan differenciálható függvények.

Tegyük fel, hogy a $p \in R^n$ ponthoz van olyan δ, $\varrho > 0$ és T szám, és $\eta: R_+ \to R_+$, R_+-on integrálható függvény, amelyekre teljesül:

(1) $V(t, x)$ alulról korlátos és
(2) $\dot{V}(t, x) \leqq -\delta |W(t, x)| + \eta(t)$
a $t \geqq T$, $x \in S^*(p, \varrho) \cap M$ halmazon.
Ekkor vagy $p \notin \Omega$ vagy $\Omega \subset W^{-1}[W(p)]$.

Bizonyítás. Tegyük fel, hogy az állítás nem igaz. Akkor létezik olyan $q \in \Omega$ és σ $(0 < \sigma < \varrho/\sqrt{k})$, hogy $q \notin S^*(p, \sigma\sqrt{k})$. Mivel $p, q \in \Omega$, és W folytonos, létezik olyan l $(1 \leqq l \leqq k)$ természetes szám és $\{t'_m\}$, $\{t''_m\}$ sorozat, amelyre:

$$T \leqq t'_1 < t''_1 < \ldots < t'_m < t''_m < \ldots; \quad \lim_{m \to \infty} t'_m = \infty; \tag{6.1}$$

$$|W(p) - W(x(t))| < \sigma\sqrt{k} \quad (t'_m \leqq t \leqq t''_m); \tag{6.2}$$

$$|W_l(x(t''_m)) - W_l(x(t'_m))| = \frac{\sigma}{2} \quad (m = 1, 2, \ldots). \tag{6.3}$$

De ekkor a (2) feltételből a $v(t) := V(t, x(t))$ függvényre a

$$v(t''_m) - v(t'_m) \leqq -\delta \frac{\sigma}{2} + \int_{t'_m}^{t''_m} \eta(t)\, dt \quad (m = 1, 2, \ldots)$$

becslést kapjuk, amelyből

$$v(t''_m) \leqq v(t'_1) - m\delta \frac{\sigma}{2} + \int_{T}^{t''_m} \eta(t)\, dt \to -\infty \quad (m \to \infty)$$

adódik, és ez ellentmond annak, hogy V alulról korlátos.

6.1. Megjegyzés. Ha W skalárértékű függvény $(k = 1)$, akkor a (2) feltételt elegendő a $[\dot{W}]_+$ függvényre megkövetelni $\dot{W}$ helyett.

6.3. TÉTEL. Legyen $H, M \subset R^n$, és tegyük fel, hogy bármely $p \in H^c$ ponthoz léteznek olyan $\varrho(p), \delta(p) > 0, T(p) \geqq 0$ számok és $\eta: R_+ \to R_+$, R_+-on integrálható függvény úgy, hogy a 6.3. lemma (1)—(2) feltétele teljesül a $t \geqq T(p)$, $x \in S^*(p, \varrho(p)) \cap M$ halmazon.

Ha $x(t)$ olyan megoldása (3.1)-nek, amelynek trajektóriája M-ben fekszik, akkor vagy $\Omega \subset H$, vagy létezik olyan $d \in R^k$, hogy $\Omega \subset W^{-1}[d]$.

Ha W skalárértékű függvény $(k = 1)$, akkor a 6.3. lemma (2) feltételét elegendő a $[\dot{W}]_+$ függvényre megkövetelni.

Bizonyítás. Ha Ω üres, akkor $\Omega \subset H$ teljesül. Tegyük fel, hogy Ω nem üres, és létezik olyan $p \in \Omega$, hogy $p \in H^c$. Ekkor a 6.3. lemma szerint $\Omega \subset W^{-1}[W(p)]$, tehát az állítás igaz.

Érdemes megjegyezni, hogy a tétel állítását megfogalmazhatjuk a következő módon is: vagy $x(t) \to H_\infty$, vagy létezik olyan $d \in R^k$, hogy $x(t) \to W^{-1}[d]_\infty$, ha $t \to \infty$ (l. 6.1. lemma, *d)* állítás).

Ez a tétel alkalmas szigorú értelemben vett stabilitási tulajdonságok tanulmányozásán kívül a megoldások más jellegű aszimptotikus viselkedésének megadására is.

Ha például $W(x)=W(y,z)=y$, $H\subset R^n$ zárt, és $M:=\{(y,z): |z|\leqq L\}$, akkor a következő eredményt kapjuk:

6.1. KÖVETKEZMÉNY. Tegyük fel, hogy bármely $\varepsilon>0$ számhoz és $K\subset R^n$ kompakt halmazhoz létezik $\delta>0$ $T\geqq 0$ szám és $\eta: R_+\to R_+$ R_+-on integrálható függvény a következő tulajdonságokkal: V alulról korlátos, és $\dot V(t,x)\leqq \leqq -\delta|Y(t,x)|+\eta(t)$ a $t\geqq T$, $x\in K\cap S^c(H,\varepsilon)$ halmazon. Ekkor (3.2) bármely z-korlátos $x(t)=(y(t),z(t))$ megoldására vagy $x(t)\to H_\infty$, vagy $y(t)$-nek létezik a véges vagy végtelen határértéke, ha $t\to\infty$. Speciálisan, ha $H:=\{(y,z): y=0\}$, akkor minden z-korlátos $(y(t),z(t))$ megoldásra $y(t)$-nek létezik a véges vagy végtelen határértéke, ha $t\to\infty$.

Most egy olyan tételt adunk adott halmaz attraktivitására, amelyben $\dot V$-tól kevesebbet követelünk, viszont megkívánjuk a (3.1) rendszer jobb oldalának egy, a V-től független regularitási tulajdonságát.

6.4. LEMMA. Legyen $x(t)$ (3.1)-nek megoldása és legyen $M\subset R^n$ olyan tartomány (összefüggő nyitott halmaz), amely tartalmazza $x(t)$ trajektóriáját.

Tegyük fel, hogy a $p\in R^n$ ponthoz létezik olyan $\varrho>0$, $T\geqq 0$ szám, hogy tetszőleges $u:[T,\infty)\to L:=S^*(p,\varrho)\cap M$ folytonos függvényre teljesülnek a következő feltételek:

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

(2) $\dot V(t,u(t))$ integrálisan negatív;

(3) $V(t,u(t))$ alulról korlátos

a $[T,\infty)$ intervallumon.

Ekkor $p\notin\Omega$.

6.5. LEMMA. A 6.4. Lemma állítása érvényben marad, ha az (1)—(2) feltételeket a következőkkel helyettesítjük: bármely $u:[T,\infty)\to L$ folytonos függvényre

$$(1')\quad \left|\int_T^\infty (|\dot W_1(t,u(t))|,\ldots,|\dot W_k(t,u(t))|)\,dt\right|<\infty;$$

$$(2')\quad \int_T^\infty \dot V(t,u(t))\,dt=-\infty.$$

A 6.4. és 6.5. lemma bizonyítása. Tegyük fel, hogy $p\in\Omega$, és legyen $\{t_m\}$ olyan sorozat, amelyre $t_m\to\infty$ és $x(t_m)\to p$ teljesül, ha $m\to\infty$. Ha $T<T^*<\omega$, és $x(t)\in S^*(p,\varrho)$ $(T^*\leqq t\leqq\omega)$, akkor (2), illetve (2') következtében

$$V(\omega,x(\omega))\leqq V(T^*,x(T^*))+\int_T^\omega \dot V(t,x(t))\,dt\to-\infty\quad(\omega\to\infty),$$

tehát (3) miatt ω nem lehet akármilyen nagy. Vagyis a megoldás trajektóriája t bármilyen nagy értékeire tetszőlegesen közel kerül p-hez, de nem maradhat a p pont $S^*(p,\varrho)$ környezetében egyetlen $[T^*,\infty)$ intervallumon sem. Ezért ugyanúgy, mint a 6.3. lemma bizonyításában, létezik olyan $\sigma>0$, l $(1\leqq l\leqq k)$ szám és $\{t_m'\}$, $\{t_m''\}$

sorozat, amely rendelkezik a (6.1)—(6.3) tulajdonságokkal. Ekkor

$$\left| \int_{t'_m}^{t''_m} \dot{W}(t, x(t))\,dt \right| \geqq \frac{\sigma}{2} \quad (m = 1, 2, \ldots) \tag{6.4}$$

adódik. Ez (1')-vel együtt nem teljesülhet, így a 6.5. lemma már bizonyítva van.

Megmutatjuk, hogy (6.4) az (1)—(3) feltételekkel is ellentmondásban van. (6.4) és (1) egyidejűleg csak úgy állhat, ha $t''_m - t'_m \geqq \delta > 0$ $(m = 1, 2, \ldots)$. De $x(t) \in S^*(p, \varrho)$, ha $t \in (t'_m, t''_m)$, ezért (2) miatt

$$V(t''_m, x(t''_m)) \leqq V(t_1, x(t_1)) + \sum_{i=1}^{m} \int_{t'_i}^{t''_i} \dot{V}(t, x(t))\,dt \to -\infty \quad (m \to \infty),$$

ami ellentmond (3)-nak.

Itt is érvényes, amit a 6.1. megjegyzésben mondtunk: ha W skalárértékű $(k = 1)$, akkor az (1), illetve (1') feltételt elegendő a $[\dot{W}]_+$ függvényre megkövetelni, és az állítás változatlanul érvényben marad.

Ezen lokalizációs lemmák alapján megadhatjuk a *Barbasin—Kraszovszkij-tétel* két általánosítását nem-autonóm differenciálegyenlet-rendszerre.

6.4. TÉTEL. Tegyük fel, hogy a (3.1) rendszerhez létezik olyan $V: R_+ \times R^n \to R$, $W: R^n \to R^k$ folytonosan differenciálható függvény és $M \subset R^n$ tartomány, amely eleget tesz a következő tulajdonságoknak:

(1) $\dot{V}(t, x) \leqq \varphi(t) U(x) + \eta(t)$, ahol $\varphi: R_+ \to R_+$, $U: R^n \to R_-$, $\eta: R_+ \to R_+$ folytonos függvények, η R_+-on integrálható;

(2) ha F jelöli U 0-helyeinek halmazát, akkor bármely $p \in F^c$ ponthoz létezik olyan $\varrho(p) > 0$, $T(p) \geqq 0$ szám, hogy

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

(3) φ integrálisan pozitív;

bármely $u: [T(p), \infty) \to L(p)$ folytonos függvényre

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

(5) $V(t, u(t))$ alulról korlátos

a $[T(p), \infty)$ intervallumon.

Ha $x(t)$ olyan megoldása (3.1)-nek, amelynek trajektóriáját M tartalmazza, akkor $\Omega \subset F$, tehát $x(t) \to F_\infty$, ha $t \to \infty$.

6.5. TÉTEL. A 6.3. tételben a (3)—(4) feltétel helyett a

$$\int_0^\infty \varphi(t)\,dt = \infty; \tag{3'}$$

$$\left| \int_T^\infty (|\dot{W}_1(t, u(t))|, \ldots, |\dot{W}_k(t, u(t))|)\,dt \right| < \infty \tag{4'}$$

feltételeket is megkövetelhetjük, az állítás változatlan marad.

A 6.4. és 6.5. tétel bizonyítása. Tegyük fel, hogy az állítás nem igaz, és legyen $p\in F^c\cap\Omega$. Akkor a feltételek szerint $\varrho(p), \delta(p)>0$, $T(p)\geqq 0$, hogy

$$\dot{V}(t, x)\leqq -\delta\varphi(t)+\eta(t) \quad (t\geqq T(p),\ x\in L(p)).$$

De ekkor a 6.4., illetve 6.5. lemma szerint $p\notin\Omega$, ami ellentmondás.

6.2. *Megjegyzés.* Ha W skalárértékű függvény ($k=1$), akkor a (4), illetve (4') feltételeket elegendő a $[\dot{W}]_+$ függvényre megkövetelni.

6.1. PÉLDA. Alkalmazzuk a 6.3—6.5. tételeket a gyakorlatban is sokszor előforduló

$$\begin{aligned}\dot{x}_1 &= -r(t)x_1+q(t)x_2\\ \dot{x}_2 &= -q(t)x_1-p(t)x_2\end{aligned} \qquad (x_1, x_2\in R) \tag{6.5}$$

rendszer tanulmányozására, ahol $p, q, r: R_+\to R$ folytonos függvények; $p(t)\geqq 0$, $r(t)\geqq 0$ ($t\in R_+$). Legyen $V(x_1, x_2):=(x_1^2+x_2^2)/2$, amelynek (6.5)-re vonatkozó deriváltja $\dot{V}(t, x_1, x_2)=-r(t)x_1^2-p(t)x_2^2$ nem-pozitív, és így (6.5) minden megoldása létezik és korlátos R_+-on. Tekintsük továbbá a $W(x_2)=x_2^2$ segédfüggvényt, amelyre $\dot{W}=-q(t)x_1x_2-p(t)x_2^2$, és jelöljük az (x_1, x_2) sík x_1-tengelyen levő pontjainak halmazát H-val. Tegyük fel, hogy létezik olyan $\alpha>0$, amellyel

$$|q(t)|<\alpha p(t) \quad (t\in R_+) \tag{6.6}$$

teljesül. Bebizonyítjuk, hogy ebben az esetben a 6.3. tétel feltételei teljesülnek. Tetszőleges $x(t)$ megoldáshoz létezik olyan C konstans, hogy $x(t)\in M:=$ $:=\{x: x_1^2+x_2^2\leqq C\}$ ($t\geqq 0$). Elegendő megmutatni, hogy bármely $\varepsilon>0$-hoz létezik egy $\delta>0$ úgy, hogy $x\in M$, $|x_2|\geqq\varepsilon$ maga után vonja a $\dot{V}(t, x)\leqq -\delta[\dot{W}(t, x)]_+$ egyenlőtlenség teljesülését ($t\geqq 0$). Legyen $\delta(\varepsilon):=2\varepsilon^2/(\alpha C)$. Ekkor (6.6) miatt $-p(t)x_2^2\leqq -\delta|q(t)|(x_1^2+x_2^2)/2$ ($t\in R_+$), amely a kívánt egyenlőtlenséget szolgáltatja. Felhasználva azt a tényt, hogy $\lim_{t\to\infty} V(t, x(t))$ létezik (Ω-nak V egy nívóhalmazán kell feküdnie), a 6.3. tételből adódik a következő eredmény:

1) *Tegyük fel, hogy* (6.6) *teljesül. Ekkor* (6.5) *bármely megoldásának mindkét komponense véges határértékhez tart, ha* $t\to\infty$. *Ha még* $\int_0^\infty p=\infty$ *is teljesül, akkor* $x_2(t)\to 0$, *ha* $t\to\infty$.

A (6.6) feltétel túl szigorúnak látszik ebben az állításban, hiszen *lokálisan* sokat követel q-tól (pl. ha $p(t_0)=0$, akkor $q(t_0)=0$), ugyanakkor az állítás a megoldások *határértékére* vonatkozik, ha $t\to\infty$. A 6.4. és 6.5. tételek segítségével ez a feltétel mellőzhető. Éspedig legyen $\varphi(t):=p(t)$, $U(x):=x_2^2$. Az $s\to[s]_+$ függvény monoton növő R-en, tehát $[\dot{W}]_+\leqq[q(t)x_1x_2]_+$, így az alábbi állításokat kapjuk:

2) *Ha* p *integrálisan pozitív, és* $\int_0^t |q(s)|\,ds$ *egyenletesen folytonos* R_+*-on, akkor* (6.5) *bármely megoldására* $x_1(t)\to$konst., $x_2(t)\to 0$ ($t\to\infty$).

3) *Ha* $\int_0^\infty p=\infty$ *és* $\int_0^\infty |q|<\infty$, *akkor* (6.5) *bármely megoldására* $x_1(t)\to$konst., $x_2(t)\to 0$ ($t\to\infty$).

Ha a *Barbasin—Kraszovszkij-tétel* olyan kiterjesztését akarjuk levezetni nem-autonóm rendszerekre, amely a 0-megoldás aszimptotikus stabilitását, illetve parciális aszimptotikus stabilitását állítja, a fenti tételek feltételeit olyanokkal kell kiegészíteni, amelyek biztosítják, hogy az Ω limeszhalmaz a $\{0\}$ halmaz legyen, illetve az $\{(y, z): y=0\}$ altérben feküdjön. Hála a fenti lokalizációs tételeknek, ezeknek a feltételeknek már csak az F halmaz környezetében kell teljesülniök, hiszen tudjuk, hogy $\Omega \subset F$.

6.1. DEFINÍCIÓ. Azt mondjuk, hogy a folytonos $A: R_+ \times R^n \to R$ függvény *y-szigorúan nem tűnik el az* $M \subset R^n$ *halmazon* (jelben: $A \neq 0$ y-szigorúan az M halmazon), ha bármely $\alpha_1 < \alpha_2$ $(0 < \alpha_1 < \alpha_2)$ számhoz van olyan $\beta > 0$ szám és $\psi: R_+ \to (0, \infty)$ $\left(\int_0^\infty \psi = \infty\right)$ függvény, hogy

$$\psi(t) \leqq \inf \{|A(t, x)|: \alpha_1 \leqq |y| \leqq \alpha_2, \varrho(x, M) \leqq \beta\}.$$

A tétel bizonyításában szükségünk lesz az egyenletes y-stabilitásnál valamivel többet kívánó stabilitási tulajdonságra, amit S-tulajdonságnak nevezünk.

6.2. DEFINÍCIÓ. Azt mondjuk, hogy a (3.2) rendszer 0-megoldása rendelkezik az *S-tulajdonság*gal, ha bármely $\varepsilon > 0$-hoz létezik olyan $\delta > 0$, hogy ha $|y_0| < \delta$, $t_0 \in R_+$, akkor $|y(t; t_0, x_0)| < \varepsilon$ a $[t_0, \infty)$ intervallumon (itt $x_0 = (y_0, z_0)$).

6.6. LEMMA. Ha a (3.2) rendszerhez létezik olyan $V: R_+ \times R^n \to R$ függvény és $a, b \in \mathcal{K}$, hogy $a(|y|) \leqq V(t, x) \leqq b(|y|)$ és $\dot{V}(t, x) \leqq 0$, akkor (3.2) 0-megoldása (S)-tulajdonságú.

Bizonyítás. Tetszőleges $\varepsilon > 0$-hoz legyen $\delta(\varepsilon) := b^{-1}(a(\varepsilon))$. Ha $|y_0| < \delta$, akkor $V(t_0, x_0) < a(\varepsilon)$. De $\dot{V} \leqq 0$, tehát $V(t, x(t; t_0, x_0)) < a(\varepsilon)$, és így $|y(t; t_0, x_0)| < \varepsilon$, ha $t \geqq t_0$.

6.6. TÉTEL. Tegyük fel, hogy (3.2) minden $(y(t), z(t))$ megoldására $z(t)$ korlátos, ha $t \geqq t_0$. Tegyük fel továbbá, hogy léteznek olyan $V: R_+ \times R^n \to R$, $W: R^n \to R^k$, $A: R_+ \times R^n \to R^l$ folytonosan differenciálható függvények, amelyek rendelkeznek a következő tulajdonságokkal:

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

$$a(|y|) \leqq V(t, x) \leqq b(|y|);$$

(2) $\dot{V}(t, x) \leqq \varphi(t) U(x)$, ahol $U: R^n \to R_-$, $\varphi: R_+ \to R_+$ folytonos függvények, φ integrálisan pozitív;

(3) ha F jelöli U 0-helyeinek halmazát, akkor minden $p \in F^c$-nek van olyan $S^*(p, \varrho)$ környezete, hogy $S^*(p, \varrho) \cap F$ üres;

(4) ha $K \subset R^n$ kompakt, akkor $A(t, x)$ korlátos az $R_+ \times K$ halmazon, továbbá $\dot{A}(t, x) \neq 0$ y-szigorúan az $F \cap K$ halmazon;

(5) bármely $u: R_+ \to K$ $(K \subset R^n$, K kompakt) függvényre $\int_0^t \dot{W}(s, u(s))\, ds$ egyenletesen folytonos a $[0, \infty)$ intervallumon.

Akkor (3.2) 0-megoldása aszimptotikusan y-stabilis.

Bizonyítás. Az (1)—(2) feltételből a 6.6. lemma szerint következik a 0-megoldás y-stabilitása, sőt S-tulajdonsága. Elegendő tehát bebizonyítani, hogy minden

$(y(t), z(t))$ megoldásra $y(t)\to 0$ $(t\to\infty)$, vagyis bármely $\varepsilon>0$-hoz van olyan T, hogy $|y(t)|<\varepsilon$, ha $t>T$. Az S-tulajdonság következtében ez biztosan teljesül, ha van olyan T, hogy $|y(T)|<\delta$ (δ-ra nézve lásd a 6.2. definíciót). Vagyis elegendő bebizonyítanunk, hogy $|y(t)|$ akármilyen kicsiny értéket is felvesz.

Ha $|x(t_0)|$ elég kicsiny, akkor $|y(t)|\leqq\alpha_2=$konstans $(t\geqq t_0)$ az y-stabilitás miatt. Mivel a megoldások z-korlátosak, van olyan $K\subset R^n$ kompakt halmaz, amely tartalmazza $x(t)$ trajektóriáját. Legyen $\alpha_1>0$ tetszőleges, és tekintsük az „$\dot{A}(t,x)\neq 0$ y-szigorúan az $F\cap K$ halmazon" tulajdonság definíciójában szereplő $\beta>0$ számot. Legyen $H:=\{x\in R^n\colon \alpha_1\leqq|y|\leqq\alpha_2, \varrho(x, F\cap K)\leqq\beta\}$. Az $x(t)$ megoldás Ω limeszhalmaza nem üres, $\Omega\subset F$ (l. 6.4. tétel), és $x(t)\to\Omega$ $(t\to\infty)$. Tehát létezik olyan t^* hogy $\varrho(x(t), F\cap K)<\beta$ a $[t^*, \infty)$ intervallumon. Ha $|y(t)|$ soha nem lenne α_1-nél kisebb, akkor $x(t)\in H$ $(t\geqq t^*)$ teljesülne, ami ellentmond (4)-nek.

Ezzel a bizonyítás teljes.

Sok esetben nehéz biztosítani, hogy a megoldások z-korlátosak legyenek, ezért megadunk egy olyan feltételrendszert is, amelyben ez a megszorítás nem szerepel. Mivel a megoldások trajektóriája ekkor általában nincs benne egy kompakt halmazban, a bizonyítás nem olyan egyszerű, mint az előbb, de lényegében ugyanazokat a gondolatokat kell felhasználni, mint a 6.4. és 6.6. tétel bizonyításában, ezért a bizonyítást elhagyjuk.

6.7. TÉTEL. Tegyük fel, hogy a (3.2) rendszerhez léteznek olyan V, W, A függvények, amelyek a 6.6. tételben szereplő (1)—(2) tulajdonságon kívül rendelkeznek a következő tulajdonságokkal:

(3) van olyan $c\in\mathscr{K}$ és $C>0$, hogy $U(x)\leqq -c(\varrho^*(x, F))$ $(|y|\leqq C)$;

(4) $\varrho^*(p, F)>0$ $(p\in F^c)$;

(5) bármely α_1, α_2 $(0<\alpha_1<\alpha_2<C)$ számhoz van olyan $\beta>0$ és $\psi\colon R_+\to R_+$ $\left(\int_0^\infty \psi=\infty\right)$ függvény, hogy

$$\psi(t)\leqq\inf\{|\dot{A}(t, x)|\colon \alpha_1\leqq|y|\leqq\alpha_2,\quad x\in S^*(F,\beta)\};$$

(6) bármely $u\colon R_+\to H:=\{x\colon |y|\leqq C\}$ folytonos függvényre $A(t, u(t))$ korlátos és $\int_0^t \dot{W}(s, u(s))\,ds$ egyenletesen folytonos R_+-on.

Akkor (3.2) 0-megoldása aszimptotikusan y-stabilis.

Eredményeink birtokában adjuk meg a 6.2. tétel általánosítását nem-autonóm rendszerekre, illetve parciális stabilitásra. Ehhez vezessük be a következő függvényosztályt.

6.3. DEFINÍCIÓ. Azt mondjuk, hogy a $\varphi\colon R_+\to R_+$ folytonos függvény az $\mathscr{F}$ *függvényosztályhoz* tartozik $(\varphi\in\mathscr{F})$, ha létezik $\varphi(t)=\varphi_1(t)+\varphi_2(t)$ alakú felbontása, ahol $\varphi_1, \varphi_2\colon R_+\to R_+$ folytonos, φ_1 korlátos R_+-on, és $\int_0^\infty \varphi_2<\infty$.

6.8 TÉTEL. Tekintsük a (3.5) mechanikai rendszert. Ha

(1) a P potenciális energiának a $q=0$ pontban minimuma van;

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

(3) a disszipáció teljes, vagyis létezik olyan β integrálisan pozitív függvény, hogy $R(t, q, \dot{q})\geqq\beta(t)|\dot{q}|^2$ $(t\in R_+)$;

(4) bármely $K \subset R^n$ kompakt halmazra

$$\max \{|G(t, q)-B(t, q)|: \ q \in K\} \in \mathscr{F},$$

akkor a (3.5) rendszer $q=\dot{q}=0$ egyensúlyi helyzete aszimptotikusan stabilis.

Bizonyítás. Tekintsük a (3.5)-tel ekvivalens (5.5) *Hamilton-rendszert,* és legyen $V(q, p):=H(q, p)=T+P$. Ez a függvény pozitív definit, hiszen T pozitív definit p-ben, P pedig q-ban. Utóbbi abból következik, hogy $P(q) \geqq 0$, $P(0)=0$, és ha a $q=0$-hoz lenne akármilyen közel P-nek 0-helye, akkor (2) nem teljesülne, hiszen P minden 0-helye egyensúlyi helyzet. Másrészt $H(0, 0)=0$ és H folytonos, így a 6.6. tétel (1) feltétele teljesül.

Legyen $W(q, p) \equiv A(t, q, p):=-p$. Ekkor

$$\dot{V}=-2R \leqq -\beta(t)\,|A^{-1}(q)\,p|^2, \quad \dot{W}=\dot{A}=\frac{\partial H}{\partial q}-(G-B)A^{-1}p,$$

tehát $U(q, p):=|A^{-1}(q)p|^2$, $F:=\{(q, p): p=0\}$, és ezért a 6.6. tétel (2)—(3) feltétele is teljesül. Megmutatjuk, hogy a (4)—(5) feltétel is teljesül. Ha $0<\alpha_1<\alpha_2$, α_2 elég kicsi, akkor létezik olyan β $(0<\beta<\alpha_1/2)$ és $\xi \in F$, hogy az $\alpha_1 \leqq |(q, p)| \leqq \alpha_2$, $|p| \leqq \beta$ halmazon

$$|\dot{A}| \geqq |\operatorname{grad} P(q)|-\left[\sum_{i=1}^{n}\left(p, \frac{\partial A^{-1}(q)}{\partial q_i}\,p\right)^2\right]^{1/2}-|G(t, q)-B(t, q)|\,|A^{-1}(q)||p| \geqq$$

$$\geqq \gamma-k_1\beta-k_2\beta\xi(t)=\gamma-k_3\beta+k_4\beta\xi_2(t) \quad (k_j=\text{állandó}),$$

ahol $\gamma:=\min\{|\operatorname{grad} P(q)|: \alpha_1/2 \leqq |q| \leqq \alpha_2\}$, $\int_0^\infty \xi_2 < \infty$. Tehát $\dot{A}(t, q, p) \neq 0$ szigorúan az F halmaz kompakt részhalmazain, vagyis (4) teljesül. Ha $K \subset R^{2n}$ kompakt, és $(q, p) \in K$, akkor létezik olyan $\zeta \in \mathscr{F}$, hogy $|\dot{W}| \leqq k_5+k_6\zeta(t)$. Mivel $\int_0^t \zeta(s)\,ds$ egyenletesen folytonos a $[0, \infty)$ intervallumon, az (5) feltétel is teljesül.

Tételünk állítása tehát a 6.6. tétel következménye.

Hasonlóan vezethető le a 6.7. tételből a (3.5) $q=\dot{q}=0$ egyensúlyi helyzetének parciális stabilitására vonatkozó következő állítás.

6.9. TÉTEL. Legyen $q=(\tilde{q}, \hat{q})$, $\tilde{q} \in R^k$, $0<k \leqq n$, és tegyük fel, hogy ha K az R^{n+k} tetszőleges kompakt részhalmaza, akkor $(\tilde{q}, \dot{q}) \in K$ esetén teljesülnek a következő feltételek:

(1) $P(0)=0$, $|\operatorname{grad} P(q)| \geqq k_1 > 0$ $(|\tilde{q}| \geqq \alpha > 0)$;

(2) $\left|\dfrac{\partial T(q, \dot{q})}{\partial q}+\operatorname{grad} P(q)\right| \leqq k_2$;

(3) van olyan $a, b \in \mathscr{K}$, hogy

$$a(|(\tilde{q}, \dot{q})|) \leqq T(q, \dot{q})+P(q) \leqq b(|(\tilde{q}, \dot{q})|);$$

(4) a disszipáció teljes (l. 6.8. tétel, (3));

(5) $\sup\{|G(t, q)-B(t, q)|: (\tilde{q}, \dot{q}) \in K\} \in \mathscr{F}$.

Akkor a (3.5) rendszer $q=\dot{q}=0$ egyensúlyi helyzete aszimptotikusan $(\tilde{q}, \dot{q})$-stabilis.

6.2. PÉLDA. *Giroszkóp aszimptotikus stabilitása.* Tekintsünk egy O pontjában rögzített, szimmetrikus merev testet, amely egy x_1, y_1, z_1 inerciarendszerben mozog. Tegyük fel, hogy az O pont egybeesik a koordináta-rendszer kezdőpontjával, és a z_1 tengely függőlegesen felfelé mutat. Tekintsünk egy olyan (mozgó) x, y, z koordináta-rendszert is, amely a testhez van rögzítve, kezdőpontja O, és z tengelye egybeesik a test szimmetriatengelyével. Jelölje ω a test pillanatnyi szögsebességvektorát, amelynek komponensei az x, y, z rendszerben legyenek p, q, r. Ekkor a test P pontjának sebességvektorát a $v_P=[\omega, \overrightarrow{OP}]$ vektorális szorzat adja, tehát a test kinetikai energiája $T=[A(p^2+q^2)+Cr^2]/2$, ahol A és C a test főtehetetlenségi nyomatékai. Könnyű kiszámolni, hogy a test impulzusmomentuma $\sigma=\operatorname{grad} T(p, q, r)$. Az impulzusmomentum-tétel szerint $\dot{\sigma}+[\omega, \sigma]=m_0$, ahol m_0 a külső erők O pontra vonatkozó momentumának eredője. Tegyük fel, hogy a testre a gravitációs erő, a szögsebességgel arányos súrlódási erő, és olyan további erők hatnak, amelyek csak a szimmetriatengelyre vonatkozóan fejtenek ki nyomatékot.

Ekkor a mozgásegyenletek alakja:

$$
\begin{aligned}
A\dot{p} &= (A-C)qr+mgz_0\gamma_2-\frac{\partial R(t, p, q)}{\partial p},\\
A\dot{q} &= (C-A)pr-mgz_0\gamma_1-\frac{\partial R(t, p, q)}{\partial q},\\
C\dot{r} &= M(t, r),\\
\dot{\gamma}_1 &= r\gamma_2-q\gamma_3,\\
\dot{\gamma}_2 &= p\gamma_3-r\gamma_1,\\
\dot{\gamma}_3 &= q\gamma_1-p\gamma_2,\\
\gamma_1^2+\gamma_2^2+\gamma_3^2 &= 1,
\end{aligned}
$$

ahol R a p, q változókban kvadratikus alak korlátos együtthatókkal; $M(t, r)$ a külső erőknek a test szimmetriatengelyére vonatkozó forgatónyomatéka, amely korlátos, ha r kompakt halmazon változik; m a test tömege; g a nehézségi gyorsulás, z_0 a test súlypontjának applikátája az x, y, z rendszerben; $\gamma_1, \gamma_2, \gamma_3$ az Oz irányba mutató egységvektor komponensei az x_1, y_1, z_1 rendszerben.

Tegyük fel, hogy a harmadik egyenletnek van egy korlátos $r=r(t)$ megoldása. Ekkor $p=q=0$, $r=r(t)$, $\gamma_1=\gamma_2=0$, $\gamma_3=+1$ megoldása a rendszernek, amely a giroszkóp függőleges irányú szimmetriatengelye körüli forgásnak felel meg. Bebizonyítjuk, hogy *ha a test a súlypontja felett van rögzítve* $(z_0<0)$ *és a disszipáció teljes* $(R(t, p, q) \geqq \geqq \beta(p^2+q^2), \beta>0)$, *akkor ez a mozgás aszimptotikusan stabilis.*

Ha végrehajtjuk az $r=r(t)$ és $\gamma_3=\sqrt{1-\gamma_1^2-\gamma_2^2}$ helyettesítést, akkor az

$$
\text{(6.7)} \qquad \begin{cases} A\dot{p} = (A-C)qr(t)+mgz_0\gamma_2-\partial R/\partial p, \\ A\dot{q} = (C-A)pr(t)-mgz_0\gamma_1-\partial R/\partial q, \\ \dot{\gamma}_1 = r(t)\gamma_2-q\sqrt{1-\gamma_1^2-\gamma_2^2}, \\ \dot{\gamma}_2 = p\sqrt{1-\gamma_1^2-\gamma_2^2}-r(t)\gamma_1 \end{cases}
$$

egyenletrendszert kapjuk. Kiszámolva a p, q, γ_1, γ_2 változókban pozitív definit

$$V = \frac{1}{2} A(p^2+q^2) - \frac{1}{2} mgz_0(\gamma_1^2+\gamma_2^2+(1-\sqrt{1-\gamma_1^2-\gamma_2^2})^2)$$

Ljapunov-függvénynek (ami a teljes mechanikai energia) a rendszer szerinti deriváltját, a

$$\dot{V} = -p\frac{\partial R}{\partial p} - q\frac{\partial R}{\partial q} = -2R \leqq -\beta(p^2+q^2)$$

formulát nyerjük. A jobb oldalon álló függvény 0-helyeinek halmaza $F := \{(p, q, \gamma_1, \gamma_2): p=q=0\}$. Olyan $A(p, q, \gamma_1, \gamma_2)$ segédfüggvényre van tehát szükségünk, amelynek deriváltja szigorúan különbözik 0-tól ezen a halmazon, tehát „nem lehet kicsi, ha p^2+q^2 kicsi, és $\gamma_1^2+\gamma_2^2$ nem kicsi". Ha az első egyenletet γ_2-vel, a másodikat γ_1-gyel szorozzuk, és a kettőt kivonjuk egymásból, akkor a jobb oldalon pontosan egy ilyen kifejezés alakul ki. Legyen tehát $A(p, q, \gamma_1, \gamma_2) := A(p\gamma_2 - q\gamma_1)$, ennek a rendszer szerinti deriváltja:

$$\dot{A}(t, p, q, \gamma_1, \gamma_2) = mgz_0(\gamma_1^2+\gamma_2^2) - Cr(t)(q\gamma_2+p\gamma_1) + \gamma_1\frac{\partial R}{\partial q} - \gamma_2\frac{\partial R}{\partial p} + A(p^2+q^2)\sqrt{1-\gamma_1^2-\gamma_2^2},$$

amely szigorúan különbözik 0-tól a $p=q=0$ halmazon (l. 6.1. def.), mivel $R(t)$ együtthatói és $r(t)$ korlátos függvények. Tehát a 6.6. tétel feltételei teljesülnek $(W \equiv y := (p, q, \gamma_1, \gamma_2))$, így állításunk bizonyítva van.

7. A különböző típusú erők hatása mechanikai rendszer egyensúlyi helyzetének stabilitására

Legelőször is adjuk meg a mechanikai rendszerekre ható erők egy osztályozását. Tekintsünk egy — a 3. pontban már vizsgált — holonom, szkleronom mechanikai rendszert, amelyre a $Q(q, \dot{q})$ általánosított erő hat:

$$\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{q}}\right) - \frac{\partial T}{\partial q} = Q(q, \dot{q}) \quad (Q(0, 0) = 0). \tag{7.1}$$

Ebben az egyenletben $q, \dot{q} \in R^n$; $T(q, \dot{q}) = (\dot{q}, A(q)\dot{q})/2$ a kinetikai energia, amely $\dot{q}$-ban pozitív definit kvadratikus alak. Ha $A(q)$ és $Q(q, \dot{q})$ elég sima, akkor ez az egyenlet az

$$A_0 q + B_1\dot{q} + C_1 q = F(q, \dot{q})$$

alakba is írható, ahol $A_0 := A(0)$, B_1, C_1 konstansokból álló $n \times n$-es mátrix, $F(q, \dot{q}) = o((q, \dot{q}))$. Legyen

$$B := \frac{1}{2}(B_1+B_1') \quad G := \frac{1}{2}(B_1-B_1'); \quad (B' = B,\ G' = -G)$$

$$C := \frac{1}{2}(C_1+C_1') \quad D := \frac{1}{2}(C_1-C_1'); \quad (C' = C,\ D' = -D).$$

Ezekkel a jelölésekkel rendszerünk egyenlete a következő:

$$A_0\ddot{q}+B\dot{q}+G\dot{q}+Cq+Dq = F(q,\dot{q}). \tag{7.2}$$

Ez az egyenlet úgy is interpretálható, hogy adott egy mechanikai rendszer a $T_0=T_0=(\dot{q}, A_0\dot{q})/2$ kinetikai energiával, amelyre a következő erők hatnak: $-Cq$ *potenciális* vagy *konzervatív* erő, $P:=(q, Cq)/2$ a potenciális energia; $-B\dot{q}$ *disszipatív erő,* ha az $R:=(\dot{q}, B\dot{q})/2$ *Rayleigh-féle függvény* pozitív szemidefinit kvadratikus alak, *gyorsító erő,* ha R negatív szemidefinit; $-G\dot{q}$ *giroszkopikus erő;* $-Dq$ *nem-konzervatív erő*[3], $F(q, \dot{q})$ *nem-lineáris erő.*

A jelen fejezetben azzal a kérdéskörrel foglalkozunk, hogy a csak konzervatív erő hatása alatt álló

$$A_0\ddot{q}+Cq = 0 \tag{7.3}$$

rendszer $q=\dot{q}=0$ egyensúlyi helyzetének stabilitási viszonyait hogyan befolyásolják az előbb bevezetett típusú kiegészítő erők. A legérdekesebb kérdések: Mikor lehet a (7.3) rendszer instabilis egyensúlyi helyzetét *stabilizálni* alkalmas giroszkopikus erő bevezetésével? Milyen disszipatív erők alkalmasak arra, hogy (7.3) stabilis egyensúlyi helyzetét *aszimptotikusan stabilissá tegyék*?

Hogy az első kérdésre választ adhassunk, be kell vezetnünk egy fogalmat: a (7.3) 0-megoldásának instabilitási fokszámát. Ha a (7.3) egyenletben végrehajtjuk a $q=Sx$ transzformációt, akkor az

$$S'A_0S\ddot{x}+S'CSx = 0$$

egyenletet kapjuk. Ismert a lineáris algebrának az a tétele, hogy tetszőleges A_0 pozitív definit szimmetrikus és C szimmetrikus mátrixokhoz létezik olyan S, nemszinguláris mátrix, hogy $S'A_0S=E$, $S'CS$ pedig diagonális [3], vagyis a mozgásegyenletek egymástól függetlenek:

$$\begin{aligned}\ddot{x}_1+\lambda_1x_1 &= 0\\ &\vdots\\ \ddot{x}_n+\lambda_nx_n &= 0.\end{aligned}$$

Az i-edik egyenlet 0-megoldása akkor és csakis akkor stabilis, ha $\lambda_i\geqq 0$. A $\lambda_1, \ldots, \lambda_n$ számokat a (7.3) rendszer *stabilitási együtthatói*nak, a köztük előforduló negatív számok számát (7.3) *instabilitási fokszám*ának nevezzük. A *Sylvester-tétel* [3] szerint az $S'CS$ negatív diagonális elemeinek száma megegyezik, a C mátrix negatív sajátértékeinek számával. Tehát (7.3) *instabilitási fokszáma megegyezik C negatív sajátértékeinek számával.*

7.1. TÉTEL. Ha (7.3) instabilitási fokszáma páratlan, és C sajátértékei különböznek 0-tól, akkor (7.3) $q=\dot{q}=0$ egyensúlyi helyzete nem stabilizálható giroszkopikus, disszpatív és nem-lineáris erő segítségével.

Bizonyítás. Megmutatjuk, hogy ha C teljesíti a tétel feltételeit, akkor az

$$A_0\ddot{q}+B\dot{q}+G\dot{q}+Cq = F(q,\dot{q}) \tag{7.4}$$

[3] Az elnevezések közül kettő nem szerencsés, hiszen minden erő „gyorsító" erő, illetve a $-Cq$ erőn kívül mindegyik „nem-konzervatív". Azért tartjuk meg mégis ezeket az elnevezéseket, mert az irodalomban általánosan elterjedtek.

egyenlet $q=\dot{q}=0$ egyensúlyi helyzete tetszőleges B, G, F esetén instabilis. (7.4) első közelítésének karakterisztikus egyenlete

$$\Delta(\lambda) := \det\left(\lambda^2 A_0+\lambda(B+G)+C\right) = 0.$$

Egyrészt, $\Delta(0)<0$, hiszen $\det C<0$. Másrészt $\Delta(\lambda)\to\infty$ $(\lambda\to\infty)$, tehát $\Delta(\lambda)$-nak van pozitív 0-helye, így a 4.2. tétel *b)* állítása szerint (7.4) 0-megoldása instabilis.

7.2. TÉTEL. Tegyük fel, hogy (7.3) instabilitási fokszáma páros, és C sajátértékei 0-tól különbözők. Ekkor van olyan G ferdén szimmetrikus mátrix, hogy az

$$A_0\ddot{q}+G\dot{q}+Cq = 0 \tag{7.5}$$

egyenlet 0-megoldása stabilis, azaz (7.3) $q=\dot{q}=0$ egyensúlyi helyzete nem-lineáris erők távollétében alkalmas giroszkopikus erővel stabilizálható.

Bizonyítás. (7.3) karakterisztikus egyenlete:

$$\Delta(\lambda) := \det(\lambda^2 A_0+\lambda G+C) = \det(\lambda^2 A_0+\lambda G+C)' =$$

$$= \det(\lambda^2 A_0-\lambda G+C) = \Delta(-\lambda),$$

tehát a $\Delta(\lambda)=0$ egyenletnek akkor és csakis akkor nem lesz pozitív valós részű gyöke, ha minden gyöke tiszta képzetes.

Legyen S olyan nem-elfajuló mátrix, hogy $S'A_0S=E$, $C^*=S'CS$ diagonális. A $q=Sx$ transzformációval (7.5) az

$$\ddot{x}+G^*\dot{x}+C^*x = 0 \tag{7.6}$$

alakba írható, ahol $G^*=S'GS$ ferdén szimmetrikus, C^* diagonális, amelynek főátlójában az első $2s$ elem negatív. (7.6) karakterisztikus egyenlete:

$$\Delta^*(\lambda) := \det(\lambda^2 E+\lambda G^*+C^*).$$

Konstruáljuk G^*-ot a következőképpen: legyen $g^*_{2i,2i-1} = -g^*_{2i-1,2i}\neq 0$ $(i=1,2,\ldots,s)$, az összes többi helyen álljon 0. Ekkor

$$\Delta^*(\lambda) = \prod_{i=1}^{s}\left[(\lambda^2+c^*_{2i-1})(\lambda^2+c^*_{2i})+\lambda^2(g^*_{2i,2i-1})^2\right]\prod_{k=2s+1}^{n}(\lambda^2+c^*_k),$$

tehát, ha G^* 0-tól különböző elemeit elég nagyra választjuk, akkor $\Delta^*(\lambda)$ minden 0-helye tiszta képzetes. Mivel (7.5) diagonizálható (l. [28]), ez azt jelenti, hogy az egyensúlyi helyzet stabilis.

A bizonyításból kitűnik, hogy alkalmas G-vel (7.5) egyensúlyi helyzete csak stabilis, de nem aszimptotikusan stabilis. Mint ahogyan a 4. pontban már láttuk, az ilyen stabilitás megszűnhet, ha az egyenletbe magasabb rendű tagok, esetünkben nem-lineáris erők lépnek be. Sőt, a stabilitás csak q-tól függő nem-lineáris erő fellépésével is megszűnhet [28, 122. o.]. Ugyanakkor felmerül a következő érdekes, nehéznek tűnő — mindmáig megoldatlan — kérdés: Legyen adva az A_0 és C mátrix, melyek a 7.2. tétel feltételeit kielégítik, továbbá egy $P: R^n\to R^n$, folytonosan differenciálható függvény, amelyre $\operatorname{grad} P(q)=o(|q|)$. Létezik-e olyan G ferdén

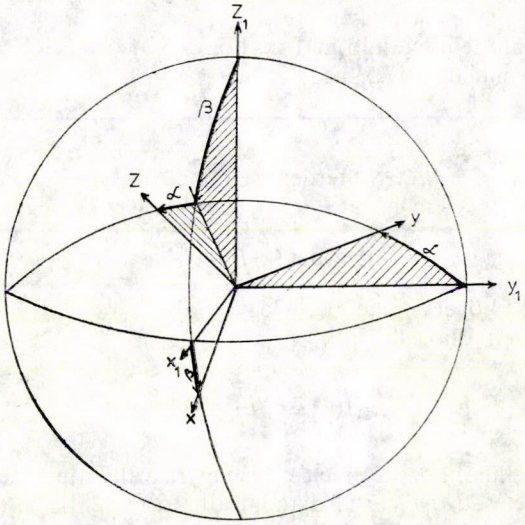
szimmetrikus mátrix, amellyel az

$$A_0\ddot{q}+G\dot{q}+Cq-\operatorname{grad}P(q)=0$$

egyenlet $q=\dot{q}=0$ megoldása stabilis?

Bebizonyítható [46], hogy (7.5) $q=\dot{q}=0$ egyensúlyi helyzetének stabilitása akkor is megszűnik, ha a rendszerre akármilyen kicsiny, de teljes disszipációjú disszipatív erő hat. (Ez egybeesik azzal a tapasztalati ténnyel, hogy ha alkatrészek forgatásával stabilizálunk egy szerkezetet, de nem gondoskodunk a mindig fellépő súrlódás miatt elhasználódott energia pótlásáról, a stabilitás megszűnik.) Ezért a 7.2. tétel alapján létrehozott stabilitást „ideiglenesnek" szokás nevezni.

7.1. PÉLDA. *A pörgettyű stabilizálása.* Vizsgáljuk újra szimmetrikus merev test mozgását, ha egy pontja rögzítve van. A mozgás leírására ugyanazt a két koordinátarendszert használjuk, mint a 6.2. példában, viszont feltesszük, hogy a rögzített pont a test súlypontja *alatt* van ($z_0>0$). Ekkor a (6.7) rendszer $p=q=\gamma_2=\gamma_3=0$ egyensúlyi helyzete instabilis, ha a test nem forog ($r(t)\equiv 0$), de a tapasztalat azt sejteti velünk, hogy stabilissá válik, ha a forgás elég gyors (pl. búgócsiga). A 7.2. tétel segítségével válaszoljunk arra a kérdésre, hogy az $r(t)=r_0$ állandó mely értékénél válik az egyensúlyi helyzet stabilissá, ha csak a lineáris erők hatását vesszük figyelembe, és a súrlódással nem számolunk. A szimmetriatengely helyzetét a 2. ábrán



2. ábra

látható α, β szöggel jellemezzük, a test szimmetriatengelye körüli elfordulásának szögét φ jelöli. A test mozgásának a 6.2. példában használt jellemzőit az új szögekkel az alábbi képletekkel fejezhetjük ki:

$$p=\dot{\alpha},\quad q=\dot{\beta}\cos\alpha,\quad r=\dot{\varphi}-\dot{\beta}\sin\alpha,$$

$$\gamma_1=\cos\alpha\sin\beta,\quad \gamma_2=-\sin\alpha,\quad \gamma_3=\cos\alpha\cos\beta.$$

A kinetikai és helyzeti energia:

$$T = \frac{1}{2} A(\dot{\alpha}^2 + \dot{\beta}^2 \cos^2 \alpha) + \frac{1}{2} C(\dot{\varphi} - \dot{\beta} \sin \alpha)^2,$$

$$P = -mgz_0(1 - \cos \alpha \cos \beta).$$

Mivel φ ciklikus koordináta ($\partial T/\partial \varphi = 0$, $\partial P/\partial \varphi = 0$), a φ-re vonatkozó *Lagrange-féle mozgásegyenlet* szerint $\dfrac{\partial T}{\partial \dot{\varphi}} = C(\dot{\varphi} - \dot{\beta} \sin \alpha) = Cr_0 =$ állandó. Hogy a mozgásegyenletek linearizált alakját megkapjuk, írjuk fel a *Lagrange-féle mozgásegyenleteket* a

$$T_0 := \frac{1}{2} A(\dot{\alpha}^2 + \dot{\beta}^2) + \frac{1}{2} C(\dot{\alpha} - \dot{\beta}\alpha)^2,$$

$$P_0 := -\frac{mgz_0}{2}(\alpha^2 + \beta^2)$$

„redukált" kinetikai és potenciális energiával:

$$A\ddot{\alpha} + Cr_0\dot{\beta} - mgz_0\alpha = 0$$

$$A\ddot{\beta} - Cr_0\dot{\alpha} - mgz_0\beta = 0.$$

Az instabilitás foka 2, tehát alkalmas $Cr_0(-\dot{\beta}, \dot{\alpha})$ giroszkopikus erővel az egyensúlyi helyzet stabilizálható. A rendszer karakterisztikus egyenlete:

$$\Delta(\lambda) = \begin{vmatrix} A\lambda^2 - mgz_0 & Cr_0 \\ -Cr_0 & A\lambda^2 - mgz_0 \end{vmatrix} = A^2\lambda^4 + (C^2r_0^2 - 2Amgz_0)\lambda^2 + (mgz_0)^2 = 0,$$

ennek gyökei tiszta képzetesek, ha

$$r_0 > \frac{1}{C}\sqrt{2mgz_0 A}.$$

Be lehet bizonyítani [46], hogy ezen feltétel teljesülése esetén az egyensúlyi helyzet a nem-lineáris tagokat figyelembe véve is stabilis.

7.3. TÉTEL. Ha a (7.3) potenciális rendszer $q = \dot{q} = 0$ egyensúlyi helyzete stabilis, akkor

a) tetszőleges giroszkopikus és disszipatív erő fellépése esetén is stabilis marad;

b) tetszőleges giroszkopikus és teljes disszipatív erő ((p, Bp) pozitív definit) hatására aszimptotikusan stabilissá válik.

Bizonyítás. a) (7.3) 0-megoldása stabilis, tehát C minden sajátértéke pozitív, így a $H = (\dot{q}, A_0\dot{q})/2 + (q, Cq)/2$ teljes energia pozitív definit. Tetszőleges giroszkopikus és disszipatív erő bevezetése után a rendszer

$$A_0\ddot{q} + B\dot{q} + G\dot{q} + Cq = 0$$

alakú. Az energia ezen rendszerre vonatkozó deriváltja: $\dot{H} = -2(\dot{q}, B\dot{q}) \leqq 0$, így a $q = \dot{q} = 0$ megoldás a 3.1. tétel szerint stabilis.

b) Ez az állítás a 6.2. tételnek speciális esete.

A fenti bizonyítás felhívja a figyelmet arra, hogy dolgozatunknak a mechanikai rendszerekre vonatkozó néhány korábbi tétele a jelen fejezet kérdésköréhez tartozik. A 6.2. tétel az éppen most bizonyított klasszikus állítást általánosítja arra az esetre, amikor q-tól függő *nem-lineáris erők* is fellépnek, a 6.9. tétel pedig ugyanezen eredmények *instacionárius* rendszerekre vonatkozó megfelelőjét adja. A 3.6. és 5.2. tételek arra adnak választ, hogy milyen *parciális stabilitási* tulajdonságok állíthatók, ha nem-lineáris potenciális erők is fellépnek, és a potenciális energia csak pozitív szemidefinit.

Az eddigi megjegyzések sejteni engedik, hogy a jelen fejezetben vázolt kérdéskör még nagyon sok érdekes problémát szolgáltat. A jelenlegi vizsgálatok egy részének célja: tisztázni a különböző típusú erők egyedi és együttes hatását az egyensúlyi helyzet stabilitására [40—41]. (A nem-potenciális erők hatásával mi itt nem foglalkoztunk. A vizsgálatok azt mutatják [46—47], hogy — ellentétben a giroszkopikus és disszipatív erőkkel — kifejthetnek stabilizáló és destabilizáló hatást is. Erre vonatkozóan l. a [20] magyar nyelvű dolgozatot.) A kutatások másik iránya az elmélet felépítése instacionárius mechanikai rendszerekre. Ezen a téren — még a csak lineáris erőket tartalmazó esetben is — nagyon kevés eredmény ismeretes. Fejezetünk hátralevő részében egy ide tartozó problémát tárgyalunk: egyensúlyi helyzet aszimptotikus stablizálásának, illetve parciális aszimptotikus stabilizálásának *szükséges* feltételét adjuk instacionárius, nem-lineáris rendszerekre.

Tegyük fel, hogy a (3.1) rendszerben szereplő X függvény folytonosan differenciálható, és tekintsük az $x(t;\ t_0, x_0)$ megoldást. Azt már korábban feltettük, hogy az $x=0$ kis környezetéből induló megoldások a $[t_0, \infty)$ intervallumon értelmezve vannak. Másrészt, X differenciálhatósága miatt $x(t;\ t_0, x_0)$ x_0-nak differenciálható függvénye, tehát tetszőleges $t, t_0 \in R_+$ $(t_0 \leqq t)$ számokra az $x(t;\ t_0, \cdot): G(0, \lambda) \to R^n$ leképezés diffeomorfizmus $(\lambda > 0)$. Jelölje ennek *Jacobi-determinánsát*

$$J(x_0;\ t,\ t_0) := \det\left(\frac{\partial(x_1(t;\ t_0, x_0), \ldots, x_n(t;\ t_0, x_0))}{\partial(x_{01}, \ldots, x_{0n})}\right).$$

Liouville tétele szerint [7]

$$J(x_0;\ t,\ t_0) = \exp\left[\int_{t_0}^{t} \sum_{i=1}^{n} \left(\frac{\partial X_i(s, x)}{\partial x_i}\right)_{x=x(s;\ t_0, x_0)} ds\right]. \tag{7.6}$$

Vezessük még be az

$$x(t;\ t_0, F) = \{x(t;\ t_0, x_0):\ x_0 \in F\} \quad (F \subset R^n)$$

$$D_t = x(t;\ 0,\ \{x_0:\ |x_0| \leqq l\}) \quad (t \geqq 0)$$

jelöléseket, ahol $l > 0$ olyan kicsiny, hogy ha $|x_0| \leqq l$, akkor $x(t;\ t_0, x_0)$ létezik a $[t_0, \infty)$ intervallumon.

7.4. TÉTEL. Ha

$$\limsup_{t \to \infty} \int_0^t \min\left\{\sum_{i=1}^{n} \frac{\partial X_i(s, x)}{\partial x_i}:\ x \in D_s\right\} ds > -\infty, \tag{7.7}$$

akkor (3.1) 0-megoldása nem lehet attraktív, nevezetesen az

$$E := \left\{x_0:\ |x_0| < l,\ \lim_{t \to \infty} |x(t;\ t_0, x_0)| = 0\right\}$$

halmaz *Lebesgue-mértéke* 0.

Bizonyítás. Először is bizonyítsuk be, hogy E *Lebesgue-mérhető.* E célból tekintsük a

$$H_m^k := \left\{x_0:\ |x_0| < l,\quad |x(t;0,x_0)| < \frac{1}{k}\quad (m \leqq t \leqq m+1)\right\}$$

$(m, k=1, 2 \ldots)$ halmazokat. Ezek a halmazok nyitottak, mivel $x(t; 0, x_0)$ a t, x_0 változóknak folytonos függvénye. Tehát az

$$E = \bigcap_{k=1}^{\infty}\left(\bigcup_{j=1}^{\infty}\bigcap_{m=j}^{\infty} H_m^k\right)$$

halmaz *Lebesgue-mérhető.*

A (7.7) feltétel és a (7.6) formula következtében létezik olyan $\{t_k\}$ sorozat és K szám, hogy $t_k \to \infty$ $(k \to \infty)$, és a D_0 halmaz tetszőleges mérhető F részhalmazára teljesül a

$$\mu[x(t_k; 0, F)] = \int_{x(t_k;0,F)} \ldots \int dx_1 \ldots dx_n = \tag{7.8}$$

$$= \int_F \ldots \int J(x_0; t_k, 0)\, dx_{01} \ldots dx_{0n} \geqq e^K \mu[F] \quad (k = 1, 2, \ldots)$$

becslés, ahol $\mu[H]$ a H halmaz *Lebesgue-féle mértékét* jelöli.

Tegyük fel, hogy a tétel állítása hamis, vagyis $\mu[E]>0$. Akkor *Jegorov tétele* [31] szerint létezik olyan mérhető $E^* \subset E$ $(\mu[E^*]>\mu[E]/2)$, hogy $|x(t_k; 0, x_0)| \to 0$ $(k \to \infty)$ az $x_0 \in E^*$ halmazon egyenletesen. Vagyis, tetszőleges $\eta>0$ számhoz létezik olyan $k(\eta)$, hogy $x(t_{k(\eta)}; 0, E^*) \subset G_\eta$, ahol G_η az O középpontú, η mértékű gömböt jelöli. Legyen

$$G_\eta^{-1} := x(0; t_{k(\eta)}, G_\eta).$$

A (7.8) becslésből következik, hogy

$$\eta \geqq \mu[x(t_{k(\eta)};\ 0, G_\eta^{-1})] \geqq e^K \mu[G_\eta^{-1}] > e^K \mu[E]/2,$$

ami ellentmond annak, hogy η tetszőlegesen kicsiny.

Ezzel a tétel be van bizonyítva.

7.2. PÉLDA. Telkintsük az $\ddot{x} - a^2 x = 0$ $(a>0, x \in R)$ egyenletet, illetve a vele ekvivalens

$$\dot{x}_1 = x_2, \quad \dot{x}_2 = a^2 x_1$$

rendszert. A (7.7) feltétel nyilvánvalóan teljesül. Az E halmaz az

$$x_2 = -a x_1, \quad \left(-\frac{l}{\sqrt{1+a^2}} < x_1 < \frac{l}{\sqrt{1+a^2}}\right)$$

szakasz pontjaiból áll.

A 7.4. tétel egy nagyon fontos következményt ad *Hamilton-féle kanonikus rendszerekre.* Tekintsük először a

$$\dot{x} = \frac{\partial H_0(x, y)}{\partial y}, \quad \dot{y} = -\frac{\partial H_0(x, y)}{\partial x} \quad (x, y \in R) \tag{7.9}$$

stacionárius *Hamilton-rendszert*, ahol $H_0 \not\equiv 0$ a $(0, 0)$ egy környezetében analitikus, $H_0(0, 0)=0$, és $x=y=0$ megoldása (7.9)-nek. A mozgások a $H_0(x, y)=c$ (c=állandó) görbék mentén történnek, mivel H_0 (7.9)-nek első integrálja, tehát az $x=y=0$ megoldás nem lehet attraktív. N. P. JERUGIN [37] vetette fel a következő problémát: Tekintsük a (7.9)-nek egy perturbációját a $H(x, y, t)=H_0(x, y)+P(x, y, t)$ $(P(0, 0, t)\equiv 0)$ *Hamilton-függvénnyel*, vagyis az

$$\dot{x} = \frac{\partial H(x, y, t)}{\partial y}, \quad \dot{y} = -\frac{\partial H(x, y, t)}{\partial x} \tag{7.10}$$

rendszert, feltéve, hogy (7.9) 0-megoldása stabilis. Létezik-e olyan P perturbáció, amelynek hatására (7.10) 0-megoldása aszimptotikusan stabilis? JERUGIN *Ljapunov-módszerrel* konstruált ilyen P függvényt, de ennek és deriváltjának bizonyos görbe mentén szakadása volt. Felvetette a kérdést, hogy lehet-e sima perturbációt adni a kívánt tulajdonsággal. A 7.4. tétel alábbi következménye mutatja, hogy a kérdésre a válasz tagadó.

7.1. KÖVETKEZMÉNY. Tekintsük a

$$\dot{q} = \frac{\partial H(t, q, p)}{\partial p}, \quad \dot{p} = -\frac{\partial H(t, q, p)}{\partial q} \quad (H(t, 0, 0) \equiv 0) \tag{7.11}$$

Hamilton-rendszert, ahol $H: R_+ \times R_n \times R_n \to R$ kétszer folytonosan differenciálható.

A (7.11) rendszer 0-megoldása nem lehet attraktív, tehát nem lehet aszimptotikusan stabilis.

Bizonyítás. Hamilton-rendszerekre a (7.7) feltétel automatikusan teljesül, hiszen a benne szereplő integrál integrandusza azonosan 0.

Megfogalmazzuk a 7.4. tétel megfelelőjét parciális stabilitásra is. A tétel bizonyítása a 7.4. tételéhez hasonló, ezért elhagyjuk.

7.5. TÉTEL. Tegyük fel, hogy a (3.1) rendszer 0-megoldása stabilis. Ha van olyan $L>0$, hogy

$$\limsup_{t\to\infty} \int_0^t \min\left\{\sum_{i=1}^n \frac{\partial X_i(s, x)}{\partial x_i}: |x| \leq L\right\} ds > -\infty,$$

akkor (3.1) 0-megoldása nem attraktív egyetlen x_i komponensre vonatkozóan sem, nevezetesen van olyan $l>0$ hogy az

$$E_i := \{x_0: |x_0| < l, \lim_{t\to\infty} x_i(t; 0, x_0) = 0\} \quad (i = 1, 2, \ldots, n)$$

halmazok Lebesgue-mértéke 0.

7.2. KÖVETKEZMÉNY. Tetszőleges (7.11) Hamilton-rendszer stabilis egyensúlyi helyzete nem lehet aszimptotikusan stabilis egyetlen q_i koordinátára és egyetlen p_j impulzusra sem.

Számítsuk ki a (7.7) feltétel integranduszát az (5.5) giroszkopikus rendszerre:

$$\sum_{i=1}^{n} \frac{\partial^2 H}{\partial p_i \partial q_i} - \sum_{i=1}^{n} \frac{\partial^2 H}{\partial q_i \partial p_i} + \sum_{i,j=1}^{n} (g_{ij} - b_{ij}) \frac{\partial^2 H}{\partial p_i \partial p_j} = -\sum_{i,j=1}^{n} b_{ij} \frac{\partial^2 H}{\partial p_i \partial p_j} =$$

$$= -\operatorname{tr}[B(t, q) A^{-1}(q)].$$

A 7.4. tételből adódik a

7.3. KÖVETKEZMÉNY. A (7.3) rendszer $q = \dot{q} = 0$ egyensúlyi helyzete nem válik aszimptotikusan stabilissá olyan $-B(t, q)\dot{q}$ kiegészítő erő hatására, amely kielégíti a

$$\liminf_{t \to \infty} \int_0^t \max \{\operatorname{tr}[B(s, q)]: |q| \leqq r\} ds < \infty$$

feltételt, bármilyen további $-G(t, q)\dot{q}$ giroszkopikus és $-\partial P(t, q)/\partial q$ potenciális kiegészítő erők hassanak is a rendszerre.

Végül, alkalmazzuk tételeinket az (5.1) változó fonalhosszúságú, súrlódó közegben mozgó inga tanulmányozására. Mint ahogyan az 5. pontban már kiszámítottuk, a

$$W = \frac{g}{l(t)} (\dot{x})^2 + (1 - \cos x)$$

függvénynek (5.1)-re vonatkozó deriváltja:

$$\dot{W} = -\frac{g}{l(t)} (\dot{x})^2 \left[3 \frac{\dot{l}(t)}{l(t)} + 2h(t) \right],$$

tehát a 3.1. tétel szerint az $x = \dot{x} = 0$ egyensúlyi helyzet stabilis, ha

$$3 \frac{\dot{l}(t)}{l(t)} + 2h(t) \geqq 0, \tag{7.12}$$

és $l(t)$ korlátos.

A 7.4. és 7.5. tételből az alábbi állításokat kapjuk:

a) Ha $\ln l^2(t) + \int_0^t h(s)\, ds \to +\infty$ $(t \to \infty)$, *akkor az* $x = \dot{x} = 0$ *egyensúlyi helyzet nem lehet attraktív.*

b) *Ha* (7.12) *teljesül, továbbá* $l(t)$ *és* $\int_0^t h(s)\, ds$ *korlátosak, akkor az* $x = \dot{x} = 0$ *egyensúlyi helyzet stabilis, de nem aszimptotikusan stabilis, sem az* x *szögkitérésre, sem az* $\dot{x}$ *szögsebességre vonatkozóan.*

8. Irodalmi megjegyzések

Az egyes témák tárgyalását nem akartuk állandóan megszakítani az ismertetett eredmények szerzőinek és a pontos irodalmi adatoknak a megadásával, így erre most, a dolgozat végén kerül sor az egyes pontok sorrendjében.

A 2. pontban bevezetett fogalom- és jelölésrendszer folyamatosan alakult ki. A stabilitás és aszimptotikus stabilitás fogalma A. M. LJAPUNOVtól [42], „egyen-

letes" megfelelőik K. P. PERSZIDSZKIJTŐL, illetve I. G. MALKINTÓL [43] erednek. A parciális stabilitásra vonatkozó fogalmakat V. V. RUMJANCEV vezette be [49, 51].

A direkt módszer 3.1—3.3. alaptételei — mint ahogyan a 3. pont elején ezt említettük — A. M. LJAPUNOVTÓL [42] és N. G. CSETAJEVTŐL [53] származnak. V. V. RUMJANCEV [49, 51] vette észre 1957-ben, hogy az alaptételek alkalmasak a parciális stabilitási tulajdonságok tanulmányozására is. A 3.6. tételt is ő közölte [51].

A 4.1—4.3. tételek klasszikus eredmények; már A. M. LJAPUNOV disszertációjában [42] is megtalálhatók, de egészen más felépítésben. A linearizálással végzett stabilitásvizsgálat általános esetére [l. (4.14), $A(t)$ nem konstans és nem periodikus] LJAPUNOV bevezette a karakterisztikus kitevő és a szabályos lineáris rendszer fogalmát, amelyek az általa „első módszer"-nek nevezett elmélet alapfogalmai. Bebizonyította, hogy ha (4.14) első közelítése szabályos rendszer, akkor lehet az első közelítés alapján stabilitásra következtetni. A 4.4. tétel I. G. MALKINTÓL származik. Ő megmutatta, hogy a tétel LJAPUNOV említett tételétől független [43]. A 4.1. és 4.2. példa az ismertetett esetekre V. I. ARNOLD [33] könyvéből való.

Az 5.1. tétel új, ebben a dolgozatban jelenik meg először. Az ingára való alkalmazása magába foglalja a [28] monográfia 38. oldalán található feltételt, amely állandó hosszúságú ingára vonatkozik. Az 5.2. tétel is új — ez inspirálta az 5.1. tétel kimondását.

A *Barbasin—Kraszovszkij-tétel* (6.1. tétel) az alaptételekkel egyenrangú a stabilitáselméletben. A bizonyítás alapgondolatának absztrahálása révén született meg a *LaSalle-féle invariancia-elv* [22—23], amely igen sikeresnek bizonyult, és egyre terjedő irodalma van [28]. A 6.4. tétel az invariancia-elv egy továbbfejlesztése, amely a szerző [18] dolgozatában jelent meg. Ugyanitt található a 6.3. tétel is, amely T. A. BURTON [5] és J. HADDOCK [12] eredményeinek általánosítása. A *Barbasin—Kraszovszkij-tételnek* nem-autonóm rendszerekre való általánosításával elsőként V. M. MATROSZOV [45] foglalkozott a CSETAJEV által felfedezett *két Ljapunov-függvényes módszerek* segítségével — tőle származik a 6.1. példa is. A 6.6—6.7. tétel — amely ebben a dolgozatban jelenik meg először — ezen irány továbbfejlesztése, amennyiben MATROSZOVÉNÁL általánosabb rendszerekre vonatkozik, néhány feltétel kevesebbet követel, és parciális stabilitásról szól. A mechanikai rendszerekre vonatkozó 6.2. tétel L. SALVADORI [29] eredménye, ezt általánosítja nem-autonóm rendszer parciális stabilitására a 6.8—6.9. tétel.

A 7. pontban vázolt kérdéskörhöz a [28, 46—47, 53] monográfiákban találhatunk anyagot. A 7.1—7.3. tételek W. THOMSON és P. TAIT nevéhez fűződnek; a direkt módszerrel való tárgyalást N. G. CSETAJEV [53] adta meg. A 7.1. példában tárgyalt pörgettyű mozgásának teljes stabilitási tárgyalása is CSETAJEV eredménye (nála tűzérségi lövedékről van szó). A 7.4. tétel a szerző és KRÁMLI ANDRÁS [19], a 7.5. tétel a szerző [52] dolgozatából való.

IRODALOM

[1] ANTOSIEWICZ, H. A., "A survey of Ljapunov's second method", *Contr. nonl. oscill.* **4** (1958) 141—156.

[2] BALLIEU, R. J. and PEIFFER, K., "Attractivity of the origin for the equation $\ddot{x}+f(t, x, \dot{x})|\dot{x}^{\alpha}|\dot{x}+g(x)=0$", *J. Math. Anal. Appl.* **65** (1978) 321—332.

[3] BELLMAN, R., *Introduction to matrix analysis* (McGraw-Hill, New York, 1970).

[4] BUDÓ, Á., *Mechanika* (Tankönyvkiadó, Budapest, 1957).

[5] BURTON, T. A., "An extension of Liapunov's direct method", *J. Math. Anal. Appl.* **28** (1969) 545—552 (l. egy korrekció, ugyanott, **32** (1970) 681—691).
[6] CESARI, L., *Asymptotic behaviour and stability problems in ordinary differential equations* (3-rd ed., Springer-Verlag, Berlin—Heidelberg—New York, 1970).
[7] CODDINGTON, E. A. and LEVINSON, N., *Theory of ordinary differential equations* (McGraw-Hill, New York, 1955).
[8] COPPEL, W. A., *Stability and asymptotic behavior of differential equations* (D. C. Heath and Company, Boston, 1965).
[9] CORNE, J. and ROUCHE, N., "Attractivity of closed sets profed by using a family of Liapunov functions", *J. Differential Equations* **13** (1973) 231—246.
[10] CSÁKI, F., *Korszerű szabályozáselmélet* (Akadémiai Kiadó, Budapest, 1970).
[11] GRIMMER, R. and HADDOCK, J., "Stability of bounded and unbounded sets for ordinary differential equations", *Ann. Mat. pura. appl.* **99** (1974) 143—145.
[12] HADDOCK, J., "Stability theory for non-autonomous systems", *Dynamical Systems, An International Symposium*, vol. II, Ed. L. Cesari, J. Hale and J. LaSalle (Academic Press, New York, 1976) 271—274.
[13] HAHN, W., *Stability of motion* (Springer Verlag, Berlin, 1967).
[14] HALANAY, A., *Differential equations: stability, oscillations, time lags* (Academic Press, New York, 1966).
[15] HALE, J. K., *Ordinary differential equations* (Pure and Applied Mathematics ser., vol. XXI, Wiley—Interscience, New York—London—Sydney—Toronto, 1969.)
[16] HATVANI, L., "On the stability of the zero solution of certain second order non-linear differential equations", *Acta Sci. Math.* **32** (1971) 1—9.
[17] HATVANI, L., "On the asymptotic behaviour of the solutions of $(p(t)x')'+q(t)f(x)=0$", *Publicationes Math. Debrecen* **19** (1972) 225—237.
[18] HATVANI, L., "Attractivity theorems for nonautonomous systems of differential equations", *Acta Sci. Math.* **40** (1978) 271—283.
[19] HATVANI, L. and KRÁMLI, A., "A condition for non-asymptotical stability", in: *Differential Equations* Ed. M. Farkas (János Bolyai Mathematical Society and North-Holland Publishing Company, Amsterdam—London, 1975) 269—276.
[20] HATVANI, L. és PINTÉR, L., „Közönséges differenciálegyenletek megoldásainak aszimptotikus viselkedése mechanikai alkalmazásokkal", a „*Differenciálegyenletek és műszaki alkalmazásaik*" c. nyári iskola kiadványában (Miskolc, 1977) 63—88.
[21] LAKSHMIKANTHAM, V. and LEELA, S., *Differential and integral inequalities, theory and applications* (Academic Press, New York—London. 1969).
[22] LA SALLE, J. P., "Stability theory for ordinary differential equations", *J. Differential Equations* **4** (1968) 57—65.
[23] LA SALLE, J. P., "Stability of nonautonomous systems", *Nonlinear Analysis, Theory, Method and Applications* **1** (1976) 83—91.
[24] LA SALLE, J. P., and LEFSCHETZ, S., *Stability by Liapunov's direct method with applications* (Academic Press, New York, 1961).
[25] MASSERA, J. L., "Contributions to stability theory", *Ann. of Math.* **64** (1956) 182—206.
[26] ONUCHIC, N., ONUCHIC, L. R. and TABOAS, P., "Invariance properties in the theory of stability for ordinary differential systems and applications, *Applicable Analysis* **5** (1975) 101—107.
[27] PONTRJAGIN, L. Sz., *Közönséges differenciálegyenletek* (Tankönyvkiadó, Budapest, 1974).
[28] ROUCHE, N., HABETS, P. and LALOY, M., *Stability theory by Liapunov's direct method* (Applied Mathematical Sciences, vol. 22, Springer-Verlag, New York—Heidelberg—Berlin, 1977).
[29] SALVADORI, L., "Sull' estensione ai sistemi dissipativi del criterio di stabilità del Routh", *Ricerche Mat.* **15** (1966) 162—167.
[30] SALVADORI, L., "Some contributions to asymptotic stability theory", *Ann. Soc. Sci. Bruxelles Sér. I.* **88** (1974) 183—194.
[31] SZŐKEFALVI-NAGY, B., *Valós függvények és függvénysorok* (Tankönyvkiadó, Budapest, 1954).
[32] YOSHIZAWA, T., *Stability theory by Liapunov's second method* (Math. Soc. of Japan, 1966).
[33] Арнольд, В. И., *Обыкновенные дифференциальные уравнения* (Наука, Москва, 1971).
[34] Барбашин, Е. А., *Введение в теорию устойчивости* (Наука, Москва, 1967).
[35] Боголюбов, Н. Н., Митропольский, Ю. А., *Асимптотические методы в теории нелинейных колебаний* (Физматгиз, Москва, 1958).
[36] Демидович, Б. П., *Лекции по математической теории устойчивости* (Наука, Москва, 1967).

[37] Еругин, Н. П., «К теории канонических систем», *Дифференциальные уравнения* **2** (1966) 1317—1332.
[38] Кордуняну, К., «Применение дефференциальных неравенств к теории устойчивости», *Analele ştiinţ. ale Univ. "Al. I. Cuza din Iasi"* **6** (1960) 47—58.
[39] Красовский, Н. Н., *Некоторые задачи теории устойчивости* (Физматгиз, Москва, 1959).
[40] Лахаданов, В. М., «О влиянии структуры сил на устойчивость движения», *Прикладная математика и механика* **38** (1974) 246—253.
[41] Лахаданов, В. М., «О стабилизации потенциальных систем», *Прикладная математика и механика* **39** (1975) 53—58.
[42] Ляпунов, А. М., *Общая задача об устойчивости движения* (Гостехиздат, Москва, 1950).
[43] Малкин, И. Г., *Теория устойчивости движения* (Наука, Москва, 1966).
[44] Марачков, В. П., «Об одной теореме устойчивости», *Изв. физ.-матем. об.-ва и НИИ математики и механики при Казанском ун-те*, сер. **3, 12** (1940) 171—174.
[45] Матросов, В. М., «Об устойчивости движения», *Прикладная математика и механика* **26** (1962) 885—895.
[46] Меркин, Д. Р., *Введение в теорию устойчивости движения* (Наука, Москва, 1971).
[47] Меркин, Д. Р., *Гироскопические системы* (Наука, Москва, 1974).
[48] Немыцкий, В. В. и Степанов, В. В., *Качественная теория дифференциальных уравнений* (Гостехиздат, Москва, 1949).
[49] Озиранер, А. С. и Румянцев, В. В., «Метод функций Ляпунова в задаче об устойчивости движения относительно части переменных», *Прикладная математика и механика* **39** (1972) 364—384.
[50] Паламодов, В. П., «Об устойчивости равновесия в потенциальном поле», *Функциональный анализ и его приложения* **11** (1974) вып. 4, 42—55.
[51] Румянцев, В. В., «Об устойчивости движения по отношению к кчасти переменных», *Вестник МГУ* 1957 вып. **4**, 9—16.
[52] Хатвани, Л., «Об отсутствии асимптотической устойчивости по части переменных», *Прикладная математика и механика* **40** (1976) 245—251.
[53] Четаев, Н. Г., *Устойчивость движения* (Гостехиздат, Москва, 1956).

(Beérkezett: 1979. május 24.)

HATVANI LÁSZLÓ
JÓZSEF ATTILA TUDOMÁNYEGYETEM, BOLYAI INTÉZET
6720 SZEGED, ARADI VÉRTANÚK TERE 1.

STABILITY AND PARTIAL STABILITY OF NONAUTONOMOUS SYSTEMS OF DIFFERENTIAL EQUATIONS

L. HATVANI

In this paper a survey of Liapunov's direct method is given. The *Lagrange—Dirichlet-theorem* and its reversibility are treated as an application of the basic theorems of the theory. We familiarize the main stability theorems using the first approximation. From the modern results several theorems on the asymptotic stability by Liapunov functions having semidefinite derivative, or being unbounded from above are proved. We discuss the effect of dissipative and gyroscopic forces produced on the stability of the equilibrium position of mechanical systems.

Alkalmazott Matematikai Lapok 5 (1979) 49—61

A REKESZRENDSZEREK INVERZ FELADATÁRÓL

TÓTH JÁNOS és HÁRS VERA
Budapest

Szükséges és elégséges feltételt adunk arra, hogy egy lineáris, állandó együtthatós differenciálegyenlet-rendszer egy rekeszrendszer determinisztikus modellje legyen. Ezen feltétel teljesülése esetén megadjuk a különböző, de lényegében ugyanazon differenciálegyenlet-rendszerrel leírható rekeszrendszerek számát zárt és nyílt rekeszrendszerek esetén.

1. Bevezetés

Az alábbiakban a reakciókinetika *inverz feladata* [l, 154. old.; 4, 1., 102., 152., 187., 211. old.; 8, 303—307. old.] egy speciális esetének egy lépésével foglalkozunk: meghatározzuk, hogy egy adott lineáris differenciálegyenlet-rendszerhez milyen feltételek esetén létezik olyan (tömeghatás kinetikájú) rekeszrendszer — és, ha létezik, hány — amelynek determinisztikus modellje éppen az adott differenciálegyenlet-rendszer. A jelen dolgozatban röviden ezt a lépést — a differenciálegyenlet-rendszerről az összetett kémiai reakcióra való következtetést — fogjuk a reakciókinetika inverz feladatának nevezni. Az irodalomban elsősorban egy másik lépéssel — a differenciálegyenlet-rendszerről a sebességi állandókra való következtetéssel szokás foglalkozni, lásd például [1]-et, [4]-et, [5]-öt; és ezek irodalomjegyzékét.

Az inverz feladat megoldásának kettős jelentősége van. *Gyakorlati* szempontból érdekes az a kérdés, hogy ha mérési adatokhoz sikerült egy adott típusú differenciálegyenlet-rendszert illeszteni, akkor származtatható-e ez a differenciálegyenlet-rendszer összetett kémiai reakció vagy reakciók modelljeként. *Elméleti* szempontból pedig azért jó tudni, hogy egy differenciálegyenlet-rendszerhez van-e reakció, mert reakciókra vonatkozó differenciálegyenlet-rendszerek kvalitatív tulajdonságaira (aszimptotikus stabilitás, multistacionaritás) egészen meglepő és erős tételek ismeretesek, mint például a zéró deficiencia tétel, vagy *Volpert tételei* (lásd például [8] irodalomjegyzékét).

Először bevezetünk néhány definíciót abból a célból, hogy feladatunkat pontosan megfogalmazhassuk. Másodszor — kizárólag elemi kombinatorikai eszközök felhasználásával — rekeszrendszerek ekvivalenciáját tanulmányozzuk. Harmadszor (triviális) szükséges és elégséges feltételt adunk az inverz feladat megoldhatóságára, majd megoldjuk azt. Legvégül néhány példát mutatunk, amelyek a megelőző definíciók és tételek jelentésének megvilágításához is hozzájárulhatnak.

2. Jelölések és definíciók

Az egyértelműség kedvéért teljesen formális definíciókat adunk, amelyek jelentését részletesebben taglalja például [2], [3] és [8]. Itt jegyezzük meg, hogy a formális definíciók és a dolgozat felépítésének szerkezete részben egy előkészületben levő, nem csak rekeszrendszerekkel foglalkozó cikk céljait szolgálják.

Az alábbiakban N, N_0, R, R^+ és $\overline{R^+}$ a pozitív egész, a nemnegatív egész, a valós, a pozitív valós és a nemnegatív valós számokat, $\mathscr{D}_f$ pedig az f függvény értelmezési tartományát jelöli.

Legyen $M, N \in N$; $\mathscr{S}$ egy tetszőleges M elemű halmaz: $\mathscr{S} =: \{\mathscr{A}(1), \ldots, \mathscr{A}(M)\}$. ($\mathscr{A}(m)$ neve: az m-edik kémiai komponens jele; $m = 1, 2, \ldots, M$.) Tegyük fel, hogy adott a $\mathscr{T} =: \{\mathscr{C}(1), \ldots, \mathscr{C}(N)\} \subset N_0^{\mathscr{S}}$ N elemű halmaz, amelynek elemeit *komplexeknek* nevezzük és az $\mathscr{R} \subset \mathscr{T} \times \mathscr{T}$ halmaz, amelyeknek elemeit reakcióknak hívjuk. Ha $(\mathscr{C}(i), \mathscr{C}(j)) \in \mathscr{R}$, akkor inkább ezt írjuk: $\mathscr{C}(i) \to \mathscr{C}(j)$ és azt mondjuk, hogy a $\mathscr{C}(i)$ *reaktáns* komplex átalakul a $\mathscr{C}(j)$ *termék* komplexszé. Feltehető, hogy minden komplex szerepel reakcióban, azaz minden $n \in \{1, 2, \ldots, N\}$-hez létezik olyan $i \in \{1, 2, \ldots, N\}$, hogy vagy $\mathscr{C}(i) \to \mathscr{C}(n)$ vagy $\mathscr{C}(n) \to \mathscr{C}(i)$ teljesül.

Az $R^{\mathscr{S}}$ lineáris tér *természetes bázisa:*

$$\mathscr{B} := \left\{\omega_{\mathscr{A}(m)};\ \omega_{\mathscr{A}(m)} \in R^{\mathscr{S}},\ \omega_{\mathscr{A}(m)}(\mathscr{A}(i)) = \delta_{im};\ m, i = 1, 2, \ldots, M\right\}$$

(ahol δ_{im} a *Kronecker-szimbólum*).

Ezekkel a függvényekkel kifejezhetjük bármelyik $\mathscr{C}(n) \in \mathscr{T}$-t:

$$\mathscr{C}(n) =: \sum_{m=1}^{M} y^m(n)\, \omega_{\mathscr{A}(m)}.$$

Feltesszük, mint szokásos, hogy a most definiált együtthatókra: $y^m(n) \in N_0$ ($m \in \{1, 2, \ldots, M\}$, $n \in \{1, 2, \ldots, N\}$). Ez utóbbi kifejezés helyett a kémiában szinte mindig, lineáris algebrában néha a következőt írják ([2], 85. old.):

$$\mathscr{C}(n) = \sum_{m=1}^{M} y^m(n)\, \mathscr{A}(m),$$

s a továbbiakban mi is azonosítjuk $R^{\mathscr{S}}$ természetes bázisának elemeit $\mathscr{S}$ elemeivel.

Legyenek az $y(n)$ *komplex vektorok* a következők:

$$y(n) := (y^1(n), \ldots, y^M(n))^T \in N_0^M,$$

amelyek koordinátái tehát a komplexek együtthatói a természetes bázisban. (Ezek kémiai elnevezése: *sztöchiometriai együttható.*) Legyen továbbá

$$Y := (y(1), \ldots, y(N))$$

a *komplex mátrix.* (Erről feltehető, hogy nincs csupa 0-kból álló sora, azaz hogy mindegyik kémiai komponens szerepel valamelyik komplexben. Csupa 0-kból álló oszlopa lehet, ez felel meg az *üres komplexnek,* amelyet $\mathcal{O}$-val jelölünk.

Legyen $V := R^M$ a *komponenstér* ($V^+ := (R^+)^M$), $W := R^N$ pedig a *komplextér,* R pedig a sebességeket megadó függvény (neve: *kinetika*):

$$\overline{(R^+)^M} =: \overline{V^+} \ni x \mapsto R(x) \in R^{N \times N}.$$

Minden $x \in \mathcal{D}_R$ esetén az $R(x)$ *sebességi mátrix* $r_{ij}(x) := (R(x))_{ij}$ elemeire teljesülnek az alábbiak:

i) $r_{ij}(x) \geqq 0 \quad (i, j = 1, 2, \ldots, N)$

ii) $r_{ii}(x) = 0 \quad (i = 1, 2, \ldots, N).$

Azt mondjuk, hogy a $\mathscr{C}(j) \to \mathscr{C}(i)$ reakció sebessége $r_{ij}(x)$, ha az $\mathscr{A}(1), \ldots, \mathscr{A}(M)$ komponensek koncentrációjából álló vektor x.

2.1. DEFINÍCIÓ. *Összetett kémiai reakciónak* vagy *mechanizmusnak* nevezzük a fentiekben leírt tulajdonságú elemekből álló $\mathscr{M} = \langle M, N, \mathscr{S}, \mathscr{T}, \mathscr{R}, R \rangle$ objektumot.

2.2. DEFINÍCIÓ. Azt mondjuk, hogy az $\mathscr{M}$ összetett kémiai reakció *tömeghatás kinetikájú,* ha van olyan $K \in \boldsymbol{R}^{N \times N}$, nemnegatív elemekkel bíró, nulla főátlójú mátrix, amellyel minden $i, j = 1, 2, \ldots, N$ esetén

$$r_{ij}(x) = k_{ij} x^{y(j)}$$

teljesül, ahol k_{ij} a K mátrix i-edik sorának j-edik eleme, a jobb oldalon szereplő második tényezőt pedig a szokásos módon ([3], 90. oldal) definiáljuk:

$$x^{y(j)} := \prod_{m=1}^{M} x_m^{y_m(j)}.$$

2.3. DEFINÍCIÓ. Az $\mathscr{M}$ mechanizmus (folytonos idejű, folytonos állapotterű) *determinisztikus modelljének* vagy *kinetikai differenciálegyenletének* nevezzük az alábbi explicit, közönséges elsőrendű differenciálegyenletet:

$$\dot{x}(t) = Y\left[R(x(t)) - R(x(t))^T\right] 1_N$$

$$(t \in \mathscr{D}_x; \quad 1_N := (1, \ldots, 1)^T \in \boldsymbol{R}^N).$$

Belátható, hogy tömeghatás kinetikájú mechanizmus esetén a kinetikai differenciálegyenlet alakja:

$$\dot{x}(t) = Y[K - \mathrm{dg}(K^T 1_N)] x(t)^Y,$$

ahol egyrészt

$$\boldsymbol{R}^N \ni z \mapsto \mathrm{dg}\, z := \begin{bmatrix} z_1 & 0 & \ldots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \ldots & 0 & z_N \end{bmatrix} \in \boldsymbol{R}^{N \times N},$$

másrészt

$$(x(t))^Y := \left((x(t))^{y(1)} \ldots (x(t))^{y(N)}\right)^T.$$

A jelen dolgozatban *csak* tömeghatás kinetikájú mechanizmusokkal foglalkozunk. Az ilyeneket $M, N, \mathscr{S}, \mathscr{T}, \mathscr{R}$ és K határozza meg, s ezért ezek jelölésére alkalmazni fogjuk az $\langle M, N, \mathscr{S}, \mathscr{T}, \mathscr{R}, K \rangle$ jelölést.

2.4. DEFINÍCIÓ. Az $\mathscr{M} = \langle M, N, \mathscr{S}, \mathscr{T}, \mathscr{R}, K \rangle$ összetett kémiai reakciót *rekeszrendszernek* (komponenseit néha *rekeszeknek*) nevezzük, ha $M \leqq N \leqq M + 1$, és $\mathscr{T} \subset \mathscr{B} \cup \{\mathscr{O}\}$. A rekeszrendszer

i) *zárt*, ha $\mathcal{T}=\mathcal{B}$;

ii) *szigorúan félig nyílt*, ha $\mathcal{T}=\mathcal{B}\cup\{\mathcal{O}\}$ és létezik olyan $m\in\{1, 2, ..., M\}$ szám, amelyre[1] $k_{0m}>0$, és minden $m\in\{1, 2, ..., M\}$ esetén $k_{m0}=0$;

iii) *szigorúan nyílt*, ha $\mathcal{T}=\mathcal{B}\cup\{\mathcal{O}\}$ és létezik legalább egy m szám, amelyre $k_{m0}>0$.

Szemléletesen: zárt a rekeszrendszer, ha a külvilágból nem lép be anyag a rendszerbe és a külvilágba nem lép ki anyag a rendszerből. Szigorúan félig nyílt rekeszrendszerbe nem kerül anyag a külvilágból, de a rendszerből távozik anyag a külvilágba. Szigorúan nyílt rekeszrendszerbe lép be anyag a külvilágból és a rendszerből távo*zhat* anyag a külvilágba.

Nem kapnánk általánosabb definíciókat, ha csak $\mathcal{T}\subset\mathcal{B}$-t, illetve $\mathcal{T}\subset\mathcal{B}\cup\{\mathcal{O}\}$-t követelnénk meg i)—iii)-nál, ugyanis feltettük, hogy minden komponens szerepel valamelyik komplexben.

Mivel $\mathcal{B}$ elemeit azonosítottuk $\mathcal{S}$ elemeivel, ezért a feltételeket $\mathcal{B}$ helyett $\mathcal{S}$-sel is megfogalmazhattuk volna.

Hangsúlyozzuk azt a tényt, hogy a rekeszrendszerek csak *speciális esetét* képezik az elsőrendű összetett kémiai reakcióknak, vagyis azoknak, amelyeknél a reaktáns komplexekről tesszük csak föl, hogy a bázis elemeivel vagy az üres komplexszel azonosak.

2.5. DEFINÍCIÓ. Egy rekeszrendszer azon $\mathcal{A}(m)$ rekeszét, amelyre

i) $k_{im}=0$ $(i=0, 1, ..., M)$ a rekeszrendszer *elsőrendű végpontjának* nevezzük, és az ilyenek számát R_1-gyel jelöljük;

ii) $k_{0m}>0$, a rekeszrendszer *másodrendű végpontjának* nevezzük, és az ilyenek számát R_2-vel jelöljük;

iii) $k_{m0}>0$, a rekeszrendszer *belépési pontjainak* nevezzük, és az ilyenek számát S-sel jelöljük.

Nyilvánvaló, hogy R_1 egyenlő K nulla oszlopvektorainak számával, R_2 egyenlő K nulladik sora pozitív elemeinek számával és S egyenlő K nulladik oszlopa pozitív elemeinek számával.

Az is nyilvánvaló, hogy a komplexeket természetes módon számozva a háromféle típusú rekeszrendszer determinisztikus modellje:

i) zárt rekeszrendszeré:

$$\dot{x}_i(t)=\left(-\sum_{j=1}^{M}k_{ji}\right)x_i(t)+\sum_{j=1}^{M}k_{ij}x_j(t), \tag{2.1}$$

$$\left(i=1, 2, ..., M;\quad t\in\overline{\boldsymbol{R}^+}\right);$$

ii) szigorúan félig nyílt rekeszrendszeré:

$$\dot{x}_i(t)=\left(-\sum_{j=0}^{M}k_{ji}\right)x_i(t)+\sum_{j=1}^{M}k_{ij}x_j(t), \tag{2.2}$$

$$\left(i=1, 2, ..., M;\quad \exists i: k_{0i}>0;\quad t\in\overline{\boldsymbol{R}^+}\right);$$

[1] Most az üres komplexet a nulladiknak tekintjük.

iii) szigorúan nyílt rekeszrendszeré:

$$\dot{x}_i(t) = \left(-\sum_{j=0}^{M} k_{ji}\right) x_i(t) + \sum_{j=1}^{M} k_{ij} x_j(t) + k_{i0}, \tag{2.3}$$

$$(i = 1, 2, \ldots, M; \quad \exists i: k_{i0} > 0; \quad t \in \overline{R^+}).$$

Feladatunk most már pontosabban megfogalmazva, így hangzik: Ha adott az

$$\dot{x}_i(t) = \sum_{j=1}^{M} a_{ij} x_j(t) + b_i \tag{2.4}$$

$$(i = 1, 2, \ldots, M; \quad a_{ij}, b_i \in \boldsymbol{R}; \quad t \in \overline{R^+})$$

differenciálegyenlet-rendszer, akkor van-e hozzá olyan rekeszrendszer (és ha igen, hány), amelynek ez a determinisztikus modellje.

Bizonyos, nem lényegesen különböző rekeszrendszereket, illetve differenciálegyenlet-rendszereket érdemes azonban azonosnak tekinteni. Ehhez további definíciókra és néhány lemmára lesz szükségünk.

3. Ekvivalens rekeszrendszerek és differenciálegyenlet-rendszerek

3.1. DEFINÍCIÓ. Egy *rekeszrendszer magja* az a rekeszrendszer, amelyet az eredetiből az összes elsőrendű végpont elhagyásával kapunk.

3.2. DEFINÍCIÓ. A (2.4) *differenciálegyenlet-rendszer magja* az a differenciálegyenlet-rendszer, amelyet az $(a_{ij})_{i,j=1}^{M}$ mátrix nulla oszlopaival azonos indexű változók elhagyásával nyerünk.

3.1. LEMMA. Egy rekeszrendszer magjának differenciálegyenlet-rendszere magja a rekeszrendszer differenciálegyenlet-rendszerének.

Két ekvivalenciarelációt definiálunk:

3.3. DEFINÍCIÓ. Két lineáris differenciálegyenlet-rendszert (két rekeszrendszert) *ekvivalensnek* nevezünk, ha magjuk közös.

3.2. LEMMA. Tekintsünk egy M rekeszből álló zárt rekeszrendszert, amelyben R_1 számú elsőrendű végpont van! Ehhez $\binom{R_1}{K_1}$ számú olyan, vele ekvivalens, szigorúan félig nyílt rekeszrendszer létezik, amelyet az eredetiből K_1 számú ($K_1 \in \{1, 2, \ldots, R_1\}$) elsőrendű végpont elhagyásával kapunk. Az új rekeszrendszer determinisztikus modellje $M - K_1$ számú függvényt tartalmaz.

Bizonyítás: Ha egy zárt rekeszrendszer elsőrendű végpontját elhagyjuk, az eredetivel ekvivalens, szigorúan félig nyílt rekeszrendszert kapunk. Az R_1 elsőrendű végpont közül $\binom{R_1}{K_1}$-féleképpen választhatjuk ki a K_1 számú elhagyandó végső rekeszt.

3.3. LEMMA. Tekintsünk egy M rekeszből álló szigorúan félig nyílt rekeszrendszert, amelyben R_2 számú másodrendű végpont van! Ehhez

$$f(R_2, K_2) := \sum_{k=1}^{K_2} \binom{K_2}{k} (-1)^{K_2-k} k^{R_2} \tag{3.1}$$

számú olyan, vele ekvivalens, zárt rekeszrendszer létezik, amelyet az eredetiből úgy kapunk, hogy a másodrendű végpontokat elsőrendű végpont hozzátételével megszüntetjük összesen K_2 számú ($K_2 \in \{1, 2, \ldots, R_2\}$) további rekesz felhasználásával. Az új rekeszrendszer determinisztikus modellje $M+K_2$ számú függvényt tartalmaz.

Bizonyítás: K_2 számú rekeszt kell hozzáillesztenünk R_2 számú másodrendű végponthoz úgy, hogy mindegyik rekeszt felhasználjuk. A logikai szitaformulát fogjuk alkalmazni a lemma bizonyítására. (Lásd például: [6], 1.8. Feladat.) Legyen S bizonyos objektumok véges halmaza és (egy rögzített $k \in N$ mellett) $T_1, T_2, \ldots, \ldots, T_k \subset S$ *tulajdonságok* (azaz ha $s \in T_i$, $i \in \{1, 2, \ldots, k\}$, akkor azt mondjuk, hogy s rendelkezik az i-edik tulajdonsággal). Jelölje a tetszőleges $A \subset S$ halmaz elemeinek számát $|A|$, S-re vonatkozó komplementumát $\bar{A}$. Ekkor

$$\left| \bigcap_{i=1}^{k} \bar{T}_i \right| = |S| - \sum_{i=1}^{k} |T_i| + \sum_{\substack{i,j=1 \\ i \neq j}}^{k} |T_i \cap T_j| - \ldots + (-1)^k \left| \bigcap_{i=1}^{k} T_i \right|.$$

Tekintsük most az M rekeszből álló szigorúan félig nyílt rendszert! Kiszámítjuk azon esemény $f(R_2, K_2)$-vel jelölt gyakoriságát, hogy az R_2 számú másodrendű végpont mindegyikéhez illesztünk rekeszt és mind a K_2 számú új rekeszt felhasználjuk. Legyen a szita-formulában szereplő S halmaz azon rekeszrendszerek halmaza, amelyekben az R_2 számú másodrendű végpont közül bizonyosakhoz hozzáillesztettünk a K_2 számú új rekesz közül bizonyosakat. Mondjuk azt, hogy egy rekeszrendszer rendelkezik az i-edik tulajdonsággal (azaz eleme T_i-nek), ha az összes másodrendű végpontjához illesztettünk új rekeszt, de csak K_2-1 számú új rekeszt használtunk fel, az $(M+i)$-ediket nem. Ekkor

$$f(R_2, K_2) = \left| \bigcap_{i=1}^{K_2} \bar{T}_i \right| = K_2^{R_2} - \sum_{i=1}^{K_2} |T_i| + \sum_{\substack{i,j=1 \\ i \neq j}}^{K_2} |T_i \cap T_j| - \ldots + (-1)^{K_2} \left| \bigcap_{i=1}^{K_2} T_i \right|, \tag{3.2}$$

ahol

$$\left| \bigcap_{j=1}^{k} T_{ij} \right| = (K_2 - k)^{R_2} \quad (1 \leqq k \leqq K_2), \tag{3.3}$$

és itt a jobb oldal értéke csak k-tól függ. Ha (3.2)-t és (3.3)-at összevetjük, akkor egyszerű átalakítások révén (3.1)-hez juthatunk.

Megjegyezzük, hogy (3.1) speciális esete a következő összefüggés:

$$f(R_2, R_2) = R_2! = \sum_{k=1}^{R_2} \binom{R_2}{k} (-1)^{R_2-k} k^{R_2}.$$

3.4. LEMMA. Tekintsünk egy M rekeszből álló szigorúan félig nyílt rekeszrendszert, amelyben R_i számú i-edrendű végpont van $(i=1, 2)$! Ehhez

$$\binom{R_1}{K_1}\binom{R_2}{J} f(J, K_2)$$

$$(K_1 \in \{0, 1, \ldots, R_1\};\ J \in \{1, 2, \ldots, R_2 - (1 - \operatorname{sign} K_1)\};\ K_2 \in \{1, 2, \ldots, J\})$$

számú olyan, vele ekvivalens, szigorúan félig nyílt rekeszrendszer létezik, amelyet úgy kapunk, hogy K_1 elsőrendű végpontot elhagyunk és K_2 új rekeszt illesztünk J számú rögzített másodrendű végponthoz. Az új rekeszrendszer determinisztikus modellje $M + K_2 - K_1$ számú függvényt tartalmaz.

Bizonyítás: Az állítások $K_1 > 0$ esetén az előző két lemmából következnek. Ha $K_1 = 0$, vagyis ha egyetlen rekeszt sem hagyunk el, akkor $J \leqq R_2 - 1$, mivel, ha ugyanekkor $J = R_2$ volna, azaz minden másodrendű végponthoz illesztenénk új rekeszt, zárt rekeszrendszert kapnánk. Az eredeti és az új rekeszrendszer magja közös, tehát ekvivalensek.

3.5. LEMMA. Tekintsünk egy M rekeszből álló szigorúan nyílt rekeszrendszert! Legyen

$$K_1 \in \{0, 1, \ldots, R_1\};\quad J \in \{1, 2, \ldots, R_2\};\quad K_2 \in \{1, 2, \ldots, J\}$$

és jelentésük legyen ugyanaz, mint az előzőekben! Ekkor létezik hozzá

$$\binom{R_1}{K_1}\binom{R_2}{J} f(J, K_2)$$

számú, vele ekvivalens, szigorúan nyílt rekeszrendszer, amelynek determinisztikus modellje $M + K_2 - K_1$ függvényt tartalmaz.

Bizonyítás: Az állítás az előző lemmákból következik. A $K_1 = 0$, $J = R_2$ esetet most nem kell kihagynunk, mivel egy szigorúan nyílt rekeszrendszer akkor is szigorúan nyílt marad, ha másodrendű végpontjainak számát 0-ra csökkentjük.

3.6. LEMMA. Tegyük fel, hogy az előző négy lemma megfelelő feltételei teljesülnek egy zárt, szigorúan félig nyílt, illetve szigorúan nyílt rekeszrendszerre. Ekkor K_1, K_2, illetve J összes lehetséges értékére egy M rekeszből álló

i) zárt rekeszrendszerből összesen

$$e(R_1) := 2^{R_1} - 1$$

számú olyan szigorúan félig nyílt rekeszrendszert;

ii) szigorúan félig nyílt rekeszrendszerből összesen

$$g(R_2) := \sum_{K_2=1}^{R_2} f(R_2, K_2)$$

számú, olyan zárt rekeszrendszert;

iii) szigorúan félig nyílt rekeszrendszerből összesen

$$h(R_1, R_2) := 2^{R_1} \sum_{J=1}^{R_2} \binom{R_2}{J} g(J) - g(R_2)$$

számú, olyan szigorúan félig nyílt rekeszrendszert;

iv) szigorúan nyílt rekeszrendszerből összesen

$$j(R_1, R_2) := h(R_1, R_2) + g(R_2) = 2^{R_1} \sum_{J=1}^{R_2} \binom{R_2}{J} g(J)$$

számú, olyan szigorúan nyílt rekeszrendszert kaphatunk, amely ekvivalens az eredetivel.

Bizonyítás:

i) A 3.2. lemmát alkalmazva kapjuk:

$$e(R_1) = \sum_{K_1=1}^{R_1} \binom{R_1}{K_1} = 2^{R_1} - 1.$$

ii) Összegezzük a 3.3. lemmában kapott eredményt K_2 összes lehetséges értékeire.

iii) Összegezzük a 3.4. lemmában kapott eredményt és vegyük figyelembe, hogy ha $K_1 = 0$, akkor $J \leqq R_2 - 1$:

$$\sum_{K_1=0}^{R_1} \sum_{J=1}^{R_2} \sum_{K_2=1}^{J} \binom{R_2}{J} \binom{R_1}{K_1} f(J, K_2) - g(R_2) = \sum_{K_1=0}^{R_1} \sum_{J=1}^{R_2} g(J) \binom{R_2}{J} \binom{R_1}{K_1} - g(R_2) =$$

$$= 2^{R_1} \sum_{J=1}^{R_2} \binom{R_2}{J} g(J) - g(R_2).$$

iv) Összegezzük a 3.5. lemmában kapott eredményt:

$$\sum_{K_1=0}^{R_1} \sum_{J=1}^{R_2} \sum_{K_2=1}^{J} \binom{R_1}{K_1} \binom{R_2}{J} f(J, K_2) = 2^{R_1} \sum_{J=1}^{R_2} \binom{R_2}{J} g(J).$$

Megjegyezzük, hogy másféle átmenet a rekeszrendszerek különféle típusai között nincs, tehát például szigorúan nyílt rekeszrendszerből nem kaphatunk zártat, s í.t.

4. Az inverz feladat megoldása

Ha az adott (2.4) differenciálegyenlet-rendszer éppen M számú ismeretlen függvényt tartalmaz, akkor a (2.1), (2.2), (2.3) differenciálegyenlet-rendszereket (2.4)-gyel összehasonlítva a következő szükséges és elégséges feltételt nyerjük az inverz feladat megoldhatóságára vonatkozóan:

4.1. TÉTEL. A (2.4) differenciálegyenlet-rendszer akkor és csak akkor determinisztikus modellje egy M rekeszből álló

i) zárt,
ii) szigorúan félig nyílt, illetve
iii) szigorúan nyílt

rekeszrendszernek, ha

i) $b_i = 0, \quad a_{ij} \geqq 0 \ (i \neq j), \quad a_{ii} = -\sum_{\substack{j=1 \\ j \neq i}}^{M} a_{ji}$

és ekkor

$$k_{ij} = a_{ij} \quad (i \neq j);$$

ii) $b_i = 0, \quad a_{ij} \geqq 0 \ (i \neq j), \quad a_{ii} \leqq -\sum_{\substack{j=1 \\ j \neq i}}^{M} a_{ji}$

$$\exists i: \ b_{ii} < -\sum_{\substack{j=1 \\ j \neq i}}^{M} a_{ji}$$

és ekkor

$$k_{ij} = a_{ij} \ (i \neq j), \quad k_{0i} = -a_{ii} - \sum_{\substack{j=1 \\ j \neq i}}^{M} a_{ji};$$

iii) $b_i \geqq 0, \quad a_{ij} \geqq 0 \ (i \neq j), \quad a_{ii} \leqq -\sum_{\substack{j=1 \\ j \neq i}}^{M} a_{ji}$

$$\exists i: \ b_i > 0,$$

és ekkor

$$k_{ij} = a_{ij} \ (i \neq j), \quad k_{0i} = -a_{ii} - \sum_{\substack{j=1 \\ j \neq i}}^{M} a_{ji}, \quad k_{i0} = b_i$$

(mindvégig $i, j = 1, 2, \ldots, M$).

4.2. TÉTEL. Tegyük fel, hogy a (2.4) differenciálegyenlet-rendszer kielégíti a 4.1. tétel i) pontjának feltételeit! Ekkor a 4.1. tétel i) pontjában definiált, M rekeszből álló rekeszrendszeren kívül még $2^{R_1} - 1$ olyan szigorúan félig nyílt rekeszrendszer létezik, amely $M - K_1$ ($K_1 \in \{1, 2, \ldots, R_1\}$) rekeszből áll és amelynek determinisztikus modellje ekvivalens az eredeti differenciálegyenlet-rendszerrel.

Bizonyítás: Az állítás a 4.1. tétel i) pontjából, a 3.6. lemma i) pontjából és — a 3.3. definíció felhasználásával — a 3.1. Lemmából következik.

4.3. TÉTEL. Tegyük fel, hogy a (2.4) differenciálegyenlet-rendszer kielégíti a 4.1. tétel ii) pontjának feltételeit! Ekkor a 4.1. tétel ii) pontjában definiált, M rekeszből álló rekeszrendszeren kívül még $g(R_2)$ számú olyan zárt rekeszrendszer létezik, amely $M + K_2 (K_2 \in \{1, 2, \ldots, R_2\})$ rekeszből áll és amelynek determinisztikus modellje ekvivalens az eredeti differenciálegyenlet-rendszerrel.

Bizonyítás: Az állítás a 4.1. tétel ii) pontjából, a 3.6. lemma ii) pontjából és — a 3.3. definíció felhasználásával — a 3.1. lemmából következik.

4.4. TÉTEL. Tegyük fel, hogy a (2.4) differenciálegyenlet-rendszer kielégíti a 4.1. tétel ii) pontjának feltételeit! Ekkor a 4.1. tétel ii) pontjában definiált, M rekeszből álló rekeszrendszeren kívül még $h(R_1, R_2)$ számú olyan zárt rekeszrendszer létezik, amely $M + K_2 - K_1$ ($K_1 \in \{0, 1, \ldots, R_1\}$, $K_2 \in \{1, 2, \ldots, R_2\}$) rekeszből áll, és amelynek determinisztikus modellje ekvivalens az eredeti differenciálegyenlet-rendszerrel.

Bizonyítás: Az állítás a 4.1. tétel ii) pontjából, a 3.6. lemma iii) pontjából és — a 3.3. definíció felhasználásával — a 3.1. lemmából következik.

4.5. TÉTEL. Tegyük fel, hogy a (2.4) differenciálegyenlet-rendszer kielégíti a 4.1. tétel iii) pontjának feltételeit! Ekkor a 4.1. tétel iii) pontjában definiált, M rekeszből álló rekeszrendszeren kívül még $j(R_1, R_2)$ számú olyan zárt rekeszrendszer létezik, amely $M+K_2-K_1$ ($K_1 \in \{0, 1, \ldots, R_1\}$, $K_2 \in \{1, 2, \ldots, R_2\}$) rekeszből áll és amelynek determinisztikus modellje ekvivalens az eredeti differenciálegyenlet-rendszerrel.

Bizonyítás: Az állítás a 4.1. tétel iii) pontjából, a 3.6. lemma iv) pontjából és — a 3.3. definíció felhasználásával — a 3.1. lemmából következik.

5. Példák

A példák a fogalmak és tételek illusztrálására szolgálnak, és egyben azt is megmutatják, hogy miként lehet „kémiai közvetítéssel" egy differenciálegyenlet-rendszer stabilitására vonatkozó állításokat nyerni.

1. PÉLDA. Tegyük fel, hogy illesztés eredményeképpen az alábbi differenciálegyenlet-rendszert kaptuk [5, 266. old.]:

$$\begin{aligned} \dot{x}_1 &= -p_2 x_1 + p_1 x_2 \\ \dot{x}_2 &= p_2 x_1 - (p_1 + p_3) x_2 \\ \dot{x}_3 &= p_3 x_2 - p_4 x_3 \\ \dot{x}_4 &= p_4 x_3 . \end{aligned} \tag{5.1}$$

Ehhez a 4.1. tétel i) pontja az alábbi zárt rekeszrendszert rendeli (a reakciók felsorolásakor minden komponens jelét csak egyszer írjuk fel, a sebességi állandókat pedig a megfelelő reakció fölé írjuk a szokással egyezően):

$$\mathscr{A}(1) \underset{p_1}{\overset{p_2}{\rightleftharpoons}} \mathscr{A}(2) \xrightarrow{p_3} \mathscr{A}(3) \xrightarrow{p_4} \mathscr{A}(4). \tag{5.2}$$

Itt $M=4$ és $R_1=1$. (5.2)-ből $2^{R_1}-1=1$ szigorúan félig nyílt rekeszrendszer nyerhető:

$$\mathscr{A}(1) \underset{p_1}{\overset{p_2}{\rightleftharpoons}} \mathscr{A}(2) \xrightarrow{p_3} \mathscr{A}(3) \xrightarrow{p_4} \mathscr{O}. \tag{5.3}$$

Az (5.1) differenciálegyenlet-rendszer magja az első három egyenletből áll, az (5.2) rekeszrendszer magja az (5.3) rekeszrendszer.

2. PÉLDA. Tekintsük az alábbi szigorúan félig nyílt rekeszrendszert:

$$\begin{array}{ccccccccc} \mathscr{A}(1) & \leftarrow & \mathscr{A}(2) & \leftarrow & \mathscr{A}(3) & \leftarrow & \mathscr{A}(4) & \rightarrow & \mathscr{A}(5) \rightarrow \mathscr{A}(6) \\ & \searrow & \swarrow & & & & & & \downarrow \\ & & \mathscr{O} & & & & & & \mathscr{A}(7) \end{array} , \tag{5.4}$$

ahol $R_1=2$, $R_2=2$. Azon szigorúan félig nyílt rekeszrendszerek száma, melyek K_1, K_2 és J különböző értékei mellett (5.4)-gyel ekvivalensek, az 1. táblázatban

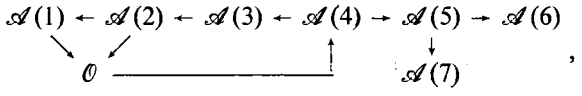
található. Összesen

$$h(2,2) = 2^2 \sum_{J=1}^{2} \binom{2}{J} g(J) - g(2) = 17 \tag{5.5}$$

(5.4)-gyel ekvivalens szigorúan félig nyílt rekeszrendszer létezik.

3. PÉLDA. Tekintsük ezt a szigorúan nyílt rekeszrendszert:

(5.6)



,

ahol $R_1=2$, $R_2=2$, $S=1$. Ez (5.4)-től csak abban különbözik, hogy $\mathscr{A}(4)$ belépési pontja. Ezért a vele ekvivalens szigorúan nyílt rekeszrendszerek az 1. táblázatból összeszámolhatók, de mivel van egy belépési pontja is, az eddigiekhez további ekvivalens rekeszrendszerek járulnak. A $K_1=0$, $K_2=1$, $J=2$ esetben további egy, a $K_1=0$, $K_2=2$, $J=2$ esetben további két rekeszrendszer. Így összesen

$$17+3 = 20 = j(2,2) = h(2,2)+g(2) \tag{5.7}$$

számú (5.6)-tal ekvivalens szigorúan nyílt rekeszrendszer létezik.

Az (5.6) rekeszrendszer determinisztikus modellje:

$$\begin{aligned}
\dot{x}_1 &= -k_{01}x_1 + k_{12}x_2 \\
\dot{x}_2 &= -(k_{02}+k_{12})x_2 + k_{23}x_3 \\
\dot{x}_3 &= -k_{23}x_3 + k_{34}x_4 \\
\dot{x}_4 &= -(k_{34}+k_{54})x_4 + k_{40} \\
\dot{x}_5 &= k_{54}x_4 - (k_{65}+k_{75})x_5 \\
\dot{x}_6 &= k_{65}x_5 \\
\dot{x}_7 &= k_{75}x_5 .
\end{aligned} \tag{5.8}$$

Írjuk most föl az 1. táblázatban szereplő egyik rekeszrendszer determinisztikus modelljét (hozzávéve egy belépési pontot). Hagyjuk el például a 6. rekeszt és csatoljuk az 1. és 2. rekeszekhez a 8. rekeszt. Az így kapott rekeszrendszer determinisztikus modellje a következő:

$$\begin{aligned}
\dot{x}_1 &= -k_{81}x_1 + k_{12}x_2; && k_{81} := k_{01} \\
\dot{x}_2 &= -(k_{82}+k_{12})x_2 + k_{23}x_3; && k_{82} := k_{02} \\
\dot{x}_3 &= -k_{23}x_3 + k_{34}x_4 \\
\dot{x}_4 &= -(k_{34}+k_{54})x_4 + k_{40} \\
\dot{x}_5 &= k_{54}x_4 - (k_{05}+k_{75})x_5; && k_{05} := k_{65} \\
\dot{x}_7 &= k_{75}x_5 \\
\dot{x}_8 &= k_{81}x_1 + k_{82}x_2 .
\end{aligned} \tag{5.9}$$

6. Összefoglalás, további feladatok

Magától értetődő szükséges és elégséges feltételt adtunk arra, hogy egy lineáris, állandó együtthatós differenciálegyenlet-rendszer egy rekeszrendszer determinisztikus modellje legyen. Ezután a feltétel teljesülése esetén megadtuk a különböző, de lényegében ugyanazon differenciálegyenlet-rendszerrel leírható rekeszrendszerek számát zárt és nyílt rekeszrendszerek esetén.

A továbbiakban először két általánosítással szeretnénk foglalkozni: az elsőrendű reakciók és az általánosított rekeszrendszerek ([7]) esetével. A gyakorlatilag legfontosabb eset következhet ezután: az, amelynél a reaktáns komplexek hossza (azaz sztöchiometriai együtthatóinak összege) kettőnél nem nagyobb.

Hasznos lenne a többi modellre (például a folytonos idejű, diszkrét állapotterűre) vonatkozó inverz feladat(ok) megoldása is. Végül gyakorlati szempontból a legfontosabb, elméletileg a legbonyolultabb a hibás mérések esete tetszőleges modellek esetén.

Köszönettel tartozunk kollégáinknak, továbbá dr. Győri Istvánnak a kézirat elolvasásakor tett megjegyzéseiért, valamint az Eü. Min. ETT támogatásáért (1-08-0201-03-0/FS).

1. TÁBLÁZAT

K_1	K_2	J	Ekvivalens szigorúan félig nyílt rekeszrendszerek száma
0	1	1	$\binom{2}{1}\binom{2}{0} f(1,1)=2$
1	1	1	$\binom{2}{1}\binom{2}{1} f(1,1)=4$
1	1	2	$\binom{2}{1}\binom{2}{2} f(2,1)=2$
1	2	2	$\binom{2}{1}\binom{2}{2} f(2,2)=4$
2	1	1	$\binom{2}{2}\binom{2}{1} f(1,1)=2$
2	1	2	$\binom{2}{2}\binom{2}{2} f(2,1)=1$
2	2	2	$\binom{2}{2}\binom{2}{2} f(2,2)=2$

IRODALOM

[1] ÉMANUEL, N. et KNORRE, D., *Cinétique chimique* (Éditions Mir, Moscou, 1975).
[2] FEINBERG, M. and HORN, F. J. M., "Chemical Mechanism Structure and the Coincidence of the Stoichiometric and Kinetic Subspaces", *Arch. Ratl. Mech. Anal.* **66** (1977) 83—97.
[3] HORN, F. and JACKSON, R., "General Mass Action Kinetics", *Arch. Ratl. Mech. Anal.* **47** (1972) 81—116.
[4] JACQUEZ, J. A., *Compartmental Analysis in Biology and Medicine* (Elsevier Publishing Company, Amsterdam, London, New York, 1972).

[5] KANYÁR, B. és TÓTH, J., "Lineáris differenciálegyenlet-rendszer illesztése gradiens módszerrel", *Alk. Mat. Lapok* **2** (1976) 259—268.
[6] LOVÁSZ, L., *Combinatorial Problems and Exercises* (Akadémiai Kiadó, Budapest és North-Holland Publishing Company, Amsterdam, New York, Oxford, 1979).
[7] TÓTH, J., "What is essential to exotic behaviour?", *React. Kinet. Catal. Lett.* **9** (1978) 377—381.
[8] TÓTH, J. és ÉRDI, P.. „A formális reakciókinetika modelljei, problémái és alkalmazásai", *A kémia újabb eredményei* **41** (1978) 226—352.

(Beérkezett: 1979. április 2.)

TÓTH JÁNOS ÉS HÁRS VERA
SEMMELWEIS OTE SZÁMÍTÁSTECHNIKAI CSOPORT
1089 BUDAPEST, VIII., KULICH GY. TÉR 5.

ON THE INVERSE PROBLEM OF COMPARTMENT SYSTEMS

J. TÓTH and VERA HÁRS

A necessary and sufficient condition is given for a linear system of differential equations with constant coefficients to be the deterministic model of a compartment system. In the case when the condition is fulfilled, the number of different (open and closed) compartment systems having essentially the same system of differential equations is given.

Alkalmazott Matematikai Lapok 5 (1979) 63—68

EGY KÉPLÉKENYSÉGTANI VIZSGÁLAT MATEMATIKAI MÓDSZERE

BÉDA GYULA
Budapest

A mechanikai kutatások előterében egyre több olyan vizsgálat jelentkezik, amelyhez véges számú felületen vagy vonalon diszkontinuitásokat mutató függvények szükségesek. Különösen figyelemre méltók azok az esetek, amikor a diszkontinuitást hordozó felület vagy vonal mozog. Erre az esetre vonatkozó általános matematikai összefüggések ismertetése után, matematikai szempontból egyszerű alkalmazásra kerül sor egy képlékenységtani vizsgálat bemutatásával.

Az utóbbi évtizedek mechanikai kutatásaiban mindinkább felmerülnek olyan kérdések, amelyek tisztázásakor egy adott tartományon belül szakadással (diszkontinuitással) rendelkező függvények lépnek fel. Ezek a diszkontinuitások véges számú felületen vagy vonalon jelentkeznek. Különösen fontosak azok az esetek, amikor a diszkontinuitásokat hordozó felületek mozognak. Ilyenkor az ilyen felületet hullámfrontnak is szokás nevezni és aszerint, hogy a diszkontinuitás az alapfüggvénynek tekintett mezőfüggvényben vagy első, illetve magasabb deriváltjaiban jelentkezik, első vagy magasabb rendű hullámról szokás beszélni. Az elmondottak értelmében vett másodrendű hullámok vizsgálatát matematikai szempontból kifogástalan alakban és általánosságban először J. HADAMARD dolgozta ki [1]. HADAMARD nevezetes munkája RIEMANN, CHRISTOFFEL [2] és HUGONIOT [3] munkáira támaszkodik. HADAMARD óta sokáig a hullámokra vonatkozó vizsgálatai mintha feledésbe merültek volna [4]. Az 50-es évektől kezdtek először ERICKSEN [5], majd T. Y. THOMAS [6] és R. HILL [4] munkássága alapján felelevenedni és gazdagodni HADAMARD hullámvizsgálati eredményei és ettől kezdve napjainkig, a mozgó szinguláris felületek vizsgálatára épülve, a mechanika újabb irányba fejlődik. Ezzel továbbfolytatódik a kezdettől fogva meglevő kölcsönhatás a matematikai és mechanikai kutatások között, most a matematika eredményei teszik lehetővé a mechanika fejlődését. A következők HADAMARD vizsgálatának alapgondolatát vázolják, megmutatva annak egy konkrét alkalmazását egy rúd képlékeny dinamikus húzásának matematikai megfogalmazásához szükséges, lehetséges anyagtörvények származtatására.

A mechanikában az alapfüggvénymező a kontinuum sebességmezeje. Az elsőrendű hullám esetében a sebességmezőnek egy mozgó felület mentén diszkontinuitása van, ezt a hullámot ütéshullámnak is szokás nevezni. A másodrendű hullám szokásos neve a gyorsuláshullám, ekkor a sebesség folytonos és véges számú felületet kivéve deriválható, a sebesség első deriváltjainak egy mozgó felület mentén diszkontinuitása van.

Az ütéshullám- és a gyorsuláshullám-vizsgálatot HADAMARD egy lemmája teszi lehetővé [7], [8]. A lemma a következőt mondja ki: legyen egy fizikai $f(x_\alpha)$ mező

egy adott β tartományban értelmezve. Ossza fel ezt a β tartományt egy $\varphi(x_\alpha)=0$ sima felület két részre, egy β^+ és β^--ra, úgy, hogy a β^+ tartományban az f és első deriváltja folytonos és f^+, illetve $f^+_{,\alpha} \equiv \left(\frac{\partial f}{\partial x_\alpha}\right)^+$ értéket veszi fel. Hasonlóképpen a β^--ban az $f_{,\alpha}$ folytonos és f^-, illetve $f^-_{,\alpha}$ értékű. Végül vegyünk fel a $\varphi(x_\alpha)=0$ felületen egy $x_\alpha = x_\alpha(s)$ görbét, akkor

$$\frac{df^+}{ds} = f^+_{,\alpha}\frac{dx_\alpha}{ds},$$

illetve

$$\frac{df^-}{ds} = f^-_{,\alpha}\frac{dx_\alpha}{ds}.$$

Itt és a továbbiakban az *Einstein-konvenciót* használjuk: egy index megjelenése kétszer egy tagban, arra az indexre vonatkozó összegzést jelent.

A lemma következményeként különböző képletek kaphatók. Vezessük be valamely függvény φ menti ugrására a [ ] jelet, tehát az f függvény ugrása:

$$[f] = f^+ - f^-,$$

az $f_{,\alpha}$ ugrása

$$[f_{,\alpha}] = f^+_{,\alpha} - f^-_{,\alpha}.$$

Tegyük fel, hogy a β tartományban, a $\varphi(x_\alpha)=0$ felületet kivéve, az f és $f_{,\alpha}$ folytonos. A lemma értelmében

$$\frac{df^+}{ds} = f^+_{,\alpha}\frac{dx_\alpha}{ds}$$

és

$$\frac{df^-}{ds} = f^-_{,\alpha}\frac{dx_\alpha}{ds}$$

a φ két oldalán. A két egyenlet különbsége, a bevezetett jelölést felhasználva

$$\frac{d}{ds}[f] = [f_{,\alpha}]\frac{dx_\alpha}{ds}.$$

Vagy másképpen felírva

$$(-[f]_{,\alpha} + [f_{,\alpha}])\,dx_\alpha = 0$$

és figyelembe véve, hogy $\varphi_{,\alpha}\,dx_\alpha = 0$, igaz a következő egyenletrendszer

$$[f_{,\alpha}] - [f]_{,\alpha} = \lambda\varphi_{,\alpha}, \tag{1}$$

ahol $\alpha = 1, 2, 3, 4$ és λ megfelelő függvény. Az f és a λ is lehet skalár, vektor vagy tenzor mennyiség. Legyen x_1, x_2, x_3 *Descartes-féle derékszögű koordináta* és

x_4 az idő. Rövidítsük ezeket így x_k $(k=1, 2, 3)$ és t. Vezessük be még a következő jelöléseket

$$[f] \equiv a; \quad \lambda\sqrt{\varphi_{,i}\varphi_{,i}} \equiv b,$$

$$n_k \equiv \frac{\varphi_{,k}}{\sqrt{\varphi_{,i}\varphi_{,i}}} \quad \text{és} \quad \frac{dx_k}{dt} n_k \equiv c_n,$$

$$\text{ahol} \quad (i, k = 1, 2, 3).$$

A bevezetettt jelölések között az n_k a t időpillanatban a hullámfront normális egységvektorának koordinátái, a c_n pedig a terjedési sebesség erre az irányra eső vetülete. Ezek után felírhatjuk az alábbi képleteket

$$[f_{,k}] = bn_k + a_{,k},$$

(2)

$$\left[\frac{\partial f}{\partial t}\right] = -bc_n + \frac{\partial a}{\partial t}.$$

A (2) első egyenletét n_k-val szorozva a

$$b = \left[\frac{\partial f}{\partial n}\right] - \frac{\partial a}{\partial n}$$

egyenlőséget kapjuk, amelyet a (2) második egyenletében felhasználva

$$\left[\frac{\partial f}{\partial t}\right] = -\left[\frac{\partial f}{\partial n}\right] c_n + \frac{\partial a}{\partial n} c_n + \frac{\partial a}{\partial t}$$

adódik. A jobb oldal második és harmadik tagjának összegét az *a Thomas-féle idő szerinti δ deriváltjának* [8] nevezik, azaz

(3)

$$\frac{\delta a}{\delta t} \equiv \frac{\partial a}{\partial n} c_n + \frac{\partial a}{\partial t},$$

vagy

$$\frac{\delta [f]}{\delta t} \equiv \left[\frac{\partial f}{\partial t}\right] + \left[\frac{\partial f}{\partial n}\right] c_n.$$

Ha gyorsuláshullám keletkezik a kontinuumban, akkor $a \equiv [f] \equiv 0$ és ezzel az (1) és (2) képletek

(4)

$$[f_{,\alpha}] = \lambda \varphi_{,\alpha}, \quad [\alpha = 1, 2, 3, 4],$$

illetve

$$[f_{,k}] = bn_k,$$

(5)

$$\left[\frac{\partial f}{\partial t}\right] = -bc_n,$$

vagy

$$[f_{,k}] = \lambda \varphi_{,k},$$

(6)

$$\left[\frac{\partial f}{\partial t}\right] = \lambda \frac{\partial \varphi}{\partial t}.$$

A röviden ismertetett képletek a *Hadamard-lemma* következményei és fontos szerepet játszanak az első és másodrendű hullámok vizsgálatánál. Hasonló képletek vezethetők be a magasabb rendű hullámokra is [8].

A gyorsuláshullámra vonatkozó (6) képlet felhasználásával fontos megállapítást tehetünk a dinamikus képlékeny húzásra vonatkozó lehetséges anyagtörvényekre. Tegyük fel, hogy egy prizmatikus rudat, amelynek tengelye az x tengely, az $x=l$ helyen befogjuk és az $x=0$ helyen olyan módon terhelünk, hogy a rúdban gyorsuláshullám keletkezzen. Számításunkban feltesszük, hogy a terhelés során egytengelyű feszültségi állapot keletkezik. Ennek megfelelően az x, y, z *Descartes-féle derékszögű koordinátarendszerben* a feszültségi tenzor mátrixa

$$\begin{bmatrix} \sigma(x,t) & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

A rúd pontjainak x tengely irányú sebességösszetevője $v(x, t)$ az x irányú fajlagos nyúlás $\varepsilon(x, t)$. Ha az x és t szerinti parciális deriváltat a σ, v és ε mellé indexbe tett x és t-vel jelöljük, akkor a kis képlékeny alakváltozást végző rúd mozgásegyenlete és a kinematikai összeférhetőségi egyenlet

$$\varrho v_t = \sigma_x \tag{7}$$

és

$$\varepsilon_t = v_x, \tag{8}$$

ahol ϱ a rúd tömegsűrűsége. A feladat megfogalmazásához a kezdeti és peremfeltételeken kívül még egy σ, ε-t, illetve deriváltjaikat összekapcsoló egyenletre, az úgynevezett anyagtörvényre van szükség. Legyen az anyagtörvény

$$\Phi(\sigma_t, \sigma_x, \varepsilon_t, \varepsilon_x, \sigma, \varepsilon) = 0 \tag{9}$$

alakú.

A gyorsuláshullám esetében σ, v és ε folytonosak, deriváltjaik azonban a $\varphi(x, t)$ mentén ugranak. A (6) felhasználásával felírhatjuk a kinematikai kompatibilitási egyenleteket:

$$[v_x] = v\varphi_x, \quad [v_t] = v\varphi_t, \quad [\sigma_x] = \mu\varphi_x, \quad [\sigma_t] = \mu\varphi_t, \quad [\varepsilon_x] = \varkappa\varphi_x$$

$$\text{és} \quad [\varepsilon_t] = \varkappa\varphi_t, \quad \text{ahol} \quad \varphi_x \equiv \frac{\partial \varphi}{\partial x} \quad \text{és} \quad \varphi_t \equiv \frac{\partial \varphi}{\partial t}.$$

A (7), (8) és (9) képletek, valamint az elmondottak alapján felírhatjuk a gyorsuláshullámra vonatkozó dinamikai feltételeket, azaz a

$$\varrho[v_t] = [\sigma_x], \tag{10}$$

$$[\varepsilon_t] = [v_x] \tag{11}$$

és

$$\Phi(\sigma_t^- + [\sigma_t], \sigma_x^- + [\sigma_x], \varepsilon_t^- + [\varepsilon_t], \varepsilon_x^- + [\varepsilon_x], \sigma, \varepsilon,) - \Phi(\sigma_t^-, \sigma_x^-, \varepsilon_t^-, \varepsilon_x^-, \sigma, \varepsilon) = 0$$

egyenleteket. A kinematikai és dinamikai kompatibilitási egyenleteket összevetve, elhagyva a Φ-ben a felső jelölést; a φ-re nézve egy nemlineáris elsőrendű parciális differenciálegyenletet kapunk. Ennek az elsőrendű parciális differenciálegyenletnek a karakterisztikus egyenletrendszere:

$$\frac{dx}{\Phi_{\varphi_x}} = \frac{dt}{\Phi_{\varphi_t}} = \frac{d\varphi}{\varphi_x \Phi_{\varphi_x} + \varphi_t \Phi_{\varphi_t}}.$$

A fenti egyenlőségekből a továbbiakban csak az elsőt vegyük figyelembe,

$$\frac{dx}{dt}(\Phi_{\sigma_t}\mu + \Phi_{\varepsilon_t}\varkappa) = \Phi_{\sigma_x}\mu + \Phi_{\varepsilon_x}\varkappa. \tag{12}$$

A (10) és (11)-ből a megfelelő kinematikai kompatibilitási feltétel felhasználásával

$$\varrho\varkappa\left(\frac{dx}{dt}\right)^2 = \mu.$$

Ezzel (12) a

$$\varrho\Phi_{\sigma_t}\left(\frac{dx}{dt}\right)^3 - \varrho\Phi_{\sigma_x}\left(\frac{dx}{dt}\right)^2 + \Phi_{\varepsilon_t}\frac{dx}{dt} - \Phi_{\varepsilon_x} = 0$$

egyenletbe rendezhető, ahol a $\frac{dx}{dt}$ a hullám terjedési sebessége. A terjedési sebességre vonatkozó harmadfokú egyenlet gyökminőségének vizsgálatával ([9]) a Φ-re, a lehetséges anyagtörvényre bizonyos következtetések vonhatók le. Ugyanis a $\frac{dx}{dt}$ terjedési sebesség valós érték, a terjedési sebesség nem lehet nulla vagy végtelen nagy, szükséges, hogy legyen a gyökök között egyidejűleg pozitív és negatív szám is. Mindezek figyelembevételével megállapítható például, hogy a $\Phi(\varepsilon_t, \varepsilon, \sigma) = 0$ alakú egyenlet nem lehet anyagtörvény [10].

A gyorsuláshullám vizsgálatára bemutatott példa mutatja, miként lehet a mozgó szinguláris felületek figyelembevételével a régóta és jelenleg is a mechanikai érdeklődés középpontjában álló anyagtörvény meghatározásához közelebb jutni.

Természetesen konkrét anyag esetében az anyagtörvény felállításához, a fenti meggondolásokkal vezetett kísérletekre van szükség.

IRODALOM

[1] HADAMARD, J., *Lecons sur la Propagation des Ondes et les Equations de l'Hydrodynamique* (Hermann, Paris, 1903).
[2] *Annali di Matematica*, tome VIII, 1877.
[3] *Journal de l'Ecole Polytechnique*, tome XXXIX, 1887 és *Journal de Math.*, tome III, serie IV, 1887.
[4] HILL, R., *Progress in Solid Mechanics*, Vol. II., Chap. 3, (North-Holland, Amsterdam, 1961).
[5] ERICKSEN, J. L., "On the propagation of waves in isotropic incompressible perfectly elastic materials", *Journal of Rat. Mech. Anal.* 2 (1953).

[6] THOMAS, T. Y., "Extended compatibility conditions for the study of discontinuity in continuum mechanics", *Journal of Moll. Phys.* 4 (1957).
[7] ERINGEN, A. C., *Continuum Physics*, Vol. II (Academic Press, New York, San Francisco, London, 1975).
[8] ERINGEN, A. C. and SUHUBI, E. S., *Elastodynamics*, Vol. I (Academic Press, New York, London, 1974).
[9] BÉDA, GY., "A dinamikus képlékeny húzás lehetséges anyagtörvényeinek differenciálegyenletei", IUTAM, MN. Képlékenységtani Kollokvium, Miskolc, 1967.
[10] BÉDA, GY., "Methode zur Bestimmung der Material-gleichung", *NME Közleményei* XXXII (1962).

(Beérkezett: 1979. március 30.)

DR. BÉDA GYULA
BME GÉPÉSZMÉRNÖKI KAR MŰSZAKI MECHANIKA TANSZÉK
1111 BUDAPEST, MŰEGYETEM RKP. 1—3.

A MATHEMATICAL METHOD FOR INVESTIGATIONS IN PLASTICITY THEORY

GY. BÉDA

A lot of questions arise in mechanical researches which need functions with discontinuity on a few lines or surfaces. Cases with moving lines or surfaces have a special interest. We give a summary of the mathematical relations for these cases and present a simple application in plasticity theory.

Alkalmazott Matematikai Lapok **5** (1979) 69—87

NYELVSTATISZTIKAI TÁBLÁZATOK

NEMETZ TIBOR és SZILLÉRY ANDRÁS
Budapest

Az írott magyar nyelv entrópiájára és az egyértelmű rekonstruálhatóság megkövetelése mellett elérhető rövidítés elméleti korlátjának megállapítására vonatkozóan végeztünk statisztikai vizsgálatokat. A fontosabbnak látszó statisztikai adatokat táblázatokban közöljük, következtetéseinket a táblázatokhoz fűzött megjegyzések formájában ismertetjük. Eredményeink következményeként magyar nyelvű újságszövegek entrópiájára 1,65 adódott felső becslésként.

1. Bevezetés

Hírközlésben, adatfeldolgozásban, adattárolásban egyaránt fontos hatékony adattömörítésre törekedni. Ismeretes, hogy az egyértelmű visszaállíthatóság megkövetelése esetén az adatsorozatok rövidítésének elméleti alsó korlátja a sorozat entrópiája. Ezért szükséges a szóban forgó sorozatok különböző rendű entrópiájának vizsgálata. Ezek a vizsgálatok különösen érdekesek írott élő nyelvek esetén. Ekkor pusztán statisztikai úton az entrópiára csupán durva becsléseket lehet nyerni a betűk statisztikai összefüggősége és a nagyszámú lehetséges betűkombináció miatt. Különböző módszereket dolgoztak ki a technikai nehézségek kiküszöbölésére. A módszerek közös vonása, hogy az adott nyelvet jól ismerő személyeknek a szóban forgó valószínűségekre vonatkozó szubjektív megítéléseit hasznosítják. Ismertetésük helyett utalunk JAGLOM, JAGLOM [6] könyvére, ahol az olvasó a téma alapos feldolgozásán kívül különböző nyelvekre vonatkozó konkrét eredményeket is megtalálhat.

Az írott magyar nyelv entrópiájára vonatkozóan NEMETZ [1] és NEMETZ, SIMON [2] dolgozata tartalmaz kezdeti eredményeket. Az ott elkezdett vizsgálatok folytatására, illetve kiegészítésére végeztünk statisztikai vizsgálatokat. Jelen dolgozatban röviden ismertetjük következtetéseinket. Elsőrendű célunk azonban a fontosabb statisztikai adatok közlése. Táblázataink többnyire jellegükben térnek el a korábbiakban publikált nyelvstatisztikai táblázatoktól, mivel rövidítés céljából készültek. Így elsősorban műszaki szakemberek, matematikusok használhatják őket. Közlésüket az elektronikában végbement nagyarányú fejlődés indokolja: az egy évtizede még gyakorlatilag kivitelezhetetlennek vagy drágának ítélt egyes adattömörítési eljárásokat ma már valószínűleg egyszerű realizálni. Megjegyezzük, hogy a magyar nyelvre vonatkozóan korábban végzett statisztikai vizsgálatokról az olvasó PETŐFI S. JÁNOS [4] témafeldolgozásából tájékozódhat.

2. A vizsgált szövegek ismertetése

Vizsgálatainkhoz 51 500 karaktert tartalmazó újságszöveget rögzítettünk 8 csatornás lyukszalagra. Ebben különböző napilapokból származó politikai és rendőri hírek, regényismertetések, a gazdasági helyzetet elemző cikkek és egy tudományos szocializmussal foglalkozó tanulmány szerepelt. Igyekeztünk az egyes témákat a napilapokban elfoglalt terjedelmüknek megfelelő arányban szerepeltetni. Mivel azonban célunk elsősorban tájékoztató jellegű eredmények nyerése volt, a reprezentativitás kérdését mélyrehatóan nem vizsgáltuk.

Az adatrögzítés módja lehetővé tette, hogy írásjegyekre, illetve különböző speciális jegyekre is kiterjesszük vizsgálatainkat. Így lehetőség volt arra, hogy alkalmas átkódolás segítségével ötcsatornás lyukszalag karaktereire vonatkozó gyakorisági táblázatot nyerjünk. Esetünkben 62 100 karakterből állt az ötcsatornás lyukszalagszöveg. Entrópia vizsgálatoknál általában figyelmen kívül hagyják az írásjeleket, számokat is. Az ékezetes ábécére szorítkozva (kettősbetűket az őket alkotó betűpárokként felfogva), s a szóközt is betűnek számítva szövegünk 48 198 betűt tartalmazott. (A lyukasztók által nem javított hibákat mi sem korrigáltuk.)

A táblázatokban a következő rövidítéseket és jelöléseket használtuk: köz= =szóköz, kv.=kocsi vissza, se.=soremelés, bv.=betűváltó, jv.=jelváltó. A nulla jelölésére a 0 szimbólumot használtuk. A köz-köz „betűpár" új bekezdést jelöl.

3. Folyamatos szövegek vizsgálata

Eredményeinket és következtetéseinket a táblázatokhoz fűzött megjegyzések formájában ismertetjük.

1. Táblázat: Szomszédos betűpárok gyakoriságai. Folyamatos szövegből, csak az ékezetes ábécé betűinek és a szóköznek a megtartásával keletkezett betűsorozat alapján készült. Az összes betűk száma 48198. A Q és W betűket figyelmen kívül hagytuk, mivel abszolút gyakoriságuk elhanyagolhatóan kicsi volt. Az *i, o, ö, u, ü* betűket hosszú párjukkal megegyezőnek tekintettük. „Ölelkező" betűpárokat tekintettünk, tehát az utolsó betű kivételével valamennyi betű egy betűpár első tagja volt. A táblázat az egyes betűpárok relatív gyakoriságait százezredekben adja meg. Az utolsó oszlop a betűk relatív gyakoriságait tízezredekben mutatja.

A táblázat adatai alapján számított entrópiák:

$$H_0 = \log_2 31 = 4{,}95,$$
$$H_1 = H(\text{betű}) = 4{,}40,$$
$$H_2 = H(\text{betű/megelőző betű}) = 3{,}70.$$

Ezek az értékek azt mondják számunkra, hogy a lehető legjobb betűnkénti kódolás sem eredményezhet 12%-osnál nagyobb átlagos rövidülést, míg betűpárok alkalmas kódolása mellett maximálisan 26%-os rövidülésre számíthatunk. Megjegyezzük, hogy az utóbbi egy kb. 900 kódszóból álló kódszótár használatát igényli, ami napjainkban már olcsón és egyszerűen megvalósítható.

2. Táblázat: Ékezetes ábécé feltételes gyakorisági sorrendje. Ez az 1. táblázat alapján készült, s azt mutatja, hogy az első oszlopban szereplő betűk után folyamatos

1/A TÁBLÁZAT

Szomszédos betűpárok gyakoriságai százezredekben 1. lap

	A	Á	B	C	D	E	É	F	G	H	I	J	K	L	M
A	38	4	79	21	296	19	6	8	256	60	231	196	573	1115	177
Á	6	2	158	42	58	—	6	4	260	4	15	52	171	481	104
B	350	42	196	2	—	496	29	8	—	2	71	—	2	23	—
C	10	17	6	—	2	4	21	2	—	—	83	—	2	—	—
D	229	179	10	2	21	265	129	2	—	10	210	23	8	10	6
E	85	38	44	8	138	13	29	35	763	54	140	160	588	1423	548
É	8	2	98	2	42	4	6	4	317	13	13	15	185	231	63
F	40	10	—	—	—	350	44	—	—	13	131	8	—	2	—
G	217	121	17	8	8	244	165	21	35	27	90	40	25	40	19
H	338	94	2	—	—	150	19	—	2	—	40	2	8	15	—
I	315	117	50	31	102	71	19	44	206	31	10	4	406	217	104
J	250	150	10	—	58	225	19	2	—	—	10	—	6	81	—
K	433	121	56	6	10	431	219	10	—	23	394	10	104	29	50
L	538	304	15	19	106	904	144	21	54	44	392	79	154	365	246
M	454	283	171	15	2	746	258	8	4	8	373	2	8	29	67
N	340	77	67	69	273	492	154	4	125	27	238	2	146	15	13
O	6	25	56	60	131	2	17	48	256	10	25	8	338	648	331
Ö	4	10	73	—	100	17	2	4	27	38	17	23	102	219	17
P	52	81	2	23	—	92	35	—	13	2	92	46	17	38	—
R	396	173	56	46	104	510	123	19	23	42	290	50	60	27	138
S	408	277	54	6	4	356	306	12	6	13	129	—	27	15	60
T	729	527	58	25	2	790	340	13	2	94	400	110	63	73	33
U	6	2	4	13	113	10	—	—	75	13	—	48	54	233	15
Ü	—	2	—	—	—	8	—	6	23	—	—	—	58	200	2
V	340	202	—	—	—	306	150	—	—	8	181	—	—	2	—
X	—	—	—	—	—	—	10	—	—	—	50	—	—	—	—
Y	163	69	23	—	—	246	47	10	4	17	113	8	13	4	21
Z	227	173	23	2	71	683	127	—	21	17	213	8	31	23	167
köz	2477	204	265	165	165	1069	481	688	121	473	442	213	967	323	1113

1/B TÁBLÁZAT

Szomszédos betűpárok gyakoriságai, *2. lap.*
Az utolsó oszlop a betűgyakoriságokat tízezredekben mutatja. A köz-köz pár új bekezdést jelent

	N	O	Ö	P	R	S	T	U	Ü	V	X	Y	Z	KÖZ	Össz:
A	733	8	4	215	396	148	615	15	—	44	2	—	750	2450	846
Á	438	—	—	4	546	463	317	8	—	31	—	—	92	44	331
B	6	79	42	2	54	10	—	46	21	—	—	—	—	113	159
C	2	8	—	—	17	308	—	—	—	—	—	—	—	27	56
D	40	240	100	—	19	50	38	58	8	17	—	2	4	125	181
E	923	629	4	50	804	504	1102	44	2	33	4	2	327	629	851
É	358	—	—	119	315	694	231	—	—	63	—	—	73	52	291
F	—	219	81	—	17	2	—	23	21	—	—	—	—	4	96
G	42	160	33	4	23	40	46	17	21	27	—	746	27	333	259
H	8	252	33	—	13	8	17	17	10	—	—	—	2	63	109
I	342	29	4	23	123	571	448	10	10	50	—	—	223	829	439
J	6	106	15	—	10	10	54	42	13	—	—	—	17	23	111
K	48	460	231	15	60	38	73	106	85	13	—	—	2	1088	412
L	83	360	240	13	25	135	456	42	17	54	8	313	13	738	588
M	19	131	17	15	10	15	33	279	31	4	—	6	17	288	329
N	108	75	60	10	6	67	390	46	15	19	—	542	23	1050	445
O	415	4	—	60	438	390	329	6	—	56	—	—	304	238	420
Ö	90	6	4	6	225	165	117	4	2	85	—	—	160	240	176
P	4	163	6	15	83	4	21	15	4	23	—	—	6	25	86
R	58	302	150	2	94	146	342	42	71	44	108	—	60	356	383
S	33	265	83	6	19	221	327	44	27	8	—	—	1392	1300	539
T	52	471	335	10	79	100	617	175	42	60	—	29	2	1690	692
U	158	2	—	6	81	263	185	—	—	2	—	—	6	38	133
Ü	54	—	—	—	17	23	25	—	—	13	—	—	46	46	52
V	2	127	21	—	2	—	4	4	6	2	—	—	—	2	136
X	4	—	—	2	—	—	2	—	—	—	—	—	—	54	12
Y	8	148	23	—	15	58	27	21	25	21	—	—	23	535	164
Z	40	402	108	6	15	108	315	50	54	25	—	—	60	710	368
köz	375	171	160	277	327	850	794	215	38	665	—	2	50	267	1335

szövegben közvetlenül következő helyen mi a betűk gyakorisági sorrendje. Az adott betűk után a mintában elő nem fordult betűket feltétel nélküli gyakoriságuk szerint rendeztük. A + jel a szóközt helyettesíti.

2. TÁBLÁZAT

Ékezetes ábécé feltételes gyakorisági sorrendje

A	+LZNT	KRDGI	PJMSB	HVACE	UFOÉÖ	ÁXYUW	Q
Á	RLSNT	GKBMZ	DJ+CV	IUAÉH	FPÁEO	ÖYÜXW	Q
B	EAB+O	IRUOÁ	ÉLÜSF	NKHPC	TZMGD	YVJXW	Q
C	SIH+É	RÁAOB	ENKDF	TLZMG	OYVUJ	PCÜXW	Q
D	EOAFÁ	É+ÖUS	NTJDR	VLBHK	ÜMZYF	CGPXW	Q
E	LTNRG	+KMSZ	JIDAH	PBUÁF	VÉEOC	ÖXÜYW	Q
É	SNRGL	TKPBZ	MV+DJ	IHÁÉE	FÁCOÖ	YUÜXW	Q
F	EOIÖÉ	AUÜRH	ÁJ+LS	TNKZM	GDYBV	EPCXW	Q
G	Y+EAÉ	OÁITN	LSJGÖ	ZVHKR	FÜMBV	DCPXW	Q
H	AOEÁ+	IÖTÉU	LÜRKN	SHGZB	JMPCV	DYFXW	Q
I	+STKN	AZLGR	ÁMDEV	BFHCO	PÉIUÜ	JÖYXW	Q
J	AEOÁL	TDU+É	ZÖÜSI	RBNKF	MGYVJ	HPCXW	Q
K	+OAEI	ÖÉUÁK	ÜTRBM	NSLHP	VDJFC	ZGYXW	Q
L	E+ATI	LOYÁM	ÖKÉSD	NJGVH	URFCÜ	BPZXW	Q
M	EAI+Á	UÉBOM	TÜLNZ	ÖSPCR	KFHYG	VJDXW	Q
N	+AÁET	ÉÖDIU	GJNSO	CKBÜY	HZVLM	FPRWX	Q
O	LRNSK	MTZG+	DPBVC	FIÁÉH	JAUOE	YÖÜWX	Q
Ö	+RLSZ	TKDNV	BHGJE	IMAOP	AÖUFÉ	ÜYCXW	Q
P	OIERÁ	AJLÉ+	VCTKU	PGZÖS	NÜBHM	DYFXW	Q
Q	+EATL	SNIOK	RZÁMÉ	GDÖYB	VUJHF	PCÜXW	Q
R	EA+TO	IÁSÖM	ÉXDRÜ	ZKNBJ	CVHUL	GFPYW	Q
S	Z+AET	ÉÁOSI	ÖMBUN	KÜRLH	VGPCD	FYJXW	Q
T	+EATÁ	OIÖÉU	JSHRL	KVBNÜ	MYCFP	GDZWX	Q
U	SLTND	RGKJ+	MHCEA	ZPBOÁ	VIÉÖY	UFÜXW	Q
Ü	LKN+Z	TGSRV	EFÁMA	IOÉDÖ	YBUJH	PCÜXW	Q
V	AEÁIÉ	OÖHÜT	U+LNR	VSKZM	GDYBJ	FPCXW	Q
W	E+ATL	SNIOK	RZÁMÉ	GDÖYB	VUJHF	PCÜXW	Q
X	+IÉTP	EALSN	OKRZÁ	MGDÖY	BVUJH	FCÜXW	Q
Y	+EOAI	ÁSÉTÜ	ZÖBMV	UHRKF	NJLGD	YPCXW	Q
Z	+EOTA	IÁMÉS	ÖDZÜU	NKVLB	GHRJP	CYFXW	Q
köz	AMEKS	TFVÉH	INRLP	BJUÁO	CDÖGZ	ÜYWQX	+

A táblázat SHANNON entrópia-becslési módszerének *(„guessing game")* kisszámú találgatásra korlátozott változatában került felhasználásra.

3. *Táblázat:* Adott távolságra levő betűpárok feltételes entrópiái a távolság és a feltétel függvényében. A számításokat az 1. táblázat kapcsán leírt anyagon végeztük el. Az utolsó oszlop az egyes betűk előfordulásának relatív gyakoriságát mutatja, míg az utolsó sor az átlagos feltételes entrópiák értékét adja meg.

Célunk annak megvizsgálása volt, hogyan csökken a statisztikai függőség az adott távolságra elhelyezkedő betűpárok között a távolság függvényében. Ismeretes, hogy két valószínűségi változó függőségét jól jellemzi kölcsönös információjuk,

3. TÁBLÁZAT

Adott távolságra levő betűpárok feltételes entrópiái

DIST.$=k=$	1	2	3	4	5	6	7	8	9	10	30	200	Prob.
T	3,681	4,173	4,226	4,319	4,395	4,392	4,391	4,380	4,408	4,417	4,396	4,386	0,06922
O	3,802	4,066	4,049	4,284	4,170	4,374	4,364	4,413	4,388	4,391	4,384	4,421	0,04204
köz	4,214	4,130	4,430	4,495	4,422	4,435	4,367	4,369	4,384	4,369	4,408	4,383	0,13358
H	3,009	3,512	4,062	4,024	4,350	4,194	4,333	4,269	4,310	4,378	4,307	4,317	0,01090
N	3,752	4,259	4,240	4,352	4,345	4,409	4,383	4,433	4,427	4,398	4,400	4,399	0,04448
M	3,516	3,994	4,308	4,176	4,342	4,327	4,366	4,379	4,408	4,427	4,382	4,426	0,03291
L	4,057	4,337	4,354	4,357	4,318	4,351	4,332	4,420	4,365	4,395	4,399	4,351	0,05880
R	4,228	4,238	4,357	4,385	4,333	4,310	4,358	4,310	4,365	4,392	4,441	4,408	0,03833
G	3,688	4,086	4,412	4,289	4,383	4,405	4,362	4,383	4,380	4,409	4,406	4,339	0,02595
I	3,841	4,267	4,263	4,245	4,397	4,408	4,393	4,342	4,329	4,372	4,353	4,397	0,04391
P	3,859	3,866	4,110	4,101	4,116	4,350	4,167	4,272	4,268	4,229	4,355	4,303	0,00861
C	2,326	3,601	3,846	4,112	4,132	4,297	4,418	4,070	4,213	4,327	4,370	4,385	0,00561
V	2,746	3,333	4,020	4,179	4,362	4,338	4,352	4,350	4,354	4,365	4,258	4,337	0,01361
E	3,734	4,076	4,240	4,353	4,362	4,355	4,411	4,397	4,380	4,378	4,377	4,396	0,08508
Z	3,725	4,268	4,315	4,330	4,269	4,266	4,314	4,391	4,395	4,444	4,393	4,414	0,03679
D	3,740	3,975	4,099	4,198	4,286	4,333	4,402	4,345	4,359	4,326	4,314	4,384	0,01807
B	3,093	3,525	3,642	4,421	4,445	4,480	4,446	4,349	4,351	4,330	4,380	4,397	0,01590
S	3,316	4,313	4,314	4,336	4,399	4,380	4,390	4,414	4,354	4,394	4,387	4,402	0,05390
Y	3,394	4,247	4,165	4,379	4,401	4,321	4,445	4,391	4,372	4,334	4,435	4,315	0,01642
F	2,732	3,245	4,285	4,193	4,164	4,161	4,191	4,378	4,195	4,294	4,366	4,362	0,00965
X	1,716	3,074	3,017	2,752	2,871	3,071	3,637	3,315	3,517	3,716	3,944	3,700	0,00123
A	3,492	4,209	4,302	4,426	4,447	4,392	4,401	4,388	4,441	4,385	4,379	4,415	0,08462
W	0,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	0,00004
J	3,373	3,756	4,085	4,156	4,188	4,296	4,383	4,384	4,304	4,353	4,342	4,263	0,01109
U	3,388	3,746	4,075	4,176	4,255	4,398	4,320	4,320	4,442	4,334	4,350	4,370	0,01323
Q	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,00000
K	3,584	4,290	4,243	4,379	4,373	4,349	4,427	4,391	4,391	4,381	4,397	4,382	0,04114
Ö	3,829	4,130	4,311	4,173	4,307	4,412	4,365	4,346	4,332	4,377	4,349	4,342	0,01757
Á	3,517	3,853	4,091	4,163	4,338	4,366	4,345	4,413	4,383	4,393	4,413	4,401	0,03306
Ü	2,965	4,033	3,970	4,235	4,011	4,265	4,189	4,172	4,420	4,235	4,357	4,272	0,00523
É	3,504	3,619	4,082	4,149	4,259	4,347	4,406	4,382	4,372	4,391	4,361	4,381	0,02905
$H(\xi_{i+k}\|\xi_i)$	3,702	4,099	4,251	4,326	4,351	4,369	4,373	4,377	4,379	4,382	4,385	4,385	4,398

4/A TÁBLÁZAT

Ötcsatornás lyukszalagszöveg karakterpárjainak gyakoriságai, 1. lap

Jel	Betű	Sorszám	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
5	T	1	475	74	365	1004	60	35	21	—	53	58	2	307	5	19	47	602
kv.		2	—	—	—	—	—	—	—	1440	—	—	—	—	—	—	—	—
9	O	3	253	16	3	174	8	320	254	—	500	338	196	19	45	47	43	2
köz		4	555	225	111	—	330	251	713	—	259	211	92	307	190	116	458	726
Ö	H	5	11	19	195	145	6	14	2	2	18	10	3	31	2	—	—	116
,	N	6	293	132	58	1364	14	84	5	—	11	3	93	182	8	51	11	376
.	M	7	23	236	101	674	6	23	63	—	23	8	3	290	11	8	2	578
se.		8	50	3	11	39	29	27	66	—	37	23	5	6	10	5	31	42
Á	L	9	354	19	277	455	32	68	182	—	270	11	42	299	8	14	39	705
4	R	10	261	8	238	249	32	45	98	—	21	71	18	232	5	35	32	394
	G	11	29	34	125	219	18	27	8	—	31	14	29	71	3	3	18	188
8	I	12	347	16	23	594	16	262	85	—	167	105	158	8	19	23	39	42
Ø	P	13	16	—	125	16	3	3	2	—	29	64	10	71	11	14	18	71
:	C	14	—	2	6	16	39	2	—	—	—	13	—	64	—	—	—	3
=	V	15	2	2	98	2	6	2	—	—	2	2	—	140	—	—	2	236
3	E	16	853	29	8	407	40	711	417	—	1100	619	592	108	39	5	21	10
+	Z	17	243	11	310	520	13	27	127	—	16	11	14	164	5	—	16	526
!	D	18	29	5	187	87	6	27	3	—	8	14	—	162	—	2	11	203
?	B	19	—	5	61	72	2	5	—	—	19	42	—	55	—	2	—	384
Ü	S	20	246	47	203	951	10	24	39	—	10	14	6	98	5	3	—	269
6	Y	21	19	16	114	396	13	6	16	—	2	10	3	87	2	—	16	185
*	F	22	—	2	169	3	10	—	—	—	2	13	—	101	—	—	—	272
/	X	23	2	—	—	42	—	3	—	—	—	—	—	39	2	—	—	—
—	A	24	462	423	6	1828	40	565	122	2	861	312	199	180	162	16	29	5
2	W	25	—	—	—	—	—	—	—	—	—	—	—	—	2	—	—	6
	J	26	42	—	82	13	—	5	—	—	63	8	—	8	—	—	—	175
jv.		27	—	3	2	16	1311	944	689	—	2500	3	13	2	—	2	—	—
7	U	28	143	5	2	27	10	117	11	—	177	64	60	—	5	10	2	8
1	Q	29	2	—	2	2	—	—	2	—	—	2	—	18	—	—	—	—
É	K	30	50	68	352	642	14	37	35	—	16	40	5	304	10	5	6	330
bv.		31	629	43	27	384	138	788	328	—	943	911	504	74	130	55	216	159

4/B TÁBLÁZAT

Ötcsatornás lyukszalagszöveg karakterpárjainak gyakoriságai, 2. lap

Jel	Betű	Sorszám	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	Össz.
5	T	1	2	—	39	71	23	6	—	510	2	82	1343	138	—	45	—	539
kv.		2	—	—	—	—	—	—	—	—	—	—	2	—	—	—	3	144
9	O	3	227	101	42	296	—	34	—	3	—	8	64	5	—	262	2	326
köz		4	27	88	183	531	2	507	—	1514	3	140	679	156	—	735	1231	1034
Ö	H	5	2	—	2	6	—	2	—	278	—	2	122	13	—	6	1202	221
,	N	6	14	211	48	51	421	3	—	245	—	2	378	35	—	108	220	442
.	M	7	13	2	130	8	5	5	—	351	—	2	481	217	—	6	19	329
se.		8	14	3	24	47	—	29	—	119	—	10	56	3	—	40	714	144
Á	L	9	10	82	11	105	241	11	6	433	—	60	740	32	—	122	2506	714
4	R	10	47	79	42	109	—	13	85	304	—	34	462	34	—	45	2	299
	G	11	21	5	10	27	578	18	—	161	—	29	341	13	—	21	6	205
8	I	12	172	79	34	441	3	35	—	220	—	3	220	8	—	307	—	343
Ø	P	13	5	—	2	3	—	—	—	48	—	35	105	13	—	13	—	68
:	C	14	—	—	3	238	—	2	—	8	—	—	37	—	—	2	—	43
=	V	15	—	—	—	—	—	—	—	264	—	—	296	3	—	—	—	105
3	E	16	251	103	35	389	2	26	3	51	—	122	185	34	—	452	—	661
+	Z	17	47	53	18	80	—	—	—	172	—	6	415	39	—	24	—	286
!	D	18	3	16	8	35	2	2	—	180	—	18	341	45	—	6	—	140
?	B	19	—	—	153	8	—	6	—	270	—	—	125	35	—	—	3	125
Ü	S	20	1086	—	37	162	—	—	—	298	—	—	669	32	—	21	367	460
6	Y	21	18	—	18	35	—	5	—	122	—	6	162	19	—	6	—	128
*	F	22	—	—	—	2	—	—	—	29	—	8	121	18	—	—	10	76
/	X	23	—	—	—	—	—	—	—	—	—	—	10	—	—	—	2	10
—	A	24	582	230	60	113	2	3	2	3	2	153	216	11	2	438	82	711
2	W	25	—	—	—	—	—	—	—	2	—	—	—	2	2	—	—	1
	J	26	13	42	8	6	—	2	—	195	—	—	164	34	—	5	—	86
jv.		27	—	—	—	399	—	—	2	425	5	—	—	2	24	2198	—	854
7	U	28	5	88	3	206	—	—	—	6	—	37	16	—	—	42	—	104
1	Q	29	—	—	—	—	—	—	—	—	2	—	—	—	2	—	—	3
É	K	30	2	6	48	23	—	—	—	343	—	6	787	79	—	77	2169	545
bv.		31	299	212	291	1203	—	50	—	555	—	101	—	24	—	471	—	854

lásd pl. RÉNYI [5], 3.8 §. Véges értékkészletű ξ és η változók esetén ez ekvivalens a

$$H(\xi)-H(\xi|\eta)$$

entrópia-különbséggel, melynek becsült értékei a táblázatból kiolvashatók. Megjegyezzük, hogy a $H(\xi_{i+10k}|\xi_i)$ átlagos feltételes entrópiák értéke 4,382 és 4,388 között változott, midőn k-t 1 és 20 között változtattuk. Első pillanatra úgy tűnhet tehát, hogy a statisztikai függőség nagyon hosszú távú. Téves lenne azonban ezt a következtetést levonni, mivel a felhasznált becslés torzított, s csak aszimptotikusan lesz torzítatlan.

A táblázatban található adatok kérdésessé teszik NEWMAN és GERSTMAN [3] megállapításainak megbízhatóságát. Ők mindössze 1000 betűből álló angol biblia szöveg alapján következtettek arra, hogy a $H(\xi_i)-H(\xi_{i+k}|\xi_i)$ különbség k növekedésével exponenciálisan csökken, s a „pontos" exponenst is megadták. Esetünkben viszont az látszik, hogy még 48 ezer betűből álló mintából sem nyerhető a $k \geqq 10$ esetben elegendő pontosságú becslés. Természetesen az első néhány érték alapján lehetne illeszteni egy exponenciális típusú közelítő-függvényt, de a közelítés jóságáról nagy k esetén semmit sem tudnánk mondani. Eredményeink tájékoztató jellegűek, a kérdés elméleti vizsgálata még nem történt meg.

4. Táblázat: Ötcsatornás lyukszalag karakterpárjainak gyakoriságai. A karakterpárok relatív gyakoriságai százezredekben vannak megadva. Az utolsó oszlop az egyes karakterek (feltétel nélküli) relatív gyakoriságát tízezredekben mutatja. A táblázat alapján nyert entrópia-értékek:

$$H(0)=\log_2 31 \quad =4{,}95,$$
$$H(1)=H\text{ (karakter)} \quad =4{,}46,$$
$$H(2)=H\text{ (karakter/megelőző karakter)} \quad =3{,}89.$$

Az utóbbi érték azt mutatja, hogy karakterpárokon alapuló egyértelmű dekódolható kódolás maximálisan kb. 20%-os rövidülést eredményezhet.

5. Táblázat: Nyolccsatornás lyukszalag karaktereinek relatív gyakoriságai tízezredekben. A táblázatban csak a szövegben előfordult jelek vannak feltüntetve. A „rang" elnevezésű oszlopban azt tüntettük fel, hogy az illető írásjegy a gyakorisági sorrendben hányadik helyen áll. Karakter-pár gyakoriságot is vizsgáltunk, ennek alapján számítottuk a H(karakter/megelőző karakter=adott) feltételes entrópiákat. Az „entrópia" elnevezésű oszlop ezeket tartalmazza a feltétel függvényében.

A kv., se. pár gyakorisága azt mutatja, hogy soronként átlagosan 57,5 karakter szerepelt. Ugyancsak kiolvasható, hogy a kiválasztott szövegben az írásjegyek kb. 97%-a betű volt, a szóközt is beleértve.

6. Táblázat: Egyes írásjegy-párok relatív gyakoriságát tartalmazza százezredekben kifejezve. A kv-se párok relatív gyakorisága 0,1738 volt. Az „egyéb" oszlopban a fel nem sorolt írásjegyek szerepelnek. Az 1. táblázattal együtt használva táblázatunk alapján az összes számottevő gyakoriságú írásjegy-pár relatív gyakoriságai meghatározhatók. Ezek alapján a 8 csatornás lyukszalagra lyukasztott újságszövegek entrópiájára a következő becslések adódnak:

$$H(0)=\log_2 256 \quad =8,$$
$$H(1)=H\text{ (karakter)} \quad =4{,}62,$$
$$H(2)=H\text{ (karakter/megelőző karakter)} \quad =3{,}72.$$

5. TÁBLÁZAT

Nyolccsatornás lyukszalag karaktereinek gyakoriságai

Betűk	Rang	Gyakoriság	Entrópia
A	3	793	3,66
Á	13	310	3,60
B	22	150	3,19
C	32	52	2,37
D	19	169	3,81
E	2	797	3,83
É	15	273	3,61
F	28	90	2,74
G	16	244	3,85
H	27	102	3,07
I	8	410	3,99
J	26	104	3,42
K	10	385	3,91
L	5	551	4,23
M	14	308	3,57
N	7	416	3,97
O	9	393	3,87
Ö	20	164	3,95
P	30	81	3,90
Q	50	0	0,00
R	11	358	4,33
S	6	505	3,49
T	4	650	3,98
U	24	125	3,48
Ü	33	49	3,06
V	23	127	2,75
W	48	0	0,00
X	34	11	1,82
Y	21	154	3,57
Z	12	349	3,84
köz	1	1240	4,28
összesen:		9360	

Jelek	Rang	Gyakoriság	Entrópia
kv	18	174	0,02
se	17	174	4,30
,	25	118	0,58
.	29	88	1,94
!	35	3	2,65
?	46	1	0,92
—	31	66	1,80
/	47	1	1,00
:	49	0	0,00
=	51	0	0,00
+	52	0	0,00
összesen:		625	—

(kv = kocsi vissza)
(se = soremelés)

Számok			
0	41	1	1,80
1	36	3	2,06
2	42	1	2,25
3	43	1	1,37
4	38	3	3,11
5	40	1	1,00
6	45	1	1,58
7	44	1	1,79
8	37	3	2,22
9	39	1	2,32
összesen:		16	—

Ebben az esetben $H(0)$ és $H(2)$ összehasonlításával nem adódhat reális kép az elérhető rövidítés mértékéről, hiszen a 8 bit könnyen ellenőrizhető hibajelzést is lehetővé tesz. Összehasonlítható viszont $H(2)$ értéke a valóban felhasznált karakterek és a paritásbittől megfosztott karakterek számának logaritmusával. A kapott

$$\frac{H(2)}{\log_2 47} \simeq 0{,}67 \quad \frac{H(2)}{7} \simeq 0{,}53$$

arányok világosan mutatják, milyen pazarlóan bánunk általában a memória-kapacitással.

6. TÁBLÁZAT

Egyes írásjegypárok gyakoriságai

Előtte							Írásjegy	Utána						
köz	.	,	—	se	szám	egyéb		köz	.	,	—	kv	szám	egyéb
2287	14	—	6	338	—	—	A	2108	111	117	31	78	—	2
188	—	—	2	29	4	—	Á	33	6	10	43	8	—	2
237	—	—	2	62	—	—	B	87	10	12	8	4	—	—
154	—	—	—	19	—	—	C	17	—	10	—	2	—	—
149	—	4	—	33	—	—	D	105	6	10	8	6	—	—
978	—	—	23	89	2	6	E	485	56	78	41	35	—	—
448	2	—	6	39	—	2	É	41	8	2	50	16	2	12
627	—	—	2	54	—	2	F	4	—	—	—	2	—	—
113	—	—	—	41	—	2	G	256	33	33	23	31	2	2
429	—	8	2	56	—	—	H	56	2	—	—	4	—	—
404	—	—	—	19	—	4	I	714	41	58	33	21	—	2
190	—	—	6	23	—	2	J	16	—	6	8	—	—	—
889	—	—	—	80	—	—	K	753	115	200	16	66	—	6
289	—	—	8	85	—	2	L	542	103	89	43	19	—	—
1016	—	—	8	149	—	—	M	250	6	17	6	16	—	—
346	—	—	4	78	—	4	N	850	58	93	21	49	—	2
151	—	—	2	19	—	4	O	210	6	9	29	19	—	—
146	—	—	—	17	2	—	Ö	210	2	10	21	19	2	4
250	—	—	—	25	—	—	P	19	2	2	4	—	—	—
289	—	—	2	62	—	4	R	301	12	25	45	10	—	—
775	2	—	8	120	—	—	S	1109	56	78	27	56	—	2
736	—	—	2	115	—	—	T	1215	194	245	23	89	—	4
198	—	—	—	14	—	—	U	27	2	9	12	4	—	—
33	—	—	—	2	2	—	Ü	39	2	2	2	—	—	4
610	—	—	2	58	—	—	V	2	—	—	—	2	—	—
—	—	—	—	—	—	—	X	49	—	2	—	—	—	—
2	—	—	—	—	—	—	Y	478	10	16	19	19	—	2
35	—	—	—	41	—	—	Z	627	8	41	23	14	—	—
—	577	1044	99	52	16	12	köz							
—	14	—	2	2	20	—	.							
—	—	—	—	—	2	—	,							
93	2	—	—	6	18	4	—							
282	274	118	458	6	2	14	kv.							
—	—	—	2	—	—	—	se.							
46	—	2	12	6	88	2	szám							
14	—	—	—	—	6	4	egyéb							

7. TÁBLÁZAT

Mássalhangzókra redukált szöveg szomszédos párjainak gyakoriságai

1.2.→ ↓ betű	B	C	D	F	G	H	J	K	L	M	N	P	R	S	T	V	X	Y	Z
B	415	45	62	74	37	16	16	111	399	99	970	16	493	127	82	25	0	0	103
C	25	4	4	41	4	103	8	29	111	29	29	0	49	637	16	4	0	0	4
D	62	86	66	21	189	53	62	280	822	321	345	12	201	555	255	90	0	4	111
F	8	4	37	4	218	49	288	74	563	8	115	0	354	53	66	25	0	0	16
G	95	29	115	74	148	111	107	197	580	189	378	45	214	473	456	140	0	1460	271
H	12	0	12	16	391	4	62	37	350	62	148	0	210	181	428	33	0	0	206
J	127	4	132	58	62	41	21	226	506	58	119	4	74	144	378	58	0	0	148
K	214	132	271	247	132	222	86	613	888	580	506	337	847	892	905	304	0	0	839
L	169	99	543	160	580	206	321	983	1188	1155	1012	354	226	1102	2081	267	16	609	428
M	432	37	136	78	794	107	99	259	641	259	1061	45	740	781	580	136	0	8	210
N	214	181	617	164	530	206	70	1320	555	596	412	136	169	674	1155	247	4	1065	345
P	8	45	66	4	41	21	103	74	247	16	140	37	391	136	255	58	0	0	45
R	247	127	436	115	308	197	247	572	757	617	551	132	325	999	1032	238	214	0	354
S	321	58	273	234	843	173	104	839	666	559	510	115	456	917	1229	251	0	4	2944
T	382	111	391	263	164	432	378	1431	1394	843	950	267	1517	1476	2475	415	8	58	518
V	74	21	99	25	267	41	16	49	839	41	325	0	201	259	230	49	4	0	136
X	0	0	0	0	0	4	4	0	0	21	16	33	8	70	8	0	0	0	86
Y	136	21	37	95	152	70	58	326	312	234	378	53	243	411	354	119	4	0	206
Z	152	95	275	210	222	99	111	584	683	720	695	99	748	596	1488	214	0	0	201
	309	110	354	188	508	215	216	801	1150	641	866	169	747	1049	1347	268	25	321	717

8. TÁBLÁZAT

Mássalhangzókra redukált szöveg xy típusú párjainak gyakoriságai

1.3. → / ↓ betű	B	C	D	F	G	H	J	K	L	M	N	P	R	S	T	V	X	Y	Z
B	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
C	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
D	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
F	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
G	13	3	21	14	25	10	10	35	64	29	48	11	37	61	55	11	1	24	39
H	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
J	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
K	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
L	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
M	11	10	37	13	21	11	20	47	57	27	50	3	66	52	84	16	22	55	38
N	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
P	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
R	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
S	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
T	45	10	25	21	75	29	32	117	155	106	128	28	102	150	180	37	—	28	79
V	8	2	6	3	19	4	3	23	22	12	15	0	11	34	49	8	0	24	24
X	—	—	—	—	3	—	1	0	2	7	6	—	2	—	3	—	—	—	1
Y	9	5	8	4	17	9	6	23	35	23	26	7	29	30	39	11	—	10	29
Z	46	9	24	10	27	12	13	57	94	27	59	16	38	86	111	18	—	23	49

9. TÁBLÁZAT

Mássalhangzókra redukált szöveg betűinek feltételes entrópiái a megelőző két betű függvényében

ξ_{i-1} → / ξ_i ↓	B	C	D	F	G	H	J	K	L	M	N	P	R	S	T	V	X	Y	Z
B	3,52	0,99	2,01	2,44	1,44	1,50	1,00	2,93	3,69	2,13	3,79	0,81	3,63	2,21	2,48	2,25	0,00	0,00	2,26
C	2,52	0,00	0,00	0,00	0,00	3,11	1,00	2,81	2,41	2,24	2,52	0,00	2,35	3,13	1,50	0,00	0,00	0,00	0,00
D	2,37	1,30	2,15	1,92	3,51	2,60	3,24	3,69	2,97	3,07	3,41	1,58	3,61	3,57	3,07	1,83	0,00	0,00	3,18
F	1,00	0,00	0,50	0,00	2,63	1,55	1,92	2,52	3,70	1,00	2,82	0,00	3,02	1,82	1,50	1,79	0,00	0,00	1,00
G	2,02	0,00	3,25	2,59	2,97	2,93	2,53	3,68	3,76	2,82	2,86	2,12	3,69	3,24	3,40	2,08	0,00	3,76	3,12
H	1,58	0,00	1,58	2,00	0,76	0,00	2,71	1,84	3,34	2,33	2,91	0,00	3,00	2,70	3,57	2,25	0,00	0,00	3,68
J	2,36	0,00	2,04	1,75	1,77	1,16	0,00	3,72	2,99	2,06	3,20	0,00	3,72	2,89	3,29	2,41	0,00	0,00	2,59
K	2,43	0,73	3,15	2,82	1,90	2,96	2,19	3,31	3,52	2,96	3,22	3,33	3,72	3,38	3,64	2,90	0,00	0,00	3,36
L	2,76	1,04	3,08	2,70	3,61	2,54	3,27	3,46	3,44	3,82	3,71	3,27	3,47	3,33	3,63	3,26	0,00	3,54	3,08
M	1,81	0,50	2,96	1,46	3,93	2,88	2,70	3,12	2,90	2,36	3,04	1,62	3,41	3,35	3,25	2,48	0,00	0,92	3,00
N	2,23	1,98	3,15	2,61	2,16	3,12	2,82	3,85	3,61	3,53	3,01	2,62	3,09	3,19	3,71	3,12	0,00	3,65	3,18
P	1,00	0,00	0,67	0,00	1,36	1,92	3,07	2,44	3,19	2,00	2,60	2,28	3,19	2,69	2,77	1,49	0,00	0,00	2,41
R	3,18	2,88	3,03	2,61	3,00	2,97	3,36	3,47	3,42	3,50	2,90	3,05	3,93	3,24	3,52	2,77	2,33	0,00	3,02
S	2,76	1,29	1,54	2,61	3,76	2,87	2,16	3,70	3,37	2,28	3,40	2,52	3,68	2,55	3,65	1,86	0,00	0,00	3,49
T	2,95	2,86	3,20	2,50	2,20	3,03	3,31	3,73	3,64	3,21	3,60	3,00	3,42	3,60	3,68	3,15	1,00	1,41	3,41
V	1,30	0,00	2,55	2,58	2,43	2,25	2,00	2,63	3,30	2,12	3,33	0,00	2,54	2,50	2,90	1,96	0,00	0,00	1,83
X	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,97	2,00	2,00	1,00	1,28	1,00	0,00	0,00	0,00	0,82
Y	2,33	1,52	2,73	1,83	2,36	2,70	2,50	3,70	3,51	3,22	3,93	2,05	3,71	2,98	3,28	2,47	0,00	0,00	2,84
Z	2,25	1,35	3,30	3,27	3,03	2,45	2,66	3,61	3,44	2,90	3,58	2,53	3,60	3,25	3,48	3,29	0,00	0,00	3,27

10. TÁBLÁZAT

Mássalhangzókra redukált szöveg betűinek feltételes entropiái a közrefogó két betű függvényében

ξ_{i-1} \ ξ_{i+1}	B	C	D	F	G	H	J	K	L	M	N	P	R	S	T	V	X	Y	Z
B	1,55	1,66	2,09	1,58	2,26	2,31	2,75	2,41	3,16	2,13	2,65	2,16	3,30	2,71	3,17	2,12	—	1,35	2,29
C	1,92	0,92	1,00	2,34	2,32	0,59	0,56	1,36	0,90	2,45	2,08	0,39	2,16	2,02	1,21	0,97	—	—	1,49
D	3,20	2,24	2,16	2,45	3,02	3,05	3,04	2,90	3,55	1,96	3,11	2,37	3,33	3,29	3,02	2,32	—	1,38	2,54
F	1,65	0,00	1,53	1,84	1,73	1,30	0,98	1,68	2,65	1,27	2,09	0,81	1,84	2,83	2,82	1,75	—	1,25	1,84
G	2,77	2,75	2,54	2,78	3,24	2,76	2,74	2,99	3,66	2,89	3,19	1,98	3,17	3,35	3,02	1,77	1,00	1,22	2,65
H	1,82	0,39	2,02	1,52	2,64	2,35	0,98	2,47	2,73	2,56	2,62	2,25	2,70	2,96	2,71	2,02	—	0,99	1,81
J	1,77	1,92	1,22	2,37	2,95	2,56	1,44	3,12	3,12	2,59	2,44	0,97	2,83	2,64	2,75	1,68	—	0,00	2,63
K	3,16	2,53	2,83	2,69	3,08	3,06	2,63	3,15	3,71	3,13	3,46	2,68	3,35	3,48	3,43	3,39	—	1,42	2,81
L	2,80	3,26	3,14	3,10	3,74	3,39	3,41	3,50	3,80	3,36	3,53	3,14	3,34	3,68	3,47	3,35	—	1,49	2,96
M	2,91	2,26	2,49	2,59	3,36	2,45	2,89	2,70	3,65	2,95	3,05	2,25	2,66	3,30	3,04	2,83	0,14	1,40	2,73
N	3,04	2,80	3,20	2,81	3,34	3,06	3,37	3,09	3,63	3,24	3,48	2,06	3,48	3,35	3,53	2,98	0,00	1,00	3,14
P	1,73	2,52	1,46	1,91	2,20	1,00	1,16	2,87	2,48	2,06	3,48	1,95	2,55	2,87	2,77	1,00	—	1,00	1,88
R	2,65	2,52	2,76	2,40	3,23	2,89	3,29	3,49	3,67	3,37	3,53	2,57	3,69	3,58	3,62	3,42	0,00	1,23	2,83
S	3,06	2,08	3,00	2,68	3,21	2,93	2,98	3,02	3,63	2,88	2,92	2,58	2,87	3,55	2,93	2,74	—	1,32	2,06
T	3,44	2,28	2,68	3,11	3,43	3,27	3,07	3,23	3,82	3,17	3,73	2,70	3,55	3,41	3,34	3,09	—	1,62	2,82
V	1,63	2,00	1,91	0,59	1,18	2,20	1,95	2,79	2,98	2,52	3,12	0,00	3,51	3,08	2,37	2,02	0,00	1,00	2,23
X	—	—	—	—	2,13	—	1,00	0,00	2,00	0,00	1,04	—	1,50	—	1,41	—	—	—	1,00
Y	2,79	2,04	3,18	2,73	2,93	2,94	2,61	3,02	3,79	3,11	3,13	2,78	3,33	3,29	3,23	2,77	—	1,42	2,23
Z	2,80	2,40	3,20	2,97	3,11	3,21	3,17	3,29	3,55	3,06	3,44	2,15	3,38	3,29	3,00	3,04	—	1,90	2,95
	3,28	3,06	3,29	3,24	3,47	3,38	3,34	3,44	3,69	3,35	3,55	3,05	3,46	3,58	3,51	3,29	0,47	2,48	3,26

11/A TÁBLÁZAT

Összefüggő szövegből csak mássalhangzók megtartásával keletkezett betűsorozat gyakoribb betűhármasai és azok relatív gyakoriságai százezredekben kifejezve (1. lap)

B	BN 103, NK 115, NM 127, NS 140, RZ 86
CS	K 123, L 197
D	LK 132, LM 293, LT 127, TT 195, SZ 115
F	LL 160, LT 82, LZ 90, RD 90
G	LS 95, NY 136, SZ 156, TT 95
GY	K 123, L 144, MS 103, NT 169, RT 111, SZ 189, T 164, Z 99
H	GY 325, LT 82
J	LN 136, LD 136, TT 82
K	BN 99, CS 115, KL 173, KR 86, LL 82, LN 86, LS 123, LT 214, MG 127, ML 86, MM 107, MN 115, NT 136, NY 98, RL 99, RS 132, RT 111, SS 136, SZ 247, TK 167, TR 111, TS 82, TT 119, VL 99, ZD 86, ZK 103, ZN 82, ZT 222, ZS 90
L	DL 111, DS 181, GY 95, KD 82, KL 115, KM 90, KS 186, KT 218, KZ 119, LK 127, LL 99, LM 123, LN 234, LS 83, LT 247, MB 95, ML 103, MN 132, MR 173, MS 107, MT 107, MZ 107, NK 127, NS 164, NT 164, PT 95, SM 90, SS 127, ST 243, SZ 255, TB 95, TK 189, TL 156, TM 95, TN 177, TR 263, TS 193, TT 407, TV 90, TZ 160, ZM 140
LY	K 103, N 123, T 82
M	BN 115, BR 243, GL 82, GS 119, KR 86, LT 103, LY 243, MN 127, ND 185, NK 169, NS 95, NT 193, NY 251, RT 127, RX 214, SK 95, SL 82, ST 148, SZ 177, TR 181, TT 90
N	BN 115, CS 119, DN 119, DL 136, DS 95, GY 259, GL 148, KK 90, KL 215, KM 115, KR 86, KS 206, KT 86, KV 82, KZ 144, LM 123, MG 99, MR 95, NK 111, SG 115, ST 140, SZ 160, TK 181, TM 99, TR 86, TS 152, TT 164, ZT 115
NY	K 103, L 127, M 99, R 103, S 173, T 95

11. TÁBLÁZAT

Összefüggő szövegből csak mássalhangzók megtartásával keletkezett betűsorozat gyakoribb betűhármasai és azok relatív gyakoriságai százezrelékben. Folytatás

P	RL	TT													
	[illegible]	70													
R	DL	GN	KT	KZ	LM	LS	LT	MN	MS	ND	NY	RD	SD	SG	SZ
	144	103	90	115	95	144	177	107	127	164	99	95	164	103	312
	TK	TL	TN	TR	TS	TT	XZ	ZS							
	103	144	169	86	140	123	86	140							
S	BN	DL	FL	GK	GS	GT	KL	KR	KT	LK	LL	LT	MG	MN	MR
	99	193	86	86	99	148	86	156	95	120	103	136	99	86	111
	MT	NK	SG	SZ	TK	TL	TN	TR	TS	TT	TZ				
	82	136	115	477	169	136	107	234	90	99	160				
SZ	B	G	K	L	M	N	R	T	[illegible]						
	90	86	247	333	800	333	489	551	132						
T	BB	BN	DM	GY	HL	JK	FL	HT	KK	KL	KM	KN	KR	KS	KT
	103	95	127	103	82	107	103	111	107	148	99	115	115	206	136
	KZ	LJ	LK	LL	LM	LN	LS	LT	LY	MG	ML	MN	MR	MT	NG
	185	82	123	132	156	247	132	136	115	123	107	144	181	82	90
	NK	NL	NT	PR	RK	RL	RM	RN	RS	RT	SG	SK	SL	SM	SN
	169	156	107	90	95	189	189	86	259	312	160	148	86	95	90
	SS	ST	SZ	TD	TH	TK	TL	TM	TN	TP	TR	TS	TT	VT	ZT
	140	177	296	107	86	275	193	214	132	86	190	338	477	99	95
V	GY	LG	LT	NY	SZ	ZT									
	127	148	247	115	123	86									
Y	SZ														
	169														
Z	KS	LL	LT	MB	MS	NB	NK	NT	NY	RK	RL	RN	RT	ST	SZ
	111	144	181	177	238	95	86	115	107	115	90	107	86	82	181
	TK	TL	TN	TR	TS	TT									
	99	222	144	86	132	399									

4. Mássalhangzókra redukált szövegek vizsgálata

Az egyik szerző és SIMON JUDIT publikálatlan eredményei azt mutatják, hogy az írásjelek, magánhangzók és szóköz elhagyásával keletkezett hiányos szövegeket jóképességű kísérleti egyének 3% alatti hibával rekonstruálni tudnak. (Ha az elhagyott betűk helyét jelezzük, a rekonstrukció általában problémamentes.) A redukált szövegek még szintén rövidíthetők a bennük előforduló betűk statisztikai függősége következtében. Egy-egy konkrét kódolásuk segítségével az elvben elérhető rövidítésre felső korlátot adhatunk. Becslés nyerhető azonban pusztán a redukált szövegek entrópiájának becslésével is. Ez volt az oka, hogy gyakorisági vizsgálatainkat kiterjesztettük mássalhangzókra redukált szövegek elemzésére is. Ennek során a *Q* és *W* betűket figyelmen kívül hagytuk, elhanyagolhatóan kicsi gyakoriságuk miatt. Így a redukált ábécé 19 betűt tartalmazott, ami lehetővé tette, hogy *„trigramm"*

(betűhármas) statisztikát is készítsünk. Természetesen a betűhármasok közül csak a gyakoribbak relatív gyakorisága fogadható el megbízhatónak, hisz az összes (elvileg) lehetséges betűhármas számához ($19^3=6859$) viszonyítva a 24 320 betűből álló redukált szöveg kicsinek tekintendő.

7. Táblázat: Mássalhangzókra redukált szöveg szomszédos párjainak relatív gyakoriságai százezredekben kifejezve. Az utolsó sor a betűk relatív gyakoriságát tízezredekben adja meg. A táblázat alapján számított entrópiák:

$$H(0)=\log_2 19 \quad =4{,}25,$$
$$H(1)=H(\text{betű}) \quad =3{,}87,$$
$$H(2)=H(\text{betű/megelőző betű}) \quad =3{,}60.$$

8. Táblázat: Mássalhangzókra redukált szöveg kettő távolságra levő $x \cdot x$ típusú betűpárjainak relatív gyakoriságai tízezredekben kifejezve.

9. Táblázat: Mássalhangzókra redukált szöveg betűinek harmadrendű $H(\xi_{i+2}|\xi_{i+1}=\cdot, \xi_i=\cdot)$ feltételes entrópiái a megelőző két betű függvényében. Az átlagos feltételes entrópia értéke

$$H(3) = H(\xi_{i+2}|\xi_{i+1}, \xi_i) = 3{,}23.$$

11/B TÁBLÁZAT

a 25 leggyakoribb betűhármas csak mássalhangzókból álló hiányos szöveg esetén (a gyakoriság százezredben van megadva)

SZT	551
SZR	489
SSZ	477
TTT	477
LTT	407
ZTT	399
TTS	338
SZL	333
SZN	333
HGY	325
RSZ	312
TRT	312
SZM	300
TSZ	296
DLM	292
TTK	275
LTR	263
NGY	259
TRS	259
LSZ	255
MNY	251
LLT	247
SZK	247
TLN	247
VLT	247

11/C TÁBLÁZAT

Csak mássalhangzókból álló hiányos szöveg betűhármasainak gyakorisági eloszlása

Relatív gyakoriság	Betűhármasok száma
≧0,00300	13
0,00247—0,00299	13
0,00201—0,00246	14
0,00150—0,00200	49
0,00100—0,00149	125
0,00090—0,00099	54
0,00080—0,00089	42
0,00071—0,00079	46
0,00061—0,00070	101
0,00051—0,00060	94
0,00041—0,00050	194
0,00031—0,00040	180
0,00021—0,00030	451
0,00011—0,00020	539
0,00006—0,00010	481
0,00001—0,00005	786
0 (egyszer sem)	3677

Tekintettel arra, hogy a mássalhangzók 0,513 relatív gyakorisággal fordulnak elő a betűkre és szóközre korlátozott szövegben (lásd 1. táblázat), ez az érték a

$$H(\text{magyar újságszöveg}) \leqq 1{,}65$$

entrópiabecslést eredményezi.

10. Táblázat: Mássalhangzókra redukált szöveg betűinek harmadrendű $H(\xi_{i+1}|\xi_i = \cdot, \xi_{i+2} = \cdot)$ feltételes entrópiái a közrefogó két betű függvényében. Az utolsó sorban az $1/2\, H(\xi_{i+2}, \xi_{i+1}|\xi_i = \cdot)$ feltételes entrópiák vannak feltüntetve. Ezek átlagos értéke $1/2\, H(\xi_{i+2}, \xi_{i+1}|\xi_i) = 3{,}41$.

11. Táblázat: Mássalhangzókra redukált szöveg gyakoribb betűhármasai alfabetikusan (11/A), a 25 leggyakoribb betűhármas gyakorisági sorrendben (11/B) és a betűhármasok gyakoriságának eloszlása (11/C táblázat). Érdemes felfigyelni arra, hogy a betűhármasoknak több mint fele egyáltalán nem fordult elő.

IRODALOM

[1] NEMETZ, T., "Entropy estimation via reconstruction of mutilated texts", *Publications de Colloque International du C.N.R.S., Cachon, France,* 4—8 Julliet 1977.

[2] NEMETZ, T. and SIMON, J., "On estimating the entropy of written Hungarian by gambling technique", *Transactions of the 8th Prague Conference on Information Theory etc. Prague,* 1978, Volume B, pp. 69—76.

[3] NEWMAN, E. B. and GERSTMAN, L. J., "A new method for analyzing printed English", *Journ. of Experimental Psychology* **44** (1952) 114—125.

[4] PETŐFI, S. J., „A nyelvstatisztikai vizsgálatok néhány kérdése", *Nyelvfeldolgozás és dokumentáció* (Szerk. Szépe György) *A tudományos tájékoztatás elmélete és gyakorlata,* **11** (1967), OMKDK, 117—140. old.

[5] RÉNYI, A., *Foundations of Probability* (Holden Day Inc., San Francisco, 1970).

[6] ЯГЛОМ, А. М. и ЯГЛОМ, И. М., Вероятность и информация, Издание третье (Наука, Москва, 1973).

(Beérkezett: 1979. február 8.)

NEMETZ TIBOR
MTA MATEMATIKAI KUTATÓ INTÉZET
1053 BUDAPEST, REÁLTANODA U. 13—15.

SZILLÉRY ANDRÁS
MTA SZÁMÍTÁSTECHNIKAI ÉS AUTOMATIZÁLÁSI KUTATÓ INTÉZET
1132 BUDAPEST, VICTOR HUGO U. 18—22.

HUNGARIAN LANGUAGE-STATISTICS

T. NEMETZ and A. SZILLÉRY

11 tables summarize the result of a statistical examination of Hungarian newspaper texts which was carried out with the aim of getting welldocumented upper bound on the entropy of written Hungarian, also aimed at searching for possibilities of effective enshortening of written texts. Brief comments are given to each individual table.

Tekintettel arra, hogy a mássalhangzók 0,513 relatív gyakorisággal fordulnak elő a betűkre és szóközre korlátozott szövegben (lásd 1. táblázat), ez az érték a

$$H(\text{magyar újságszöveg}) \leqq 1{,}65$$

entrópiabecslést eredményezi.

10. Táblázat: Mássalhangzókra redukált szöveg betűinek harmadrendű $H(\xi_{i+1}|\xi_i=\cdot, \xi_{i+2}=\cdot)$ feltételes entrópiái a közrefogó két betű függvényében. Az utolsó sorban az $1/2\, H(\xi_{i+2}, \xi_{i+1}|\xi_i=\cdot)$ feltételes entrópiák vannak feltüntetve. Ezek átlagos értéke $1/2\, H(\xi_{i+2}, \xi_{i+1}|\xi_i)=3{,}41$.

11. Táblázat: Mássalhangzókra redukált szöveg gyakoribb betűhármasai alfabetikusan (11/A), a 25 leggyakoribb betűhármas gyakorisági sorrendben (11/B) és a betűhármasok gyakoriságának eloszlása (11/C táblázat). Érdemes felfigyelni arra, hogy a betűhármasoknak több mint fele egyáltalán nem fordult elő.

IRODALOM

[1] Nemetz, T., "Entropy estimation via reconstruction of mutilated texts", *Publications de Colloque International du C.N.R.S., Cachon, France*, 4—8 Juillet 1977.

[2] Nemetz, T. and Simon, J., "On estimating the entropy of written Hungarian by gambling technique", *Transactions of the 8th Prague Conference on Information Theory etc. Prague*. 1978, Volume B, pp. 69—76.

[3] Newman, E. B. and Gerstman, L. J., "A new method for analyzing printed English", *Journ. of Experimental Psychology* **44** (1952) 114—125.

[4] Petőfi, S. J., „A nyelvstatisztikai vizsgálatok néhány kérdése", *Nyelvfeldolgozás és dokumentáció* (Szerk. Szépe György) *A tudományos tájékoztatás elmélete és gyakorlata*. 11 (1967), OMKDK, 117—140. old.

[5] Rényi, A., *Foundations of Probability* (Holden Day Inc., San Francisco, 1970).

[6] Яглом, А. М. и Яглом, И. М., Вероятность и информация, Издание третье (Наука, Москва, 1973).

(Beérkezett: 1979. február 8.)

NEMETZ TIBOR
MTA MATEMATIKAI KUTATÓ INTÉZET
1053 BUDAPEST, REÁLTANODA U. 13—15.

SZILLÉRY ANDRÁS
MTA SZÁMÍTÁSTECHNIKAI ÉS AUTOMATIZÁLÁSI KUTATÓ INTÉZET
1132 BUDAPEST, VICTOR HUGO U. 18—22.

HUNGARIAN LANGUAGE-STATISTICS

T. Nemetz and A. Szilléry

11 tables summarize the result of a statistical examination of Hungarian newspaper texts which was carried out with the aim of getting welldocumented upper bound on the entropy of written Hungarian, also aimed at searching for possibilities of effective enshortening of written texts. Brief comments are given to each individual table.

TÖMEGKISZOLGÁLÁSI MODELLEK ALKALMAZÁSA SZÁMÍTÓGÉPES RENDSZEREKRE

ASZTALOS DOMONKOS

Budapest

Jelen dolgozat célja az egy központi egységből és több perifériális egységből álló multiprogramozott számítógépek működésének vizsgálata véges forrású tömegkiszolgálási modellek felhasználásával. Két probléma tárgyalása képezi a dolgozat gerincét. Az első a központi egység foglaltsági periódusának vizsgálata, amely során bizonyításra kerül, hogy statisztikailag különböző fogyasztók (programok) kiszolgálása során a foglaltsági periódus várható értéke a kiszolgálási diszciplínák egy széles osztályára, amely tartalmazza az abszolút prioritásos és FIFO diszciplínákat is, azonos. A második probléma az abszolút prioritásos diszciplína esetén a központi egységnél levő fogyasztók számának várható értékét minimalizáló prioritások meghatározásával foglalkozik. Numerikus példák szemléltetik az eredményeket.

1. Bevezetés

Az egy központi egységből és több perifériális egységből álló multiprogramozott számítógépek működése jól modellezhető zárt tömegkiszolgálási hálózatokkal. Mindegyik program megfelel egy fogyasztónak, a központi egység és a perifériák pedig a kiszolgáló egységeknek.

Ha a gyakorlatban az esetek túlnyomó többségében egy adott perifériát csak egy program használhat, akkor az összes periféria együttesen jól reprezentálható egy korlátlan kapacitású kiszolgáló egységgel. A központi egységnek egy másik kiszolgáló egység felel meg, amelynél a számolásra várakozó programok sorbaállnak. Egy adott számítógépes rendszerben az éppen végrehajtás alatt álló programok száma, azaz a multiprogramozás szintje, viszonylag hosszabb ideig állandó. Minden ilyen időintervallumra az előbb felvázolt zárt modell egy véges forrású tömegkiszolgálási rendszerként tárgyalható, amelyben a forrás a korlátlan kapacitású kiszolgáló egységet helyettesíti. A kiszolgáló egység a központi egységnek felel meg. A továbbiakban a dolgozat a véges forrású tömegkiszolgálási rendszerek foglaltsági periódusának várható értékének kiszámításával, valamint optimális vezérlésének meghatározásával foglalkozik a stacionárius esetben. A kapott eredmények felhasználhatók konkrét számítógépes rendszerek működésének közelítő leírására [1].

Jelen dolgozat exponenciális struktúrájú rendszerekkel foglalkozik, ami azt jelenti, hogy a fogyasztó által mind a kiszolgálási egységnél, mind a forrásnál eltöltött idő exponenciális eloszlású valószínűségi változó. A dolgozatban ismertetett eredmények olyan véges forrású tömegkiszolgálási rendszerekre vonatkoznak, amelyekben a fogyasztók statisztikai leírásában csak a kiszolgáló egységnél van eltérés, a forrásnál eltöltött idő minden fogyasztóra nézve azonos paraméterű exponenciális eloszlású valószínűségi változó. A továbbiakban az egyszerűbb hivatkozás kedvéért a fenti típusú véges forrású tömegkiszolgálási rendszereket homogén

forrású rendszereknek nevezzük. Az olyan rendszereket, amelyekben a forrás is inhomogén, egyszerűen inhomogén rendszereknek, amelyekben a forrás is és a kiszolgálás is homogén, homogén rendszereknek nevezzük, végül a homogén kiszolgálású, de inhomogén forrású rendszereket homogén kiszolgálású rendszereknek nevezzük. Az inhomogén rendszerek viszonylag kimerítő tárgyalását adja JAISWAL könyve [3], különös hangsúllyal a prioritásos kiszolgálási diszciplinákra vonatkozóan. Az inhomogén rendszerekben a kiszolgáló egység foglaltsági periódusának vizsgálatával foglalkozik különböző kiszolgálási diszciplinák figyelembevételével a [4] és [5] dolgozat. Ezekben a szerző algoritmust ad a foglaltsági periódus várható értékének meghatározására, amely egyben lehetővé teszi a kiszolgáló egység kihasználtságának meghatározását. A [4, 5] dolgozatok numerikus eredményei arra utalnak, hogy homogén forrású rendszerben a foglaltsági periódus várható értéke nem függ a kiszolgálási diszciplinától, ha az az abszolút prioritásos, FIFO, osztott processzoros diszciplinák valamelyike, valamint abszolút prioritásos kiszolgálás esetén független a prioritások szétosztásának módjától. A dolgozat első részében bebizonyítjuk, hogy homogén forrású rendszerekben a fenti kiszolgálási diszciplinák bármelyikére a foglaltsági periódus várható értéke azonos, ha a fogyasztók összetétele változatlan marad. Ezekből az eredményekből kiindulva bizonyítható, hogy homogén forrású rendszerekben a foglaltsági periódus várható értéke invariáns az ún. konzervatív, osztatlan kiszolgálójú diszciplinák osztályában.

Miután ismert, hogy abszolút prioritásos[1] homogén forrású rendszerben a prioritások szétosztásának módja nincs hatással a kiszolgáló egység kihasználtságára, természetszerűleg vetődik fel a probléma, miként kell a prioritások szétosztását elvégezni a kiszolgáló egységnél tartózkodó fogyasztók átlagos számának minimalizálásához. A [2] dolgozatban lett először bizonyítva, hogy a végtelen tárolójú exponenciális struktúrájú abszolút vagy viszonylagos[2] prioritásos rendszereknél a fenti feladat optimalizálásához a prioritások szétosztását a kiszolgálási idők várható értékének növekvő sorrendjében kell végrehajtani. A dolgozat második részében megmutatjuk, hogy véges homogén forrású rendszereknél is a fenti szabály alkalmazása adja az optimális megoldást. Megemlítjük, hogy a [6] dolgozatban a szerző bebizonyította, hogy olyan inhomogén rendszerben, amelyben a kiszolgáló egység terhelése alacsony, az optimális prioritáselosztást szintén a kiszolgálási idők várható értékének növekvő sorrendje határozza meg. A dolgozat harmadik része a fenti eredményeket szemléltető numerikus eredményeket tartalmaz.

2. A foglaltsági periódus vizsgálata

Egy véges homogén forrású exponenciális struktúrájú tömegkiszolgálási rendszer elemei a következők. Adva van egy szériális kiszolgáló egység amelynek rögzített számú, N darab fogyasztót kell kiszolgálnia. Az N darab fogyasztóhoz hozzárendeljük sorszámként az első N pozitív egész számot. Az i-edik fogyasztó egyszeri kiszolgálásának ideje μ_i paraméterű exponenciális eloszlású valószínűségi változó.

[1] Ebben a dolgozatban, ha abszolút prioritásos kiszolgálásról beszélünk, mindig feltesszük, hogy a megszakított fogyasztó kiszolgálása a megszakítás helyétől folytatódik *(preemptive-resume)*.

[2] Viszonylagos prioritásos rendszerben az eredmény igaz marad, ha a kiszolgálási idők általános eloszlású valószínűségi változók.

Miután egy fogyasztó kiszolgálása befejeződött, azonnal átkerül a forrásba, ahol eltölt egy bizonyos időt, majd ezután megjelenik a kiszolgáló egységnél újabb szolgáltatási igénnyel. A forrásnál eltöltött egyszeri idő minden fogyasztó részére egy λ paraméterű exponenciális eloszlású valószínűségi változó. A fenti valószínűségi változók kölcsönösen függetlenek. Könnyen belátható, hogy a leírt rendszer ekvivalens egy olyan két kiszolgáló egységet tartalmazó zárt tömegkiszolgálási hálózattal, amelyben a forrásnak egy végtelen kapacitású kiszolgáló egység felel meg. A tömegkiszolgálási rendszer meghatározását a kiszolgálási diszciplina leírása teszi teljessé.

Tekintsük először az *osztott processzoros* kiszolgálási diszciplinát inhomogén rendszerben. Az osztott processzoros kiszolgálás azt jelenti, hogy a kiszolgálási igénnyel rendelkező fogyasztók kiszolgálása megszakítás nélkül, de változó intenzitással megy végbe, éspedig, ha a τ időtartam alatt a kiszolgálási fázisban levő fogyasztók száma mindvégig k, akkor ezen fogyasztók kiszolgálási ideje nem τ-val, hanem csak τ/k értékkel halad előre. Jelölje $S=(b_1, \ldots, b_N)$ a rendszer állapotát, ahol $b_i=0$ vagy 1 annak megfelelően, hogy az i-edik fogyasztó a forrásnál, illetve a kiszolgálási egységnél található. A rendszernek összesen 2^N állapota van. Felírjuk a $P_t(S)$ valószínűségekre a differenciálegyenletek rendszerét:

$$P_t'(S) = -P_t(S)\left(\sum_{i=1}^{N} b_i\mu_i \Big/ \sum_{i=1}^{N} b_i + \sum_{i=1}^{N} |b_i-1|\lambda_i\right) +$$

$$+ \sum_{j:b_j=0} P_t(S^{j+})\mu_j \Big/ \left(\sum_{i=1}^{N} b_i+1\right) + \sum_{j:b_j=1} P_t(S^{j-})\lambda_j.$$

Az S^{j+} és S állapotok csak a j-edik komponensükben különbözhetnek. Ha $b_j=0$, akkor $b_j^{j+}=1$. Ha $b_j=1$, akkor a két állapot azonos. Hasonlóan értelmezhető az S^{j-} és S állapotok viszonya. Ha $b_j=1$, akkor $b_j^{j-}=0$, ha $b_j=0$, akkor a két állapot azonos. Az adott rendszernek léteznek a stacionárius állapot valószínűségei, vagyis léteznek a $\lim_{t\to\infty} P_t(S)=P(S)$ határértékek és $\lim_{t\to\infty} P_t'(S)=0$.

Így egy 2^N ismeretlenes lineáris algebrai egyenletrendszert kapunk, amelynek alakja:

$$P(S)\left(\sum_{i=1}^{N} b_i\mu_i \Big/ \sum_{i=1}^{N} b_i + \sum_{i=1}^{N} |b_i-1|\lambda_i\right) = \tag{2.1}$$

$$= \sum_{j:b_j=0} P(S^{j+})\mu_j \Big/ \left(\sum_{i=1}^{N} b_i+1\right) + \sum_{j:b_j=1} P(S^{j-})\lambda_j.$$

Egyszerű helyettesítéssel könnyen ellenőrizhető, hogy a fenti egyenletrendszer megoldása

$$P(S) = C^{-1}\left(\sum_{i=1}^{N} b_i\right)! \prod_{i=1}^{N}\left(\frac{1}{\mu_i}\right)^{b_i} \prod_{i=1}^{N}\left(\frac{1}{\lambda_i}\right)^{|b_i-1|}, \tag{2.1'}$$

ahol C a normalizáló konstans,

$$C = \sum_{S}\left(\sum_{i=1}^{N} b_i\right)! \prod_{i=1}^{N}\left(\frac{1}{\mu_i}\right)^{b_i} \prod_{i=1}^{N}\left(\frac{1}{\lambda_i}\right)^{|b_i-1|}. \tag{2.2}$$

Ezt az összeget átírhatjuk az alábbi formába:

$$C = \prod_{i=1}^{N} \frac{1}{\lambda_i} \left(1 + 1! \sum_{j=1}^{N} \frac{\lambda_j}{\mu_j} + 2! \sum_{j_1, j_2} \frac{\lambda_{j_1}\lambda_{j_2}}{\mu_{j_1}\mu_{j_2}} + \dots + N! \prod_{j=1}^{N} \frac{\lambda_j}{\mu_j} \right). \tag{2.2}$$

Legyen $v_i = \lambda_i/\mu_i$, $i = 1, \dots, N$. Ekkor

$$C = \prod_{i=1}^{N} \frac{1}{\lambda_i} \left[1 + \sum_{j=1}^{N} j!\, C_j(\bar{v}) \right], \tag{2.3}$$

ahol $C_j(\bar{v})$ az N darab v_i értékből képezhető j darab különböző v_i tényezőből álló összes lehetséges szorzatok összegét jelöli. A $C_j(\bar{v})$ kiszámítására az alábbi egyszerű rekurzív összefüggés használható:

$$C_0^{(0)}(\bar{v}) = 1, \quad C_0^{(i)}(\bar{v}) = 0,$$

$$C_j^{(i)}(\bar{v}) = C_j^{(i-1)}(\bar{v}) + v_i C_{j-1}^{(i-1)}(\bar{v}),$$

ahol $i = 1, \dots, N$ és $j = 1, \dots, i$ és $C_j^{(i)}(\bar{v})$ azt jelöli, hogy a j tényezős szorzatok képzése során a $\bar{v}$ vektornak csak az első i darab komponensét vesszük figyelembe. Ha $j > i$, akkor $C_j^{(i)}(\bar{v}) = 0$.

Ezután a foglaltsági periódus várható értéke (2.1′) és (2.3) alapján a stacionárius esetben

$$E\delta_{OP}^{(N)} = \frac{1}{\sum_{i=1}^{N} \lambda_i} \cdot \frac{1 - P(0)}{P(0)} = \frac{1}{\sum_{i=1}^{N} \lambda_i} \sum_{j=1}^{N} j!\, C_j(\bar{v}), \tag{2.4}$$

ahol $P(0)$ annak stacionárius valószínűsége, hogy a kiszolgáló egység nyugalmi állapotban található. A várható értéknél az OP index az osztott processzoros diszciplínát jelöli.

Így az [5] dolgozatban leírt algoritmusnál (2.4) alapján egy sokkal hatékonyabb eljárásunk van az $E\delta_{OP}^{(N)}$ meghatározására. Az eljárás számítógépes realizációjában $7N$ tárolóegységre (szó) van szükség az adatok részére. Egy másik dolgozatban bemutatjuk, hogy a fenti rendszer egyéb paramétereinek (pl. az i-edik fogyasztó ciklusideje) meghatározására is megadhatók a fentihez hasonló hatékony számítási eljárások.

Tekintsük most a véges homogén forrású rendszert. Ebben az esetben a $\lambda_i = \lambda$, $i = 1, \dots, N$ helyettesítéssel (2.4)-ből

$$E\delta_{OP}^{(N)} = \frac{1}{N\lambda} \sum_{j=1}^{N} j!\, C_j(\bar{v}), \quad v_j = \lambda/\mu_j, \quad j = 1, \dots, N. \tag{2.5}$$

A későbbiekben erre az eredményre lesz szükségünk.

Tekintsünk egy olyan véges homogén forrású tömegkiszolgálási rendszert, amelyben a kiszolgálás az *abszolút prioritás* elvén alapszik. Ennek értelmében az N darab fogyasztó mindegyikéhez hozzárendeljük az első N pozitív egész szám valamelyikét. Ugyanaz a szám nem lehet két fogyasztóhoz hozzárendelve. Az i-edik fogyasztóhoz rendelt számot a fogyasztó prioritásának nevezzük és j_i-vel jelöljük. Ha $j_i < j_l$, akkor azt mondjuk, az i-edik fogyasztó magasabb prioritású, mint az

l-edik fogyasztó. Az abszolút prioritásos rendszer azt jelenti, hogy ha egy j_l prioritású fogyasztó érkezik a kiszolgáló egységhez és egy alacsonyabb prioritású j_i ($j_i > j_l$) fogyasztó van kiszolgálás alatt, akkor az l-edik fogyasztó kiszolgálása megszakad és azonnal elkezdődik az l-edik fogyasztó kiszolgálása. Az l-edik fogyasztó kiszolgálása a megszakítás helyétől folytatódik az első olyan időpontban, amikor nincs magasabb prioritású fogyasztó a kiszolgáló egységnél. Exponenciális struktúrájú véges inhomogén rendszerekre a foglaltsági periódus várható értéke a stacionárius esetben két rekurzív formula segítségével határozható meg [4]. Ugyanebben a dolgozatban meg lett mutatva, hogy $N=2$ és homogén forrású rendszer esetén a foglaltsági periódus várható értéke nem változik a prioritások felcserélése során. A következőkben megmutatjuk, hogy ez N tetszőleges értékére is igaz.

2.1. TÉTEL. Exponenciális struktúrájú, véges, homogén forrású rendszerben abszolút prioritásos kiszolgálási diszciplína esetén tetszőleges N-re ($N=1, 2, \ldots$) a kiszolgáló egység foglaltsági periódusának várható értéke $E\delta_{\mathrm{AP}}^{(N)}$, a stacionárius esetben független a prioritások elosztásától.

Bizonyítás: Tekintsük a [4]-ben közölt rekurzív formulákat (ahol a bizonyítás során az AP indexet elhagyjuk):

$$(2.6)\qquad E\delta^{(N)}=\frac{N-1}{N}\left[E\delta^{(N-1)}+\frac{1}{\mu_N}\left(1-\varphi_{N-1}(\lambda)\right)\left(1+(N-1)\lambda E\delta^{(N-1)}\right)\right]+\frac{1}{N\mu_N}\left(1+(N-1)\lambda E\delta^{(N-1)}\right),$$

$$(2.7)\qquad \varphi_N(s)=\frac{N-1}{N}\left[\varphi_{N-1}(s+\lambda)+\frac{\mu_N[\varphi_{N-1}(s)-\varphi_{N-1}(s+\lambda)]}{\mu_N+\lambda(N-1)\left(1-\varphi_{N-1}(s)\right)+s}\right]+\frac{1}{N}\cdot\frac{\mu_N}{\mu_N+\lambda(N-1)\left(1-\varphi_{N-1}(s)\right)+s},$$

ahol $\varphi_N(s)$ az N fogyasztót tartalmazó homogén forrású rendszer foglaltsági periódusának *Laplace-transzformáltját* jelöli, μ_N az N prioritású fogyasztó kiszolgálási intenzitását jelöli.

A bizonyítás az alábbi lemmán alapul.

LEMMA. A $\varphi_N(s)$ értéke tetszőleges N-re a $k\lambda$ pontokban ($k=0, 1, \ldots$) a $\mu_1, \mu_2, \ldots, \mu_N$ kiszolgálási intenzitások szimmetrikus függvénye és

$$(2.8)\qquad \varphi_N(k\lambda)=\frac{1}{N}\cdot\frac{\sum_{i=0}^{N-1}(N-i)\lambda^i C_{N-i}^{(N)}(\bar{\mu})\prod_{j=0}^{i-1}(k+j)}{\sum_{i=0}^{N}\lambda^i C_{N-i}^{(N)}(\bar{\mu})\prod_{j=0}^{i-1}(k+j)}.$$

Ha $i-1<0$, akkor $\Pi=1$ és ha $i=N$, akkor $C_0(\bar{\mu})=1$. $C_{N-i}^{(N)}(\bar{\mu})$ itt is a $\bar{\mu}=(\mu_1, \ldots, \mu_N)$ N értékből képezhető $N-i$ különböző tényezőjű összes lehetséges szorzat összegét jelöli.

Megemlítjük, hogy a lemma bizonyításával a fenti állítás egyben a $\varphi_N(s)$, $s\in[0, \infty)$ értékekre is igaz.

Bizonyítás. A lemmát a teljes indukció módszerével bizonyítjuk. Felhasználva, hogy a μ paraméterű exponenciális eloszlású valószínűségi változó *Laplace-transzformáltja* $\varphi(s)=\mu/(\mu+s)$ és azt behelyettesítve a $\varphi_N(s)$-t megadó (2.7) formulába $N=2$ esetén kapjuk, hogy

$$\varphi_2^{(1,2)}(k\lambda) = \varphi_2^{(2,1)}(k\lambda) = \frac{1}{2}\cdot\frac{2\mu_1\mu_2+k\lambda(\mu_1+\mu_2)}{\mu_1\mu_2+k\lambda(\mu_1+\mu_2)+k(k+1)\lambda^2},$$

ahol a felső zárójeles index a prioritások elosztását jelöli. Az első helyen álló szám a magasabb prioritású fogyasztó sorszámát jelöli. Így $N=2$ esetén a lemma állítása igaz. Tegyük fel, hogy igaz tetszőleges $N=n$ értékre. Felhasználva a $\varphi_{n+1}(s)$ és $\varphi_n(k\lambda)$ értékekre vonatkozó (2.7) és (2.8) formulákat, bizonyítani kell, hogy $\varphi_{n+1}(k\lambda)$ esetén is igaz a lemma állítása. A (2.7) rekurzív összefüggés alapján

$$(2.9)\quad \varphi_{n+1}(k\lambda) = \frac{n}{n+1}\left[\frac{\varphi_n((k+1)\lambda)[(k+n)\lambda-n\lambda\varphi_n(k\lambda)]+\mu_{n+1}\varphi_n(k\lambda)+\mu_{n+1}/n}{\mu_{n+1}+(k+n)\lambda-n\lambda\varphi_n(k\lambda)}\right].$$

Végezzük el a helyettesítést először a számláló első tagjára:

$$(2.10)\quad \varphi_n((k+1)\lambda)[(k+n)\lambda-n\lambda\varphi_n(k\lambda)] = \frac{\lambda}{n}\cdot\frac{\sum_{i=0}^{n-1}(n-i)\lambda^i C_{n-i}^{(n)}(\bar{\mu})\prod_{j=0}^{i-1}(k+j+1)}{\sum_{i=0}^{n}\lambda^i C_{n-i}^{(n)}(\bar{\mu})\prod_{j=0}^{i-1}(k+1+j)}\times$$

$$\times\frac{(k+n)\sum_{j=0}^{n}\lambda^i C_{n-i}^{(n)}(\bar{\mu})\prod_{j=0}^{i-1}(k+j)-\sum_{j=0}^{n-1}(n-i)\lambda^i C_{n-i}^{(n)}(\bar{\mu})\prod_{j=0}^{i-1}(k+j)}{\sum_{i=0}^{n}\lambda^i C_{n-i}^{(n)}(\bar{\mu})\prod_{j=0}^{i-1}(k+j)} =$$

$$= \frac{\lambda k}{n}\cdot\frac{\sum_{i=0}^{n-1}(n-i)\lambda^i C_{n-i}^{(n)}(\bar{\mu})\prod_{j=0}^{i-1}(k+1+j)}{\sum_{i=0}^{n}\lambda^i C_{n-i}^{(n)}(\bar{\mu})\prod_{j=0}^{i-1}(k+j)},$$

Eközben felhasználtuk, hogy a második tényező számlálójában levő különbség az azonos tagok összevonása után:

$$\sum_{i=0}^{n}\lambda^i C_{n-i}^{(n)}(\bar{\mu})\prod_{j=0}^{i}(k+j) = k\sum_{i=0}^{n}\lambda^i C_{n-i}^{(n)}(\bar{\mu})\prod_{j=0}^{i-1}(k+1+j).$$

Legyen

$$G = \sum_{i=0}^{n}\lambda^i C_{n-i}^{(n)}(\bar{\mu})\prod_{j=0}^{i-1}(k+j).$$

Ekkor a (2.9) számlálójában levő tagokat közös nevezőre hozva és n-nel egyszerűsítve kapjuk (2.10)-re a következő kifejezést:

(2.11)

$$G^{-1}\left[k\lambda\sum_{i=0}^{n-1}(n-i)\lambda^i C_{n-i}^{(n)}(\bar{\mu})\prod_{j=0}^{i-1}(k+1+j)+\sum_{i=0}^{n}(n-i+1)\lambda^i C_{n-i}^{(n)}(\bar{\mu})\mu_{n+1}\prod_{j=0}^{i-1}(k+j)\right]=$$

$$=G^{-1}\left[\sum_{i=1}^{n-1}(n-i+1)\lambda^i C_{n-i+1}^{(n)}(\bar{\mu})\prod_{j=0}^{i-1}(k+j)+\sum_{i=1}^{n-1}(n-i+1)\lambda^i C_{n-i}^{(n)}(\bar{\mu})\mu_{n+1}\prod_{j=0}^{i-1}(k+j)+\right.$$

$$\left.+(n+1)C_{n+1}^{(n+1)}(\bar{\mu},\mu_{n+1})+\mu_{n+1}\lambda^n\prod_{j=0}^{n-1}(k+j)+\lambda^n C_1^{(n)}(\bar{\mu})\sum_{j=0}^{n-1}(k+j)\right]=$$

$$=G^{-1}\sum_{i=0}^{n}(n+1-i)\lambda^i C_{n+1-i}^{(n+1)}(\bar{\mu},\mu_{n+1})\prod_{j=0}^{i-1}(k+j).$$

Eközben felhasználtuk, hogy

$$C_{n+1-i}^{(n+1)}(\bar{\mu},\mu_{n+1})=\mu_{n+1}C_{n-i}^{(n)}(\bar{\mu})+C_{n+1-i}^{(n)}(\bar{\mu}).$$

Tekintsük most a (2.9) kifejezésben szereplő nevezőt az $n+1$ szorzó nélkül:

(2.12)
$$\mu_{n+1}+(k+n)\lambda-n\lambda\varphi_n(k\lambda)=G^{-1}\left[\mu_{n+1}\sum_{i=0}^{n}\lambda^i C_{n-i}^{(n)}(\bar{\mu})\prod_{j=0}^{i-1}(k+j)+\right.$$

$$\left.+\lambda(k+n)\sum_{i=0}^{n}\lambda^i C_{n-i}^{(n)}(\bar{\mu})\prod_{j=0}^{i-1}(k+j)-\sum_{i=0}^{n-1}(n-i)\lambda^{i+1}C_{n-i}^{(n)}(\bar{\mu})\prod_{j=0}^{i-1}(k+j)\right]=$$

$$=G^{-1}\left[\sum_{i=0}^{n}\lambda^i C_{n-i}^{(n)}(\bar{\mu})\mu_{n+1}\prod_{j=0}^{i-1}(k+j)+\sum_{i=0}^{n}\lambda^{i+1}C_{n-i}^{(n)}(\bar{\mu})\prod_{j=0}^{i}(k+j)\right]=$$

$$=G^{-1}\left[\sum_{i=1}^{n}\lambda^i C_{n-i+1}^{(n+1)}(\bar{\mu},\mu_{n+1})\prod_{j=0}^{i-1}(k+j)+\lambda^{n+1}\prod_{j=0}^{n}(k+j)+C_{n+1}^{(n+1)}(\bar{\mu},\mu_{n+1})\right]=$$

$$=G^{-1}\sum_{i=0}^{n+1}\lambda^i C_{n+1-i}^{(n+1)}(\bar{\mu},\mu_{n+1})\prod_{j=0}^{i-1}(k+j).$$

Így (2.11) és (2.12) alapján

$$\varphi_{n+1}(k\lambda)=\frac{1}{n+1}\cdot\frac{\sum_{i=0}^{n}(n+1-i)\lambda^i C_{n+1-i}^{(n+1)}(\bar{\mu},\mu_{n+1})\prod_{j=0}^{i-1}(k+j)}{\sum_{i=0}^{n+1}\lambda^i C_{n+1-i}^{(n+1)}(\bar{\mu},\mu_{n+1})\prod_{j=0}^{i-1}(k+j)}.$$

A kapott formula μ_i-kre vonatkoztatott szimmetrikussága a $C_{n-i}^{(n)}(\bar{\mu})$ kifejezések meghatározása alapján, azok μ_i-kre vonatkoztatott szimmetrikusságából következik. Ezzel a lemmát bizonyítottuk.

Ezután a 2.1. tételt is a teljes indukció módszerével bizonyítjuk. A [4] dolgozatban $N=2$ esetén az állítást bizonyították. Tegyük fel, hogy igaz az állításunk tetsző-

darab fogyasztóhoz miként lettek a prioritások hozzárendelve. Bővítsük ki a fogyasztók halmazát egy α paraméterű fogyasztóval és legyen a sorszáma és prioritása n. Ekkor az $E\delta^{(N)}$-re adott (2.6) rekurzív formulából, a $\varphi_{n-1}(k\lambda)$ szimmetrikusságából és az indukciós állításból következik, hogy

$$E\delta^{(n)}(1, 2, \ldots, n-1, n) = E\delta^{(n)}(j_1, j_2, \ldots, j_{n-1}, n),$$

ahol $j_1, j_2, \ldots, j_{n-1}$ az $1, 2, \ldots, n-1$ számok egy permutációja és j_k $(k=1, 2, \ldots, n-1)$ a k prioritású fogyasztó sorszámát jelöli. Ugyanez az összefüggés igaz, ha az n fogyasztó közül bármelyiket rögzítjük a legalacsonyabb prioritású helyen, azaz

$$E\delta^{(n)}(i_1, i_2, \ldots, i_{n-1}, n) = E\delta^{(n)}(j_1, j_2, \ldots, j_{n-1}, n).$$

(2.13)

A következő lépésben megmutatjuk, hogy ha az n fogyasztó közül kettőt kiválasztunk és azokhoz a két legalacsonyabb prioritást rendeljük hozzá ($n-1$ és n), akkor

$$E\delta^{(n)}(i_1, i_2, \ldots, i_{n-2}, j, l) = E\delta^{(n)}(i_1, i_2, \ldots, i_{n-2}, l, j),$$

(2.14)

vagyis a két legalacsonyabb prioritású fogyasztó prioritásának felcserélése nem változtatja meg a foglaltsági periódus várható értékét. Valóban, a bizonyítás elején adott rekurzív formulák ismételt használatával adódik, hogy

$$E\delta^{(n)}(i_1, i_2, \ldots, i_{n-2}, j, l) = \frac{n-2}{n} \cdot \frac{(\lambda)}{\mu_j \mu_l} [\mu_j \mu_l + (n-1)\lambda(\mu_j + \mu_l) +$$

$$+ (n-1)n\lambda^2 + (n-2)^2\lambda^2\varphi_{n-2}(\lambda) - (n-2)\lambda\varphi_{n-2}(\lambda)(\mu_j + \mu_l + n\lambda) +$$

$$-(n-2)(n-1)\lambda^2\varphi_{n-2}(2\lambda)] + (n-1)(\mu_j + \mu_l + n\lambda) + (n-2)^2\lambda\varphi_{n-2}(\lambda)\varphi_{n-2}(2\lambda) -$$

$$-(n-2)[\varphi_{n-2}(\lambda)(\mu_j + \mu_l + n\lambda) + (n-2)(n-1)\lambda\varphi_{n-2}(2\lambda)] = E\delta^{(n)}(i_1, i_2, \ldots, i_{n-2}, l, j).$$

Ez utóbbi egyenlőség igaz, mivel a két egyenlőségjel között szereplő kifejezés μ_j és μ_l értékekre nézve szimmetrikus, az $E\delta^{(n-2)}$- és $\varphi_{n-2}(\cdot)$ értékek függetlenek a j-edik és l-edik fogyasztótól.

A várható értékekre vonatkozó (2.13) és (2.14) egyenlőségekből már nyilvánvalóan adódik, hogy a foglaltsági periódus várható értéke független a prioritások szétosztásától.

KÖVETKEZMÉNY.

$$E\delta_{\mathrm{AP}}^{(N)} = E\delta_{\mathrm{FIFO}}^{(N)} = E\delta_{\mathrm{LIFO}}^{(N)}.$$

A FIFO kiszolgálási diszciplína megfelel az érkezési sorrendben való kiszolgálásnak *(First-In-First-Out)*, a LIFO esetén pedig egy érkezés kiszolgálása rögtön megkezdődik és az esetleg megszakított fogyasztó kiszolgálása a megszakítás helyétől folytatódik a megszakítást okozó fogyasztó kiszolgálása után.

Bizonyítás. Az AP diszciplína esetén $E\delta_{\mathrm{AP}}^{(N)}$ független a prioritások kiosztásától, így a kiszolgálás időpontban történő megváltoztatása sincs hatással $E\delta_{\mathrm{AP}}^{(N)}$ értékére. A FIFO diszciplínával azonos kiszolgálást kapunk ha egy érkezéskor az éppen beérkező fogyasztónak rendelt prioritás megegyezik azzal a számmal, hogy hányadikként érkezett a kiszolgáló egységhez, és a korábban már a kiszolgáló

egységnél levő fogyasztók prioritását nem változtatjuk meg, illetve egy fogyasztó távozásakor a kiszolgáló egységnél maradt fogyasztók mindegyikének prioritását eggyel csökkentjük. Mindkét esetben a fennmaradt prioritásértékek tetszőlegesen oszthatók szét a forrásnál tartózkodó programok között. A LIFO diszciplinával azonos kiszolgálást kapunk, ha egy érkezéskor az éppen beérkező fogyasztó prioritása egy lesz és a korábban már a kiszolgáló egységnél tartózkodó fogyasztók prioritását eggyel növeljük, egyébként az eljárás megegyezik a FIFO kiszolgálásnál leírtakkal.

Korábban már utaltunk arra, hogy a numerikus számítások során, rögzített $N, \bar{\mu}, \lambda$ paraméterekre, az $E\delta_{AP}$ és $E\delta_{OP}$ értékek megegyeztek (l. [4, 5]).

2.2. TÉTEL. Exponenciális struktúrájú, véges, homogén forrású tömegkiszolgálási rendszerben adott $N, \bar{\mu}, \lambda$ paraméterekre $E\delta_{AP}=E\delta_{OP}$.

Bizonyítás. A bizonyítás N szerinti teljes indukcióval történik.

Ha $N=1$, akkor $E\delta_{AP}^{(1)}=E\delta_{OP}^{(1)}=\mu_1^{-1}$. Tegyük fel (2.5) felhasználásával, hogy $N=n-1$ esetén $E\delta_{AP}^{(n-1)}=E\delta_{OP}^{(n-1)}=\dfrac{1}{(n-1)\lambda}\sum_{i=1}^{n-1} i!\, C_i^{(n-1)}(\bar{\nu})$, ahol $\nu_i=\lambda/\mu_i$, $i=1,\ldots,n$. Az $E\delta_{AP}^{(n)}$-re adott (2.6) rekurzív formulát használva kapjuk, hogy

$$E\delta_{AP}^{(n)} = \frac{1}{n\lambda}\sum_{j=1}^{n-1} i!\, C_i^{(n-1)}(\bar{\nu}) + \tag{2.15}$$

$$+\frac{1}{n\lambda}\cdot\frac{n\sum_{j=0}^{n-1}\lambda^i C_{n-1-i}^{(n-1)}(\bar{\mu})\prod_{j=0}^{i-1}(j+1)-\sum_{i=0}^{n-2}(n-1-i)\lambda^i C_{n-1-i}^{(n-1)}(\bar{\mu})\prod_{j=0}^{i-1}(j+1)}{\frac{1}{\lambda}\sum_{i=0}^{n-1}\lambda^i C_{n-1-i}^{(n-1)}(\bar{\mu})\mu_n\prod_{j=0}^{i-1}(j+1)}\times$$

$$\times\left(1+\sum_{i=1}^{n-1} i!\, C_i^{(n-1)}(\bar{\nu})\right) = \frac{1}{n\lambda}\sum_{i=1}^{n-1} i!\, C_i^{(n-1)}(\bar{\nu}) +$$

$$+\frac{1}{n\lambda}\cdot\frac{\sum_{i=0}^{n-1}\lambda^{i+1} C_{n-(i+1)}^{(n-1)}(\bar{\mu})(i+1)!}{\sum_{i=0}^{n-1}\lambda^i C_{n-(i+1)}^{(n-1)}(\bar{\mu})\mu_n i!}\cdot\left(1+\sum_{i=1}^{n-1} i!\, C_i^{(n-1)}(\bar{\nu})\right).$$

Felhasználva, hogy

$$C_i^{(n)}\left(\frac{1}{\bar{\mu}}\right) = \frac{C_{n-i}^{(n)}(\bar{\mu})}{C_n^{(n)}(\bar{\mu})}$$

és

$$C_i^{(n)}(\bar{\nu}) = C_i^{(n-1)}(\bar{\nu}) + \nu_n C_{i-1}^{(n-1)}(\bar{\nu}),$$

ahol

$$\bar{\mu}=(\mu_1,\ldots,\mu_n),\quad 1/\bar{\mu}=(\mu_1^{-1},\ldots,\mu_n^{-1})\quad\text{és}\quad \bar{\nu}=(\lambda/\mu_1,\ldots,\lambda/\mu_n),$$

(2.15)-ből kapjuk:

$$E\delta_{\mathrm{AP}}^{(n)} = \frac{1}{n\lambda}\sum_{i=1}^{n-1} i!\, C_i^{(n-1)}(\bar{\nu}) + \frac{1}{n\lambda}\cdot\frac{\sum_{i=0}^{n-1}(i+1)\,C_{n-1}^{(n-1)}(\bar{\mu})\,\nu_n\,C_i^{(n-1)}(\bar{\nu})\,i!}{\sum_{i=0}^{n-1} i!\,C_{n-1}^{(n-1)}(\bar{\mu})\,C_i^{(n-1)}(\bar{\nu})}\times$$

$$\times\left(1+\sum_{i=1}^{n-1} i!\,C_i^{(n-1)}(\bar{\nu})\right) = \frac{1}{n\lambda}\left[\sum_{i=1}^{n-1} i!\,C_i^{(n-1)}(\bar{\nu}) + \sum_{i=0}^{n-1}(i+1)!\,C_i^{(n-1)}(\bar{\nu})\,\nu_n\right] =$$

$$= \frac{1}{n\lambda}\sum_{i=1}^{n} i!\,C_i^{(n)}(\bar{\nu}) = E\delta_{\mathrm{OP}}^{(n)}.$$

A következőkben további kiszolgálási diszciplinákra szélesítjük ki a 2.1. tétel következményét és a 2.2. tételt. Egy tetszőleges kiszolgálási diszciplina egyetlen kiszolgálót tartalmazó rendszerben nem más, mint egy döntési eljárás, amely meghatározza minden időpillanatra, hogy a kiszolgáló egységnél levő fogyasztók közül melyiket, vagy melyeket kell kiszolgálni. Jelölje T a rendszer működési idejének egy tetszőleges intervallumát. Ezután a kiszolgálási diszciplinák két csoportra oszthatók.

Az első csoportba azok tartoznak, amelyeknél bármely véges T intervallum felosztható véges számú diszjunkt intervallumok olyan sorozatára, hogy mindegyik intervallumban csak egy adott fogyasztó részesül kiszolgálásban. Ezeket a kiszolgálási diszciplinákat osztatlan kiszolgálójú diszciplináknak nevezzük. Ilyen kiszolgálási diszciplina pl. a FIFO, LCFS, AP, a kvantált v. ciklikus kiszolgálási diszciplina.

A második csoportba tartozik az összes többi diszciplina. Ilyen például az osztott processzoros kiszolgálás, amely a kvantált kiszolgálás határesete, mikor az időkvantum értéke nullához tart.

Egy kiszolgálási diszciplina konzervatív, ha a kiszolgáló egységnél nem vész el és nem keletkezik kiszolgálási igény.

2.3. TÉTEL. Exponenciális struktúrájú, véges, homogén forrású $N, \bar{\mu}, \lambda$ paraméterű tömegkiszolgálási rendszerben a foglaltsági periódus várható értéke a stacionárius esetben azonos minden konzervatív osztatlan kiszolgálójú diszciplinára és

$$E\delta(N, \bar{\mu}, \lambda) = \frac{1}{N\lambda}\sum_{i=1}^{N} i!\,C_i^{(N)}(\bar{\nu}),$$

ahol $\bar{\nu}$ a $(\lambda/\mu_1, \ldots, \lambda/\mu_N)$ N darab szám összességét jelöli.

Bizonyítás. A bizonyítás a 2.1. tétel eredményén alapul. Könnyen belátható, hogy az abszolút prioritásos kiszolgálási diszciplina, amelynél egy fogyasztó megszakított kiszolgálása a megszakítás helyétől folytatódik, konzervatív és osztatlan kiszolgálójú. A 2.1. tétel alapján AP diszciplina esetén $E\delta(N, \bar{\mu}, \lambda)$ értéke független a prioritások szétosztásától és így egyben független attól is, ha a prioritások szétosztása tetszőleges időpontban megváltozik. Az osztatlan kiszolgálójú diszciplinák definíciója alapján tetszőleges véges T intervallumban véges azoknak az eseteknek a száma, amikor a kiszolgálás egyik fogyasztóról átvált egy másikra. Ennek megfelelően az a rendszer, amelyben az eredeti helyett az abszolút prioritásos diszciplinát

alkalmazzuk és az eredeti diszciplina döntésének megfelelően megváltoztatjuk a prioritások elosztását, ugyanúgy viselkedik mint az eredeti rendszer (l. pl. a 2.1. tétel következménye bizonyítását). A tétel megfogalmazásában szereplő egyenlőség egyszerűen a 2.2. tétel következménye.

3. Optimális vezérlés

A 2.1. tétel alapján a kiszolgáló egység kihasználtsága független a prioritások elosztásától. Ekkor, a végtelen tárolójú rendszerekhez hasonlóan, megfogalmazható az alábbi feladat. A prioritások milyen szétosztása mellett lesz a kiszolgáló egységnél tartózkodó fogyasztók számának várható értéke, Eq, *a* stacionárius esetben minimális? Ez egy optimális vezérlési feladat, amennyiben a prioritások minden egyes elosztását vezérlésnek tekintjük. A j-edik prioritású fogyasztónak a kiszolgáló egységnél tartózkodó fogyasztók számának várható értékéhez való hozzájárulása [3] IV. fejezet alapján:

$$Eq_j = 1 - \frac{\mu_j}{\lambda}(e_{j-1} - e_j),$$

ahol

$$e_j = (1 + j\lambda E\delta^{(j)})^{-1}$$

azon esemény stacionárius valószínűsége, hogy a kiszolgáló egység szabad, ha a rendszerben j különböző prioritású fogyasztó van jelen.

3.1. TÉTEL. Exponenciális struktúrájú, véges, homogén forrású tömegkiszolgálási rendszerben abszolút prioritásos kiszolgálási diszciplina esetén tetszőleges N-re az $Eq = \sum_{j=1}^{N} Eq_j$ összeg akkor minimális, ha a fogyasztókhoz a prioritásokat a μ_j értékek csökkenő sorrendjében rendeljük hozzá.

Bizonyítás. Tekintsük az $Eq_j + Eq_{j+1}$ összeget tetszőleges $1 \leqq j < N$ esetén. Jelölje μ_j a j-edik prioritású fogyasztó kiszolgálási intenzitását. Megmutatjuk, hogy

$$(Eq_j + Eq_{j+1})_{\mu_j > \mu_{j+1}} < (Eq_j + Eq_{j+1})_{\mu_j < \mu_{j+1}}. \tag{3.2}$$

Ebből már állításunk következni fog.

$$Eq_j + Eq_{j+1} = 2 - \frac{1}{\lambda}(\mu_j e_{j-1} - \mu_{j+1} e_{j+1} - (\mu_j - \mu_{j+1}) e_j). \tag{3.3}$$

Tekintsük most csak a zárójelben szereplő kifejezést és e_j értékét meghatározó (3.1) formulában helyettesítsük $E\delta^{(j)}$ helyére az $E\delta^{(j)}$ értékét meghatározó (2.6) rekurzív formula jobb oldalát. Így

$$e_j = \frac{\mu_j e_{j-1}}{\mu_j + [j - (j-1)\varphi_{j-1}(\lambda)]\lambda}. \tag{3.4}$$

Legyen $C = \lambda[j - (j-1)\varphi_{j-1}(\lambda)]$. Nyilvánvalóan $C \geqq 0$ minden $\lambda \geqq 0$ értékre, mivel $\varphi_{j-1}(\lambda)$ a kiszolgáló egység foglaltsági periódusának mint valószínűségi változónak a Laplace-transzformáltja és így $\varphi_{j-1}(\lambda) \in (0, 1]$ minden véges λ-ra.

Ekkor a (3.3)-ban szereplő zárójeles kifejezés (3.4) alapján átírható a következő alakra:

$$\mu_j e_{j-1} - \mu_{j+1} e_{j+1} - (\mu_j - \mu_{j+1}) e_j = \mu_j e_{j-1} \frac{\mu_{j+1}+C}{\mu_j+C} - \mu_{j+1} e_{j+1} =$$

$$= \mu_j \mu_{j+1} \left(e_{j-1} \frac{\mu_{j+1}+C}{(\mu_j+C)\mu_{j+1}} - e_{j+1} \frac{1}{\mu_j} \right).$$

Az e_j-t meghatározó (3.1) formulából következik: $e_{j-1} > e_{j+1}$. e_{j-1} értéke független μ_j és μ_{j+1} értékektől és e_{j+1} értéke a 2.1. tétel értelmében nem változik meg, ha a j és $j+1$ prioritású fogyasztókat felcseréljük. Ugyanakkor, ha $\mu_j > \mu_{j+1}$, akkor $(\mu_{j+1}+C)/(\mu_{j+1}(\mu_j+C)) > 1/\mu_j$ és ha $\mu_j < \mu_{j+1}$, akkor $(\mu_{j+1}+C)/((\mu_j+C))\mu_{j+1}) < 1/\mu_j$. Ebből már következik a bizonyítás elején felírt (3.2) egyenlőtlenség.

A 3.1. tétel gyakorlati jelentősége, hogy az általa meghatározott vezérlés esetén lesz a kiszolgáló rendszer áteresztőképessége maximális. Állításunk igaz marad akkor is, ha az abszolút prioritásos diszciplinák helyett a konzervatív, osztatlan kiszolgálójú diszciplinák osztályát tekintjük (l. a 2.3. tétel bizonyításának gondolatmenetét).

4. Numerikus eredmények

Ebben a részben az előző fejezetek megállapításait szemléltető néhány numerikus eredményt mutatunk be. Az 1. táblázatban szereplő adatok $N=3$ esetre vonatkoznak és a kiszolgálás intenzitása helyett a kiszolgálási idők várható értékei szerepelnek. A numerikus eredmények az abszolút prioritásos (AP) és osztott processzoros (OP) diszciplinákra vonatkoznak. Az 1. táblázatban ϱ a kiszolgáló egység kihasználtságát jelöli, ahol $\varrho = E\delta^{(3)}/(E\delta^{(3)} + (3\lambda)^{-1})$. A táblázat utolsó oszlopa a vezérlés hatékonyságát szemlélteti az osztott processzoros diszciplinához viszonyítva, ahol $\Delta E = 100(Eq_{AP} - Eq_{OP})/Eq_{OP}$.

Minél kisebb ΔE értéke, annál hatékonyabb a prioritások elosztása. Érdekes, hogy az egyenlő kiszolgálást biztosító OP diszciplina az Eq kritérium szerint lényegesen közelebb van az optimális AP diszciplinához, mint a legrosszabbhoz. Ugyanakkor észrevehető, hogy amikor a központi egység terhelése (kihasználtsága) alacsony, akkor az Eq érték szerint nincs lényeges különbség az optimális AP és az OP diszciplinák között. Ez bizonyos fokig csökkenti az inhomogén rendszerekre vonatkozó [6]-ban közölt eredmény gyakorlati jelentőségét.

A dolgozat eredményeinek alkalmazását multiprogramozott számítógépek modellezésénél az [1] cikk tárgyalja. Elmondható, hogy ma már viszonylag fejlett matematikai eszközök állnak rendelkezésre a modellezők részére. A bonyolultabb matematikai modellek alkalmazását viszont sok esetben éppen az akadályozza, hogy nem állnak rendelkezésre azok a mérési adatok, melyek a modellek paraméterezéséhez szükségesek.

A tapasztalat alapján a dolgozatban ismertetett egyszerű modell alkalmazása jó kompromisszumos megoldás a könnyű kezelhetőség és a pontosság szempontjából.

Köszönetnyilvánítás. Ezúttal is szeretnék köszönetet mondani ARATÓ MÁTYÁSnak és TOMKÓ JÓZSEFnek azért az aktív figyelemért és sok értékes javaslatért, amellyel munkám során támogattak.

1. TÁBLÁZAT

Kiszolgáló egység kihasználtsága, vezérlések összehasonlítása ($N=3$, $\lambda^{-1}=30$)

$\bar{\mu}^{-1}$ 1 pr.	2 pr.	3 pr.	ϱ	Eq_{OP}	Eq_{AP}	$\Delta E(\%)$
2	3	4	0,2664	0,3166	0,3082	−2,63
2	4	3			0,3145	−0,66
3	2	4			0,3126	−1,26
3	4	2			0,3245	2,52
4	2	3			0,3243	2,45
4	3	2			0,3301	4,26
5	15	25	0,7616	1,2649	1,1510	−9,00
5	25	15			1,2398	−1,97
15	5	25			1,2376	−2,15
15	25	5			1,4480	14,47
25	5	15			1,4147	11,84
25	15	5			1,5309	21,03

IRODALOM

[1] ASZTALOS, D., "A hybrid simulation/analytical model of a batch computer system", Proc. of 4th International Simposium on Modelling and Performance Evaluation of Computer Systems, Vienna, February 6—8, 1979.
[2] BROSH, I. and NAOR, P., "On optimal disciplines in priority queueing", *Bull. Inst. Internat. Statist.* **40** (1963) 593—609.
[3] JAISWAL, N. K., *Priority Queues* (Academic Press, 1968).
[4] TOMKÓ, J., „Számológépek központi egységének kihasználtságáról, I.", *Alk. Mat. Lapok* **1** (1975) 319—331.
[5] TOMKÓ, J., „Számológépek központi egységének kihasználtságáról, II.", *Alk. Mat. Lapok* **3** (1977) 83—96.
[6] Веклеров, Е. Б., «Об управляемой замкнутой системе обслуживания» *Пробл. передачи информации* 1972, № 3.

(Beérkezett: 1979. március 5.)

ASZTALOS DOMONKOS
ORSZÁGOS TERVHIVATAL SZÁMÍTÁSTECHNIKAI KÖZPONTJA
1149 BUDAPEST, ANGOL U. 27.

APPLICATION OF FINITE SOURCE QUEUEING MODELS FOR COMPUTER SYSTEMS

D. ASZTALOS

The paper deals with the application of finite source queueing models of multiprogrammed computer systems consisting of one CPU and many I/O peripherals. It is shown that for any given parameter set the expected busy period of the CPU is invariant for a wide class of scheduling rules, containing the FIFO, processor sharing and preemptive-resume priority disciplines. An optimal priority allocation is derived for the preemptive-resume priority discipline minimising the expected number of requests in the CPU queue. Numerical examples illustrate the drawn results.

Alkalmazott Matematikai Lapok 5 (1979) 103—122

TÖMEGKISZOLGÁLÁS SZIMULÁCIÓJA SORBAKAPCSOLT KISZOLGÁLÓHELYEK ESETÉN

NÉMETH GYÖRGY
Budapest

A cikk sorbakapcsolt kiszolgálóhelyekkel rendelkező tömegkiszolgálási rendszerek három, különböző speciális feltételeknek eleget tevő modelljét tartalmazza. Az első két modellt mintegy bevezetésnek szánva analitikus megoldás megadása után sztochasztikus szimulációval oldjuk meg, így mód nyílik a két különböző módszerrel kapott eredmények összehasonlítására. A harmadik modell megoldását sztochasztikus szimulációval nyerjük. E módszer főleg olyan esetekben indokolt, amikor — mint itt — klasszikus számítási eljárást nem ismerünk.

1. Bevezetés

Ha valamely tömegkiszolgálási rendszer olyan m kiszolgálóhelyből (állomásból) áll, amely rendszerbe belépő minden kiszolgálandó egységnek (1) minden kiszolgálóhelyen (2) az előre egységesen definiált sorrendben végig kell haladnia teljes kiszolgálása céljából, sorbakapcsolt kiszolgálóhelyek rendszeréről beszélünk. A kiszolgálásra jelentkező egységek valamiféle sorbanállási szisztéma szerint lépnek be és haladnak át, állomásról állomásra a rendszeren.

Ilyen rendszerek néhány alapvető törvényszerűségét P. M. MORSE [29, 34. o.] leírja; meg kell azonban jegyeznünk, hogy e rendszerek analitikus tárgyalása nehéz.

Az alábbiakban néhány speciálisan megválasztott kiszolgálóhelyekkel rendelkező tömegkiszolgáló rendszert szimulációs módszerrel tárgyalunk. A tárgyalt rendszerek specialitása egyrészt m megválasztásától, másrészt a sorbanállási rendszer definíciójától függ. (A továbbiakban feltételezzük, hogy minden kiszolgálóhelyen egyszerre csak egy egységet szolgálnak ki.) Három speciális szekvenciális sorbanállási rendszert P. M. MORSE [29, 35. o.] közöl. E rendszerek abban különböznek, hogy mi történik a kiszolgálandó egységgel, miután a k-adik állomáson éppen kiszolgálták.

Ha a $(k+1)$-edik állomás már befejezte az előző egység kiszolgálását és szabad, akkor — a rendszer definíciójától függően — fogadhatja (vagy sem) a k-adik állomáson már kiszolgált egységet. Tegyük fel azonban, hogy a $(k+1)$-edik állomás foglalt. Ekkor a k-adik állomáson már kiszolgált egység

(1) elhagyhatja a k-adik állomást és a $(k+1)$-edik állomás előtt várakozhat addig, amíg az fogadni nem tudja; vagy
(2) továbbra is a k-adik állomáson marad, blokkolva az állomást, hogy a következő egységet fogadja, ezért ez az állomás egészen addig nem folytathatja a kiszolgálást — még akkor sem, ha egy másik egység a $(k-1)$-edik állomáson várakozik —, amíg a $(k+1)$-edik állomás fel nem szabadul, hogy a várakozó egységet fogadja és így feloldja a k-adik állomást is; vagy

(3) minden egységnek, amely az m kiszolgálóhely valamelyikén tartózkodik, ott kell maradnia addig, amíg az összes állomáson ki nem szolgálják, ezután az összes egység egyszerre lép tovább, mindegyik a következő kiszolgálóhelyre, az m-edik állomáson levő elhagyja a rendszert, egy egység pedig a várakozó sorból belép az első állomásra (ha a várakozó sorban egyetlen egység sincs, akkor mindegyikük addig várakozik, amíg egy újabb egység nem érkezik a rendszerbe, ennek beérkezése után azonban minden egység a következő kiszolgálóhelyre kerül).

Az (1) esetre T. H. NAYLOR adott egy megoldást [31, 141—151. o.]. Az említett problémát a (2) bekezdésben vázolt feltételekkel két speciális esetre (1. és 2. modell), majd ezenkívül a (3)-ban megadott feltételek mellett (3. modell) is megoldjuk szimuláció felhasználásával.

A (2) esetben feltételezzük, hogy az összes állomás kiszolgálási idejének valószínűségeloszlása exponenciális, μ átlagos kiszolgálási rátával, továbbá, hogy a beérkezések *Poisson-eloszlást* követnek λ átlagos beérkezési rátával.

2. Két kiszolgálóhellyel rendelkező sorbanállási rendszer, amelyben egyik kiszolgálóhely előtt sem állhat várakozó sor (1. modell)

A modell

Első alkalmazásként olyan két-kiszolgálóhelyes sorbanállási rendszert szimulálunk, amelyben sem az első állomás előtt, sem a két állomás között nem állhat várakozó sor. (2) típusú rendszereknek egy partikuláris állapotát definiálhatjuk, ha a két kiszolgálóhely lehetséges állapotaira vonatkozóan kikötéseket teszünk. A 2. kiszolgálóhely lehet üres (0) vagy foglalt (1), az 1. kiszolgálóhely pedig üres (0) vagy foglalt és éppen kiszolgál (1) vagy blokkolt (b) (már befejezte ugyan az ott levő egység kiszolgálását, de a 2. kiszolgálóhely foglalt).

Az előbbi feltételek szerint egy adott időpillanatban a (0, 0), (1, 0), (0, 1), (1, 1) és (b, 1) állapotok valamelyike következik be (ahol a két számjegy az első, illetve második kiszolgálóhely állapotára utal). Ha a valószínűségeket a két kiszolgálóhely lehetséges állapotainak megfelelő indexszekkel látjuk el, akkor a rendszer a következőképpen írható le analitikusan [21, 381—382. o.]:

$$P_{00} = \frac{2}{F}, \quad P_{01} = \frac{2\Psi}{F}, \quad P_{10} = \frac{\Psi^2+2\Psi}{F}, \quad P_{11} = P_{b1} = \frac{\Psi^2}{F},$$

ahol $\Psi = \lambda/\mu$ és $F = 3\Psi^2+4\Psi+2$.
Az egységek várható száma a rendszerben

$$E(k) = \frac{5\Psi^2+4\Psi}{F} = \begin{cases} 2\Psi - \frac{3}{2}\Psi^2 & (\Psi \ll 1) \\ 1 & (\Psi = 1) \\ \frac{5}{3} - \frac{32}{45\Psi} & (\Psi \gg 1) \end{cases}$$

A foglalt kiszolgálóhelyek várható száma:

$$E(SB) = \frac{4\Psi^2+4\Psi}{F} = \begin{cases} 2\Psi - 2\Psi^2 & (\Psi \ll 1) \\ \frac{8}{9} & (\Psi = 1) \\ \frac{4}{3} - \frac{4}{9\Psi} & (\Psi \gg 1) \end{cases}$$

Annak valószínűsége, hogy ügyfél érkezik az 1. kiszolgálóhelyre, de kiszolgálására nincs lehetőség:

$$P_R = \frac{3\Psi^2+2\Psi}{F} = \begin{cases} \Psi - \frac{1}{2}\Psi^2 & (\Psi \ll 1) \\ \frac{5}{9} & (\Psi = 1) \\ 1 - \frac{2}{3\Psi} & (\Psi \gg 1) \end{cases}$$

A szimuláció folyamán a (0, 0), (1, 0), (0, 1), (1, 1), (b, 1) állapotoknak megfelelő periódusokat T_{00}, T_{10}, T_{01}, T_{11} és T_{b1}-gyel jelöljük. Nyilvánvalóan $T_{00}+T_{10}+T_{01}+T_{11}+T_{b1}=TATS$, ahol $TATS$ a szimuláció teljes időtartama. A szimulált valószínűségek az alábbi egyenletek alapján számíthatók:

$$P_{ik}^{(s)} = \frac{T_{ik}}{TATS} \quad (i, k = 0, 1)$$

és

$$P_{b1}^{(s)} = \frac{T_{b1}}{TATS}.$$

A többi megfelelő statisztika pedig

$$E(k)^{(s)} = P_{01}^{(s)} + P_{10}^{(s)} + 2(P_{11}^{(s)} + P_{b_1}^{(s)})$$

$$E(SB)^{(s)} = P_{01}^{(s)} + P_{10}^{(s)} + P_{b1}^{(s)} + 2P_{11}^{(s)}$$

$$P_R^{(s)} = P_{10}^{(s)} + P_{b1}^{(s)} + P_{11}^{(s)},$$

ahol a felső (s) indexek a megfelelő statisztikák szimulált értékeire utalnak.

Átmenet-állapotok

A modellben az alábbi változókat használjuk:

RAR: két soron következő beérkezés közötti véletlen időintervallum,
RSR_h: a kiszolgálás véletlen időtartama a h-adik kiszolgálóhelyen ($h=1, 2$),
TP: a következő beérkezés időpontja,
TS_h: valamely egység kiszolgálásának befejezési időpontja a h-adik kiszolgálóhelyen ($h=1, 2$),
i: a rendszerben éppen jelenlevő egységek száma,
L: a szimulációban részt vevő egységek száma.

Ha a fenti feltételek alapján definiált sorba kapcsolt kiszolgálóhelyekkel rendelkező sorbanállási rendszer szimulálása céljából folyamatábrát akarunk felrajzolni, akkor mindenekelőtt figyelembe kell vennünk a lehetséges átmeneteket a rendszer egyik állapotából a másikba. Ezek:

(0, 0) állapotból (1, 0) állapotba;
(1, 0)-ból (0, 1)-be;
(0, 1)-ből vagy (1, 1)-be, vagy (0, 0)-ba;
(1, 1)-ből vagy (1, 0)-ba, vagy (b, 1)-be és
(b, 1)-ből (0, 1)-be.

A fentieken kívül még további 5 átmenet történhet meg elméletileg 0 valószínűséggel, nevezetesen akkor, amikor legalább két véletlen esemény ugyanabban az időpillanatban következik be. Az említett átmenetek:

(1, 0) állapotból (1, 1) állapotba, ha $TP^{(k+1)}=TS_1^{(k)}$ (a k-adik egység kiszolgálásának befejezési és a $(k+1)$-edik egység beérkezési időpontja egybeesik az első kiszolgálóhelyen),

$$(0,1)\text{-ből} \quad (1,0)\text{-ba}, \quad \text{ha} \quad TP^{(k+1)} = TS_2^{(k)},$$

$$(1,1)\text{-ből} \quad (0,1)\text{-be}, \quad \text{ha} \quad TS_1^{(k+1)} = TS_2^{(k)},$$

$$(b,1)\text{-ből} \quad (1,1)\text{-be}, \quad \text{ha} \quad TP^{(k+1)} = TS_2^{(k-1)},$$

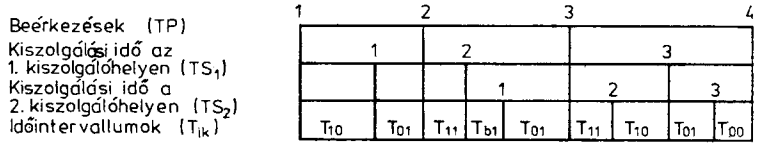
$$(1,1)\text{-ből} \quad (1,1)\text{-be}, \quad \text{ha} \quad TP^{(k+1)} = TS_1^{(k)}$$

és

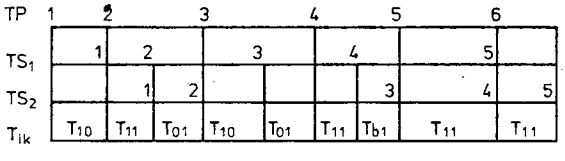
$$TS_1^{(k)} = TS_2^{(k-1)}.$$

Két olyan eset fordulhat elő a szimuláció folyamán, amikor egy új egység nem csatlakozhat a rendszerhez, mégpedig akkor, ha az első kiszolgálóhely (1) vagy (b) állapotban van (azaz amikor az első állomás foglalt és éppen kiszolgál vagy amikor blokkolt). Ha valamely egység beérkezésekor e két feltétel bármelyike fennáll, az egység kiszolgálását visszautasítja a rendszer.

Az összes lehetséges „szokásos" átmenet egy sorozatát az 1. ábrán, a „valószínűtlen" átmenetek egy sorozatát pedig a 2. ábrán illusztráljuk, ahol az első három egyenesen szereplő egész számok az egységek sorszámát jelzik.



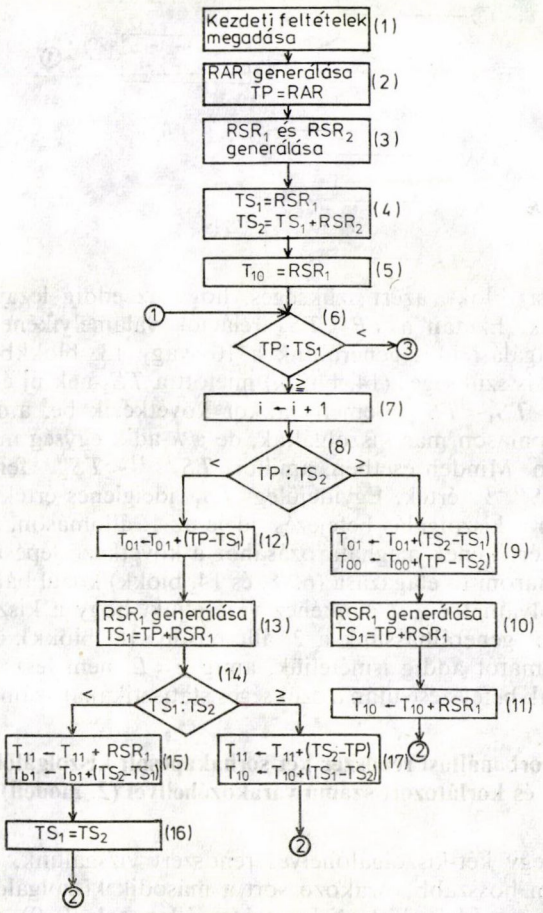
1. ábra



2. ábra

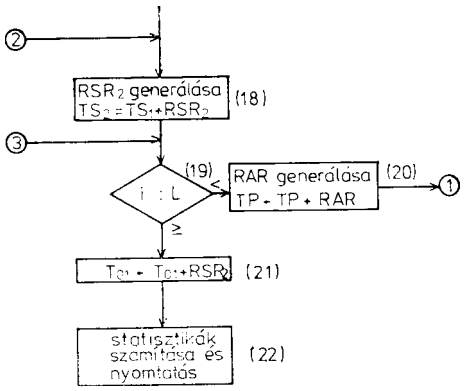
Feltételezzük, hogy amikor az első egység a rendszerbe lép a 0 időpontban: $i=0$, $T_{00}=0$, $T_{10}=0$, $T_{01}=0$, $T_{11}=0$ és $T_{b1}=0$.

Az 1. modell folyamatábrája



A 2. blokkban az 1. és a 2. egység beérkezése közötti időintervallumot generáljuk, majd ezt az értéket TP felveszi. Ezután mindkét állomás számára generálunk kiszolgálási időt és kijelöljük az első kiszolgálások befejező időpontját (TS_1, TS_2) a 3. és 4. blokkban. A szimuláció kezdetén a rendszer először (1, 0) állapotba kerül.

A 6. blokk segítségével döntjük el, hogy egy új érkező egység csatlakozhat-e a rendszerhez vagy sem. Ha $TP<TS_1$, akkor az érkező egység kiszolgálását a rendszer megtagadja, majd a 20. blokkban RAR-nek és TP-nek új értéket generálunk és mindaddig az 1 címkéhez vezéreljük a folyamatot, amíg $TP \geqq TS_1$ feltétel nem teljesül. A rendszer által „visszautasított" egységeket a továbbiakban nem vesszük számításba, a belépő egységek számát azonban mindig 1-gyel növeljük (7. blokk).



A 8. (döntési) blokk azért szükséges, hogy az eddig lezajlott időtartamokat összeszámolhassuk. Ezután a $TP \gtrless TS_2$ relációk valamelyikének bekövetkezésétől függően új kiszolgálási időt generálunk a 10. vagy 13. blokkban. Ha $TP < TS_2$, egy másik döntés is szükséges (14. blokk) mielőtt a TS_2-nek új értéket adunk. Nyilvánvaló, hogy a $TS_1 < TS_2$ esemény akkor következik be, amikor a $(k+1)$-edik egységet az 1. állomáson már kiszolgálták, de a k-adik egység még kiszolgálás alatt áll a 2. állomáson. Minden esetben, amikor $TS_1^{(k+1)} < TS_2^{(k)}$ fennáll, kiszámítandó a $T_{b1} = TS_2^{(k)} - TS_1^{(k+1)}$ érték. Egyidejűleg TS_1 ideiglenes értéket kap (16. blokk), amely ugyan nem a kiszolgálás befejezési ideje az 1. állomáson, de szükséges a T_{01} periódus korrekt értékének meghatározásához a következő lépésben.

A folyamat három fő elágazása (6., 8. és 14. blokk) közül bármelyiknek a végrehajtása után a folyamatot a 2 címkéhez vezéreljük, hogy a kiszolgálás számára új befejező időpontot generálhassunk a 2. állomáson (18. blokk).

A leírt folyamatot addig ismételjük, amíg $i = L$ nem lesz, majd a szimuláció utolsó szakaszának befejezése után a szükséges statisztikákat számítjuk ki (22. blokk)

3. Sorbanállási rendszer két sorbakapcsolt kiszolgálóhellyel és korlátozott számú várakozóhellyel (2. modell)

A modell

Most ismét egy két-kiszolgálóhelyes rendszert vizsgálunk, ahol megengedünk egy egységnél nem hosszabb várakozó sort a második kiszolgálóhely előtt, de nem lehet várakozó sor az első előtt. A lehetséges állapotok (0, 0), (1, 0), (0, 1), (1, 1), (0, 2), (1, 2) és $(b, 2)$, ahol a $(b, 2)$ állapot akkor következik be, ha a 2. állomás előtt sor várakozik, az 1. állomáson már befejezték a kiszolgálást, így az 1. állomás blokkolva van.

A modell analitikus leírása ismert [29, 37—38. o.]. A rendszert leíró mennyiségek közül a $P_{00}, P_{01}, P_{02}, P_{10}, P_{11}, P_{12}, P_{b2}$ valószínűségek a legfontosabbak, de megemlítünk még néhány további jellemző mennyiséget is, mint például a rendszerben levő egységek várható számát, a fogialt kiszolgálóhelyek várható számát, annak valószínűségét, hogy egység érkezik az 1. állomáshoz, de nincs lehetőség kiszolgálásra stb. A szimulált valószínűségek és statisztikák

$$P_{ik}^{(s)} = \frac{T_{ik}}{TATS} \quad (i, k = 0, 1) \quad \text{és} \quad P_{j_2}^{(s)} = \frac{T_{j2}}{TATS} \quad (j = 0, 1, b)$$

$$E(k)^{(s)} = P_{10}^{(s)} + P_{01}^{(s)} + 2(P_{11}^{(s)} + P_{02}^{(s)}) + 3(P_{12}^{(s)} + P_{b2}^{(s)})$$

$$E(SB)^{(s)} = P_{10}^{(s)} + P_{01}^{(s)} + P_{02}^{(s)} + P_{b2}^{(s)} + 2(P_{12}^{(s)} + P_{11}^{(s)})$$

$$P_R^{(s)} = P_{10}^{(s)} + P_{11}^{(s)} + P_{12}^{(s)} + P_{b2}^{(s)}.$$

Átmenet-állapotok

Ennél a modellnél az alábbi átmenetek lehetségesek:

(0, 0) állapotból (1, 0) állapotba, (1, 0)-ból (0, 1)-be, (0, 1)-ből vagy (1, 1)-be vagy (0, 0)-ba, (1, 1)-ből vagy (1, 0)-ba vagy (0, 2)-be, az (1, 2)-ből vagy (b, 2)-be vagy (1, 1)-be, a (0, 2)-ből vagy (1,2)-be vagy (0, 1)-be, (b, 2)-ből (0, 2)-be.

Ezenkívül még a következő átmenetek fordulhatnak elő nulla elméleti valószínűséggel: (1, 0)-ból (1, 1)-be, ha $TP^{(k+1)} = TS_1^{(k)}$ (a $(k+1)$-edik egység ugyanabban az időpillanatban érkezik a rendszerbe, amikor a k-adik egység kiszolgálása az 1. állomáson befejeződött és egyidejűleg elkezdődött kiszolgálása a 2. állomáson),

(0, 1)-ből (1, 0)-ba, ha $TP^{(k+1)} = TS_2^{(k)}$

(1, 1)-ből (1, 2)-be, ha $TP^{(k+1)} = TS_1^{(k)}$

és

$$TS_1^{(k)} < TS_2^{(k-1)}$$

(1, 1)-ből (1, 1)-be, ha $TP^{(k+1)} = TS_1^{(k)}$

és

$$TS_1^{(k)} = TS_2^{(k-1)}$$

(1, 1)-ből (0, 1)-be, ha $TS_1^{(k)} = TS_2^{(k-1)}$

(1, 2)-ből (0, 2)-be, ha $TS_1^{(k)} = TS_2^{(k-2)}$

(1, 2)-ből (1, 2)-be, ha $TP^{(k+1)} = TS_1^{(k)}$

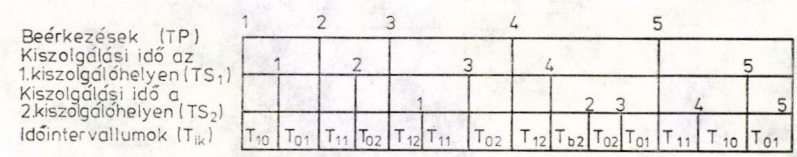
és

$$TS_1^{(k)} = TS_2^{(k-2)}$$

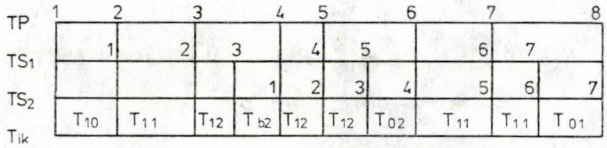
(0, 2)-ből (1, 1)-be, ha $TP^{(k+1)} = TS_2^{(k-1)}$

(b, 2)-ből (1, 2)-be, ha $TP^{(k+1)} = TS_2^{(k-2)}$.

Az összes lehetséges „szabályos", illetve „valószínűtlen" esetet magába foglaló átmeneti állapot egy-egy sorozatát a 3., illetve 4. ábrán illusztráljuk.

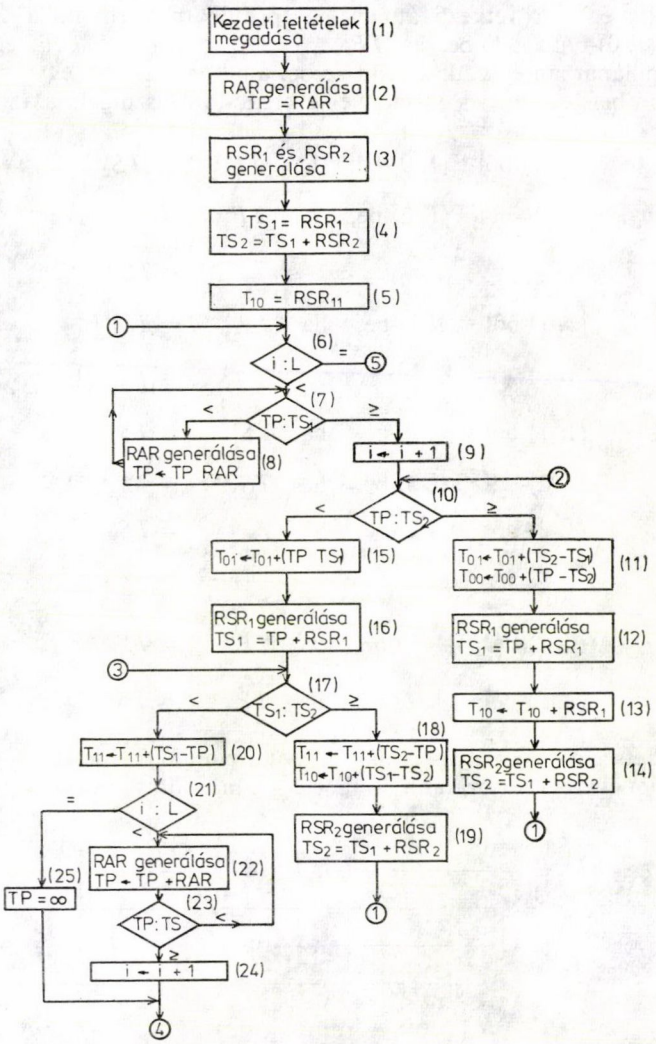


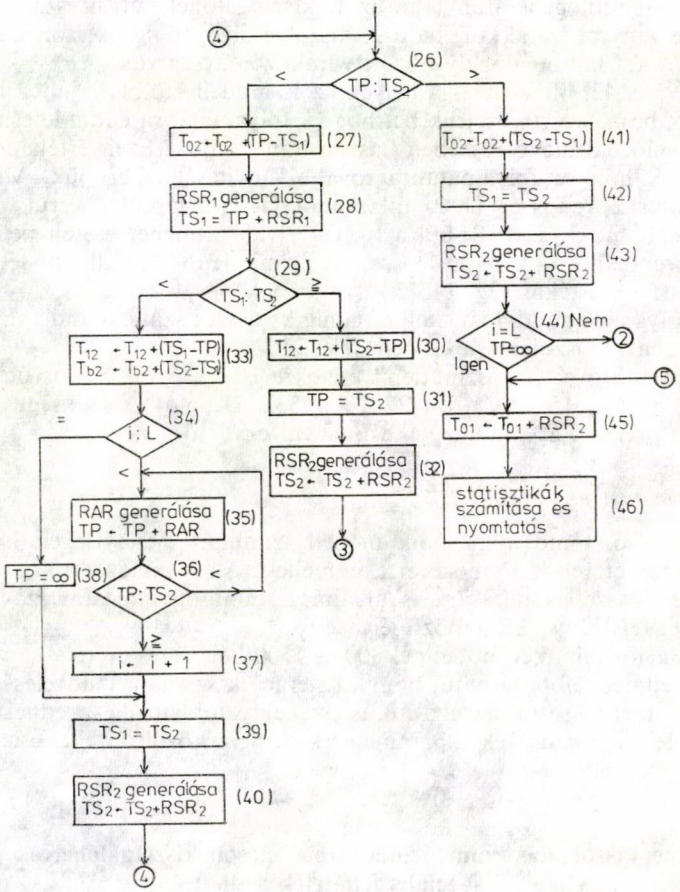
3. ábra



4. ábra

A 2. modell folyamatábrája





A 2. és 1. modell folyamatábrájának első része megegyezik attól a módosítástól eltekintve, hogy most a beérkezések közötti időintervallumokat külön generáljuk a 8. és 22. blokkban (később még a 35. blokkban is).

$TP = \infty$ jelöléssel utalunk arra (25. és 38. blokk), hogy TP-hez igen nagy értéket rendelünk, ami akkor következik be, ha $i = L$. Ezzel a folyamatot (a 26. blokknál) a szimuláció befejező részéhez vezéreljük (5 címke).

Azokban az időintervallumokban, amikor az (1, 2), (0, 2), (b, 2) állapotok nem fordulnak elő — az 1. modellhez hasonlóan — az 1.—25. blokkokat használjuk. A folyamatábra további részében az 1. modell kiterjesztése és ismétlése történik két új blokk (26. és 29.) bevezetésével, amelyek a 10. és 17. blokk döntéseihez hasonlóak. Ugyanakkor ezek a döntések különböznek a 10. és 17. blokk döntéseitől, hiszen most a rendszerben eggyel több egység van, ezért a számításokhoz szükséges T_{02} T_{12} és T_{b2} időtartamok csak akkor határozhatók meg, ha ezeket a döntéseket is figyelembe vesszük.

Kiszolgálási időt generálunk, ha az 1. kiszolgálóhely üres, és a rendszerbe új egység lép be vagy pedig akkor, ha a 2. kiszolgálóhelyen egy kiszolgálás éppen befejeződött és a két kiszolgálóhely között várakozó egység van.

A 31., 39. és 42. blokkot — amelyek az 1. modell 16. blokkjához hasonló szerepet töltenek be — azért vezettük be, hogy a folyamatábra eddig létesített elágazásait a szimuláció hátralevő részében hasznosítani tudjuk. Ezen értékadó utasítások teszik lehetővé, hogy a folyamatábra további bővítését elkerüljük. Minden olyan esetben, amikor a rendszer (1, 2) állapotból (1, 1) állapotba kerül TP felveszi TS_2 értékét (31. blokk) a (0, 2)-ből a (0, 1)-be való átmenet esetén pedig TS_1-hez TS_2 értékét rendeljük (42. blokk). Ezenkívül, minden (b, 2) állapot létrejötte után TS_1 felveszi TS_2 értékét (39. blokk).

A leírt folyamatot addig ismételjük, amíg az összes szimulálandó egység be nem lép a rendszerbe és kiszolgálása be nem fejeződik.

A folyamatábra a „valószínűtlen" esetekre is érvényes. Ekkor ugyanis a T_{ik} kiszámításához (11., 15., 18., 20., 27., 30., 33., 41. blokk) szükséges $|TP-TS_1|$, $|TP-TS_2|$ és $|TS_1-TS_2|$ mindegyike nullával egyenlő.

Numerikus eredmények

Mind az első, mind a második modell szimulált mennyiségeit kiszámítottuk a λ és μ paraméterek különböző értékei mellett. Az 1. táblázat a szimulált és analitikus értékek összehasonlítására is alkalmas, minthogy tartalmazza a megfelelő mennyiségeket $\Psi=\lambda/\mu$ különböző értékeire.

A szimulációt mindkét modellnél 300 és 1500 beérkezés esetére hajtottuk végre. A táblázat eredményeiből látható, hogy az iterációk számának növelésével a szimulációval nyert mennyiségek az analitikus értékekre jobban illeszkednek.

A megfelelő mennyiségek a paraméterek itt nem közölt értékei esetén is kivétel nélkül jól egybeesnek.

4. Tömegkiszolgálás szimulációja sorbakapcsolt kiszolgálóhelyek esetén speciális feltételek mellett

A modell

A továbbiakban feltételezzük, hogy minden egyes kiszolgálóhely saját kiszolgálási rátával rendelkezik, de sem a beérkezések közötti, sem a kiszolgálások időtartamának valószínűségeloszlását nem specifikáljuk.

E cikk bevezetőjének (3) szakaszában definiált szekvenciális sorbanállási modell megfogalmazásához a következő jelöléseket használjuk:

m: a kiszolgálóhelyek száma;
RAR_i: az $(i-1)$-edik és i-edik egység beérkezése közötti időintervallum $(i=2, \ldots, L)$;
RSR_{ih}: az i-edik egység kiszolgálási ideje a k-adik kiszolgálóhelyen, $(i=1, \ldots, L$ és $h=1, \ldots, m)$;
TS_{i0}: az i-edik egység rendszerbe lépésének időpontja, $(i=1, \ldots, L)$;
TS_{ih}: az i-edik egység kiszolgálásának befejező időpontja a h-adik kiszolgálóhelyen $(i=1, \ldots, L$ és $h=1, \ldots, m)$;

1. TÁBLÁZAT

		$\Psi=0{,}3$			$\Psi=0{,}5$			$\Psi=1{,}0$			$\Psi=4{,}5$		
		Szimulált		Analitikus	Szimulált		Analitikus	Szimulált		Analitikus	Szimulált		Analitikus
		$L=300$	$L=1500$		$L=300$	$L=1500$		$L=300$	$L=1500$		$L=300$	$L=1500$	
1. Modell	P_{00}	0,5377	0,5693	0,5764	0,3959	0,4153	0,4211	0,2057	0,2166	0,2222	0,0305	0,0264	0,0247
	P_{10}	0,2008	0,2004	0,1989	0,2433	0,2618	0,2632	0,3050	0,3300	0,3334	0,3113	0,3542	0,3622
	P_{01}	0,2003	0,1779	0,1728	0,2184	0,2114	0,2105	0,2347	0,2243	0,2222	0,1092	0,1142	0,1115
	P_{11}	0,0256	0,0267	0,0259	0,0692	0,0565	0,0526	0,1186	0,1172	0,1111	0,2585	0,2566	0,2508
	P_{b1}	0,0358	0,0256	0,0259	0,0726	0,0548	0,0526	0,1349	0,1118	0,1111	0,2891	0,2484	0,2508
	$E(k)$	0,52	0,48	0,48	0,75	0,70	0,68	1,05	1,01	1,00	1,52	1,48	1,48
	$E(SB)$	0,49	0,46	0,45	0,67	0,64	0,63	0,91	0,90	0,89	1,23	1,23	1,23
	P_R	0,2615	0,2527	0,2507	0,3850	0,3732	0,3684	0,5586	0,5590	0,5556	0,8589	0,8591	0,8638
2. Modell	P_{00}	0,5445	0,5680	0,5713	0,3726	0,4012	0,4092	0,1706	0,1933	0,2000	0,0122	0,0143	0,0149
	P_{10}	0,1896	0,1979	0,1989	0,2339	0,2602	0,2614	0,2885	0,3172	0,3200	0,2992	0,2669	0,2975
	P_{01}	0,1938	0,1743	0,1714	0,2160	0,2087	0,2045	0,1868	0,2019	0,2000	0,0699	0,0628	0,0670
	P_{11}	0,0345	0,0286	0,0275	0,0759	0,0599	0,0568	0,1359	0,1278	0,1200	0,2348	0,2345	0,2305
	P_{02}	0,0310	0,0252	0,0239	0,0676	0,0462	0,0455	0,1049	0,0864	0,0800	0,0716	0,0744	0,0709
	P_{12}	0,0026	0,0026	0,0036	0,0152	0,0104	0,0114	0,0563	0,0381	0,0400	0,1461	0,1699	0,1596
	P_{b2}	0,0036	0,0034	0,0036	0,0179	0,0132	0,0114	0,0559	0,0350	0,0400	0,1645	0,1769	0,1596
	$E(k)$	0,53	0,50	0,49	0,84	0,75	0,74	1,29	1,17	1,16	1,91	1,99	1,92
	$E(SB)$	0,49	0,46	0,46	0,72	0,67	0,66	1,02	0,97	0,96	1,37	1,39	1,38
	P_R	0,2302	0,2325	0,2335	0,3429	0,3437	0,3409	0,5366	0,5182	0,5200	0,8447	0,8482	0,8472

WT_{ih}: az az időtartam, ameddig az i-edik egységnek várakoznia kell, hogy beléphessen a h-adik kiszolgálóhelyre ($i=1, \ldots, L$ és $h=1, \ldots, m$);

IDT_{ih}: az az időtartam, amíg a h-adik kiszolgálóhely üres és várakozik az i-edik egység megérkezésére, ($i=1, \ldots, L$ és $h=1, \ldots, m$).

A megadott feltételek mellett egy új kiszolgálási sorozat kezdődhet rögtön azután, mihelyt az összes állomás kiszolgálta az éppen ott levő egységet, egyébként egyik állomáson sem kezdődik kiszolgálás, amíg a következő egység be nem lép a rendszerbe. Ebből következik, hogy várakozási és tétlen idő is előfordulhat ugyanazon a kiszolgálóhelyen. Várakozási idő akkor fordul elő, ha egy egység beérkezett a rendszerbe, de még nincs lehetőség kiszolgálására az 1. kiszolgálóhelyen vagy pedig akkor, ha valamely egységet a h-adik állomáson már kiszolgálták ugyan, de kiszolgálása a $(h+1)$-edik állomáson nem kezdődhet meg.

Tétlen idő viszont akkor következik be, ha a kiszolgálás befejeződött valamelyik állomáson, de a következő egység még nem érkezett meg vagy nem léphet be a kiszolgálóhelyre.

A teljes szimuláció három részre osztható:

(1) kezdeti szakasz: a 0 időpontból kezdődően addig tart, amíg megkezdődik az első egység kiszolgálása az m-edik kiszolgálóhelyen,
(2) a működés szakasza: az az időtartam, amely a kezdeti szakasz vége és az első állomáson az utolsó szimulált egység kiszolgálásának kezdési időpontja között eltelik,
(3) befejező szakasz: a szimuláció hátralevő része.

A „kezdeti" szakaszban egymást követő egységek esetén a várakozási és tétlen idők a következőképpen számíthatók:

$$\begin{aligned} WT_{i,1} &= TSMAX - TS_{i,0} \\ WT_{i-1,2} &= TSMAX - TS_{i-1,1} \\ &\vdots \\ WT_{1,i} &= TSMAX - TS_{1,i-1} \end{aligned}$$

és

$$\begin{aligned} IDT_{i,1} &= TSMAX - TS_{i-1,1} \\ IDT_{i-1,2} &= TSMAX - TS_{i-2,2} \\ &\vdots \\ IDT_{1,i} &= TSMAX - TS_{0,i}, \end{aligned}$$

ahol $i=2, \ldots, m$ és $TSMAX=$maximum $(TS_{i,0}, TS_{i-1,1}, \ldots, TS_{1,i-1})$ és a $TS_{0,2}, \ldots, TS_{0,m}$ mindegyike — amelyek az $IDT_{1,2}, \ldots, IDT_{1,m}$ kiszámításához szükségesek —, nullával egyenlő.

Nyilvánvaló, hogy

$$\begin{aligned} WT_{i-1,2} &= IDT_{i,1} \\ WT_{i-2,3} &= IDT_{i-1,2} \\ &\vdots \\ WT_{1,i} &= IDT_{2,i-1}. \end{aligned}$$

Az előbbieket összefoglalva írhatjuk

$$WT_{i-l,l+1} = TSMAX - TS_{i-l,l} \tag{4.1}$$

és

$$IDT_{i-l,l+1} = TSMAX - TS_{i-(l+1),l+1}, \tag{4.2}$$

ahol minden soron következő i-re ($i=2, \ldots, m$), $l=0, 1, \ldots, i-1$.

A „működés" szakasza és a „befejező" szakasz együtt tárgyalható.

A várakozási és tétlen idők kiszámításához szükséges formulák a fentiekkel majdnem megegyeznek, egy csekély módosítás bevezetésével ugyanis

$$WT_{i-l,l+n+1} = TSMAX - TS_{i-l,l+n} \tag{4.3}$$

és

$$IDT_{i-l,l+n+1} = TSMAX - TS_{i-(l+1),l+n+1}, \tag{4.4}$$

ahol minden soron következő i-re ($m<i\leqq L$), $l=0, 1, \ldots, m-n-1$. n értéke 0-tól m-ig változhat, és a „működés" szakaszában nyilván $n=0$. Látható, hogy WT_{ih} és IDT_{ih} értékeit ugyanúgy nyerjük, mint a „kezdeti" szakaszban, de most az összes kiszolgálóhelynél meg kell határozni ezeket a mennyiségeket (a „kezdeti" és „befejező" szakaszban azonban nem minden kiszolgálóhelynél). A „befejező" szakaszban (amikor $i=L$) $n>0$, ezért várakozási idő nem fordul elő azokon a kiszolgálóhelyeken, ahol minden kiszolgálás már befejeződött.

Egy új változó, k bevezetésével a (4.3) és (4.4) formulák a teljes szimulációra kiterjeszthetők és mindhárom szakaszra érvényes egységes formulák adhatók meg. Ezek:

$$WT_{i-l,k+l+1} = TSMAX - TS_{i-l,k+l} \tag{4.5}$$

és

$$IDT_{i-l,k+l+1} = TSMAX - TS_{i-(l+1),k+l+1}, \tag{4.6}$$

ahol minden soron következő i-re ($i=2, \ldots, m$) $l=0, 1, \ldots, m-n-1$, azzal a megszorítással, hogy a szimuláció különböző szakaszaiban n és k értékei csak a 2. táblázatban megadottak szerint változhatnak.

2. TÁBLÁZAT

Szakasz / Változó	Kezdeti	Működő	Befejező
n	$m-i$	0	1-től m-ig változik
k	0	0	n

A szimuláció folyamán n azoknak az állomásoknak a számát jelenti, ahol éppen nem folyik kiszolgálás, k pedig a kiszolgálásokat befejezett állomások számát adja.

A szimuláció folyamatának befejezése után kiszámítható a teljes várakozási és tétlen idő minden kiszolgálóhelynél. Ha ezeket WT_h-val és IDT_h-val jelöljük, akkor

$$WT_h = \sum_{i=1}^{L} WT_{ih}, \quad h = 1, \ldots, m$$

és

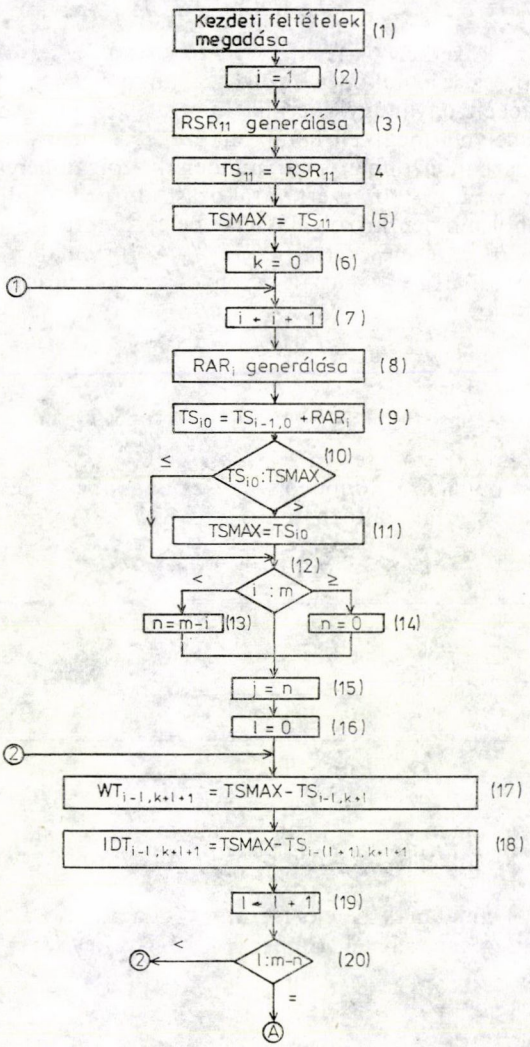
$$IDT_h = \sum_{i=0}^{L} IDT_{ih}, \quad h = 1, \ldots, m,$$

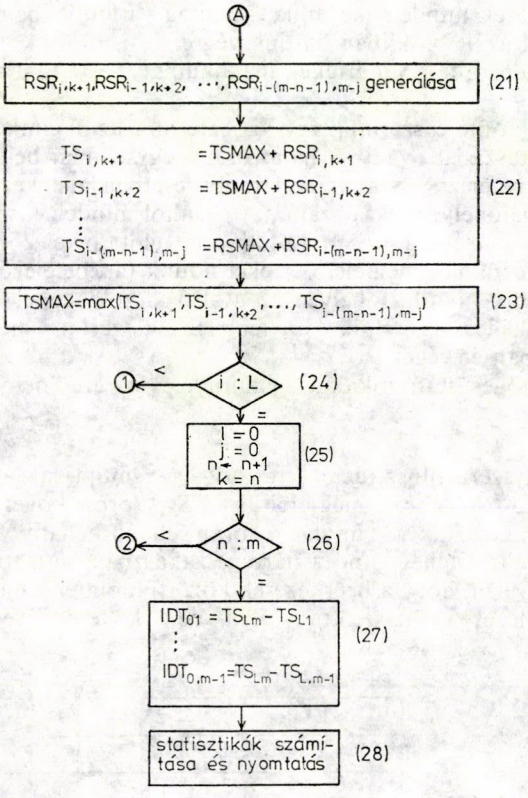
ahol IDT_{0h} akkor következik be, ha már az utolsó egységet is kiszolgálták a h-adik kiszolgálóhelyen és

$$IDT_{0h} = TS_{Lm} - TS_{Lh} \quad (h = 1, \ldots, m).$$

Nyilvánvaló, hogy $WT_{11}=0$ és $IDT_{11}=0$.

A 3. modell folyamatábrája





A szimuláció nulla időpontban indul, így $TS_{1,0}=0$. A kezdeti feltételeket az (1) blokkban adjuk meg, majd az első kiszolgálási idő generálása után értékét a TS_{11}-hez, illetve $TSMAX$-hoz rendeljük (3.—5. blokk). A „kezdeti" szakaszban és a „működés" szakaszában $k=0$ (6. blokk).

Minden beérkezés alkalmával i értékét 1-gyel megnöveljük, új RAR_i értéket generálunk és meghatározzuk az i-edik egység beérkezési időpontját (7.—9. blokkok). Ha $TS_{i0}>TSMAX$, akkor $TSMAX$ felveszi TS_{i0} értékét.

A 12. blokk segítségével dönthetjük el, hogy a szimulációs folyamat „kezdeti" szakaszban van-e ($i<m$) vagy sem. E döntéstől függően — a 2. táblázat alapján — adunk megfelelő értéket n-nek. E pillanatban $j=n$. A továbbiakban a j index kiszolgálási idők egy újabb sorozatának és befejezési időpontjainak meghatározásában játszik szerepet (21. és 22. blokk).

A 2 címke beiktatásával a 17.—20. blokkokban egy ciklust létesítettünk, hogy a (4.5) és (4.6) formula felhasználásával az aktuális várakozási és tétlen időket kiszámíthassuk. Ezt a ciklust addig ismételjük, amíg a várakozási és tétlen időket minden állomásnál meg nem határozzuk. A szimuláció során WT_{ih} és IDT_{ih} értékeit, WT_{11} és IDT_{11} kivételével, i és h összes lehetséges kombinációjára kiszámítjuk. $IDT_{12}, \ldots, IDT_{1m}$ meghatározásakor (18. blokk) a $TS_{02}, \ldots, TS_{0m}$ mennyiségek is

előfordulnak, de ezek mindegyike nulla, — ahogy a fentiekben már megemlítettük. (Ezt az értékadást az 1. blokkban hajtjuk végre.)

A legutóbb generált TS_{ih} értékek maximumát $TSMAX$-t a 23. blokkban határozzuk meg.

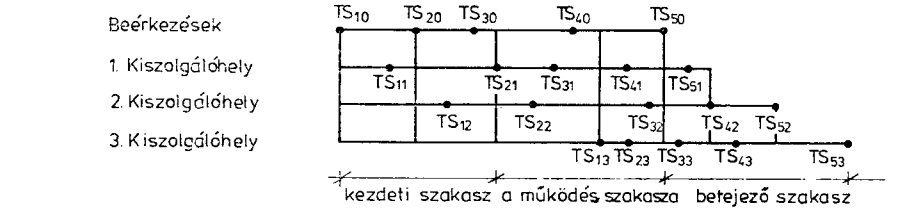
A 24. blokk döntése szerint vagy visszatérünk az 1 címkéhez (a „kezdeti" és a „működés" szakaszában) vagy — ha az összes egység már belépett a rendszerbe — végrehajtjuk a „befejező" szakaszt. Utóbbi esetben megkezdjük összeszámlálni azoknak a kiszolgálóhelyeknek a számát (n), ahol minden kiszolgálási tevékenység befejeződött vagy a jelenlegi $TSMAX$ időpont előtt befejeződik. A 25. blokkban az l, j és k változóknak megfelelő értéket adunk (a „befejező" szakaszban $j=0$).

Amíg az m-edik kiszolgálóhelyen az utolsó kiszolgálás el nem kezdődik, a folyamatot a 2 címkéhez vezéreljük és a fent leírtak szerint ismételjük. A szimuláció a TS_{Lm} időpontban ér véget.

Végül a szükséges statisztikákat számítjuk és az eredményeket nyomtatjuk ki.

Példa

Az 5. ábra egyszerű illusztráció arra, hogy megmutassuk a folyamat egy részét, ha $m=3$ és $L=5$. Az ábrán megjelöltük TS_{ih} soron következő értékeit. A továbbiakban meghatároztuk a szimuláció során fellépő néhány változó, a várakozási és tétlen idő aktuális értékét mind a három szakaszban — a folyamatábra alapján. Szükséges megjegyezni, hogy a beérkezések közötti időintervallumokat és a kiszolgálási időket tetszőlegesen választottuk, azok nem követnek semmiféle elméleti eloszlást.



5. ábra

Kezdeti szakasz ($k=0$):

$i = 2, \quad n = 1, \quad j = 1$

$TSMAX = TS_{20}$

$l = 0$:

$WT_{21} = TS_{20} - TS_{20}$

$IDT_{21} = TS_{20} - TS_{11}$

$l = 1$:

$WT_{12} = TS_{20} - TS_{11}$

$IDT_{12} = TS_{20} - TS_{02}$

$i = 3, \quad n = 0, \quad j = 0$

$TSMAX = TS_{21}$

$l = 0$:

$WT_{31} = TS_{21} - TS_{30}$

$IDT_{31} = TS_{21} - TS_{21}$

$l = 1$:

$WT_{22} = TS_{21} - TS_{21}$

$IDT_{22} = TS_{21} - TS_{12}$

$l = 2$:

$WT_{13} = TS_{21} - TS_{12}$

$IDT_{13} = TS_{21} - TS_{03}$

A működés szakasza ($k=0, n=0, j=0$):

$i = 4, \quad TSMAX = TS_{13}$

$l = 0$:

$WT_{41} = TS_{13} - TS_{40}$

$IDT_{41} = TS_{13} - TS_{31}$

$l = 1$:

$WT_{32} = TS_{13} - TS_{31}$

$IDT_{32} = TS_{13} - TS_{22}$

$l = 2$:

$WT_{23} = TS_{13} - TS_{22}$

$IDT_{23} = TS_{13} - TS_{13}$

$i = 5, \quad TSMAX = TS_{50}$

$l = 0$:

$WT_{51} = TS_{50} - TS_{50}$

$IDT_{51} = TS_{50} - TS_{41}$

$l = 1$:

$WT_{42} = TS_{50} - TS_{41}$

$IDT_{42} = TS_{50} - TS_{32}$

$l = 2$:

$WT_{33} = TS_{50} - TS_{32}$

$IDT_{33} = TS_{50} - TS_{23}$

Befejező szakasz ($i=5, j=0$):

$n = 1, \quad k = 1$

$TSMAX = TS_{42}$

$l = 0$:

$WT_{52} = TS_{42} - TS_{51}$

$IDT_{52} = TS_{42} - TS_{42}$

$l = 1$:

$WT_{43} = TS_{42} - TS_{42}$

$IDT_{43} = TS_{42} - TS_{33}$

$n = 2, \quad k = 2$

$TSMAX = TS_{52}$

$l = 0$:

$WT_{53} = TS_{52} - TS_{52}$

$IDT_{53} = TS_{52} - TS_{43}$.

5. Kiegészítések

Egyenletes eloszlású pszeudo-véletlen számok előállítására a *Multiplikatív Kongruencia Módszert* alkalmaztuk az

$$x_{i+1} \equiv a x_i \pmod{m} \tag{5.1}$$

kongruencia felhasználásával (ahol x_i, a, m és x_{i+1} pozitív egészek és m nagyobb mint x_i, a és x_{i+1}). Az (5.1)-ből nyert sorozatot az x_i/m transzformációval a [0, 1) intervallumra képeztük le. Maximális periódust (5.1)-ből akkor kapunk, ha

(1) $m=2^{\alpha}$, ahol α pozitív egész;
(2) x_0 (kezdő érték) m-hez relatív prím;
(3) a m-hez relatív prím;
(4) $a \equiv \pm 3 \pmod 8$ azaz $a=8t\pm 3$, ahol t pozitív egész;
(5) a-nak $2^{\alpha/2}$ közelében kell lennie. a ily módon választott értéke kielégíti a COVEYOU [8]—GREENBERGER [15] feltételt.

Ezen feltételek mellett elérhető maximális periódus $p=2^{\alpha-2}$ (HULL és DOBELL [19]).

Egyrészt a fenti meggondolások miatt, másrészt mivel e cikk megírásához használt programokat 24-bites szavakkal rendelkező számítógépen futtattuk, az (5.1) kongruencia alapján a kezdeti értékeket az alábbiak szerint választottuk:

$m=2^{23}$
$x_0=2^{23}-1$ (vagy bármely páratlan szám)
$a=2^{11}+3+8t$ (ahol t egész és $0\leqq t\leqq 64$).

m, x_0 és a ilyen értékei mellett a periódus hossza: $p=2^{21}$.

A fenti módszer, amelyet számos statisztikai teszttel is ellenőriztünk, mindig megbízható eredményeket szolgáltatott.

A módszer alkalmazásával létesített véletlenszám-generátorokat szubrutinként használva nyertünk sztochasztikus *(exponenciális* és *Poisson-eloszlású)* változókat „inverz transzformációs technika" [31] felhasználásával. Ha ugyanis egy statisztikai populáció $F(x)$ eloszlásfüggvénye explicit formában létezik, ezzel a technikával valószínűségi változók generálhatók a populációból, ha ismételten használjuk az $x_i=F^{-1}(u_i)$ relációt, ahol u_i egyenletes eloszlású pszeudo-véletlen számokat jelent a $[0, 1)$ intervallumon.

Exponenciális vagy *Poisson-eloszlás* esetén

$$x_i=-\left(\frac{1}{\lambda}\right)\log_e u_i$$

vagy

$$\prod_{i=0}^{x} u_i \geqq e^{-\lambda} > \prod_{i=0}^{x+1} u_i \quad (x>0, \text{ egész})$$

és így u_i minden értékéhez x_i és x egyetlen értéke tartozik, melyeknek sűrűségfüggvénye $\lambda e^{-\lambda x}$ vagy $\frac{\lambda^x}{x!}e^{-\lambda}$.

IRODALOM

[1] AHRENS, J. H. and DIETER, U., "Computer methods for sampling from gamma, beta, Poisson and binomial distributions", *Computing* **12** (1974) 223—246.

[2] ATKINSON, A. C. and PEARCE, M. C., "The computer generation of beta, gamma and normal random variables", *J. Roy. Stat. Soc. A.* **139** (1976) 431—461.

[3] ASHOUR, S. and JHA, R. D., "Numerical transient-state solutions of queueing systems", *Simulation* **21** No. 4., Oct. 1973.

[4] BLACK, J. R. and EVEN, J. C. JR., "Computer solutions of queueing models", *Simulation* **21** No. 4., Oct. 1973.

[5] BLAKE, K. a and GORDON, G., "Systems simulation with digital comuters", IBM Systems *Journal* **3** No. 1., 1964.

[6] CHORAFAS, D. N., *Systems and Simulation* (Academic Press, New York, 1965).

[7] CHU, K. and NAYLOR, T. H., "Two alternative methods for simulating waiting line models", *Journ. of Industrial Engineering* Nov.—Dec. 1965.

[8] COVEYOU, R. R., "Serial correlation in the generation of pseudorandom numbers", *Journ. of Assoc. for Comp. Mach.* **7** (1960).

[9] DIETER, U. "Pseudo-random numbers: The exact distribution of pairs," *Mathematics of Computation* **25** No. 116., Oct. 1971.

[10] DUTTON, J. M. and STARBUCK, W. H., *Computer Simulation of Human Behavior* (Wiley, 1971).

[11] DWASS, M., *Probability and Statistics* (W. A. Benjamin, Inc., New York, 1970).
[12] FELLER, WM., *An Introduction to Probability Theory and its Applications* (Wiley, 1957).
[13] GOOD, I. J., "The serial test for sampling numbers and other tests for randomness", *Proc. Camb. Phil. Soc.* **49** (1953) 276—284.
[14] GOOD, I. J., "On the serial test for random sequences", *Annals. Maths. Stat.* **28** (1957) 262—264.
[15] GREENBERGER, M., "Notes on a new pseudo-random number generator". *J. Assoc. Comp. Mach.* **8** (1961) 163—167.
[16] HALTON, J. H., "A retrospective and prospective survey of the Monte Carlo method", *SIAM Review* **12** (1970) 1—63.
[17] HAMMERSLEY, J. M. and HANDSCOMB, D. C., *Monte Carlo Methods* (J. Wiley, 1964).
[18] HULL, T. E. and DOBELL, A. R., "Random number generators", *SIAM Review* **4** (1962) 230—254.
[19] HULL, T. E. and DOBELL, A. R., "Mixed random number generators for binary machines", *J. Assoc. for Comp. Mach.* **11** (1964) 31—40.
[20] HUTCHINSON, D. W. "A new uniform pseudo-random number generator", *Communications of ACM* **9** (1966) 432—433.
[21] KAUFMANN, A., *Methods and Models of Operations Research* (Englewood Cliffs, N. J., Prentice-Hall, 1963).
[22] KENDALL, D. G. and HARDING, E. F., *Stochastic Analysis* (J. Wiley, 1973).
[23] LEE, A. M., *Applied Queueing Theory* (MacMillan Co. of Canada Ltd., St Martin's Press Inc., 1968).
[24] LEHMER, D. H., "Mathematical methods in large-scale computing units", *Annals. Computer Laboratory, Harward Univ.* XXVI (1951).
[25] MACLAREN, M. D. and MARSAGLIA, G., "Uniform random number generators", *Journal of the ACM* XII. No. 1., 1965.
[26] MAISEL, H. and GHUGNOLI, G., *Simulation of Discrete Stochastic Systems* (Chicago, Sci. Res. Assoc., 1972).
[27] MARTIN, F. F., *Computer Modelling and Simulation* (J. Wiley, 1968).
[28] MIZE, J. H. and COX, J. G., *Essentials of Simulation* (Prentice-Hall Intern. Inc., 1968).
[29] MORSE, P. M., *Queues, Inventories and Maintenance* (J. Wiley, 1958).
[30] NAYLOR, T. H., *Computer Simulation Experiments with Models of Economic Systems* (J. Wiley, 1971).
[31] NAYLOR, T. H., BÁLINTFY, J. L., BURDICK, D. S. and CHU, K., *Computer Simulation Techniques* (J. Wiley, 1966).
[32] PANICO, J. A., *Queueing Theory* (Prentice-Hall, Inc., 1969).
[33] PRÉKOPA, A., *Valószínűségelmélet* (Műszaki Könyvkiadó, Budapest, 1962).
[34] REITMAN, J., *Computer Simulation Applications* (J. Wiley, 1971).
[35] RUDOLPH, E. and HAWKINS, D. M., "Random number generators in cyclic queueing applications", *J. Stat. Comp. Simul.* **5** (1976) 65—71.
[36] SAATY, T. L., *Mathematical Methods of Operations Research* (McGraw-Hill Book Co., Inc., 1959).
[37] SAATY, T., *Elements of Queueing Theory* (McGraw-Hill Book Co., Inc., 1961).
[38] SASIENI, M., YASPAN, A. and FRIEDMAN, L., *Operations Research: Methods and Problems* (J. Wiley, 1959).
[39] SOWEY, E. R., "A chronological and classified bibliography on random number generation and testing", *Int. Stat. Rev.* **40** (1972) 355—371.
[40] SREJGYER, JU. A., *Monte Carlo-módszerek* (Műszaki Könyvkiadó, Budapest, 1965).
[41] TAHA, H. A., *Operations Research* (MacMillan Co. New York, 1971).
[42] TOCHER, K. D., *The Art of Simulation* (D. Van Nostrand Co. Inc. Princeton, N. J., 1967).

(Beérkezett: 1979. február 12.)

NÉMETH GYÖRGY
EGYETEMI SZÁMÍTÓKÖZPONT
1093 BUDAPEST IX., DIMITROV TÉR 8.

SIMULATION OF WAITING LINES IN SERIES

G. NÉMETH

Some basic rules for sequential service lines are known, however, the analytical discussion of the system in series is difficult. In this paper we confine ourselves to some particular cases for which solution will be given by stochastic simulation.

The problem is what happens to a unit after it has completed service by the k^{th} station in line. We suppose that the $(k+1)^{st}$ station is not free to accept the next unit. Then we distinguish 3 different cases (models).

In the first 2 models their analytical descriptions are given as well. In the 3rd model, assuming that each separate service station has its own service rate without specifying particular probability distributions for either the interarrival or the service times, the following conditions are given: all units in service have to stay in their respective stations until all stations complete service on the units in them, when all units move at once, each into the next station, the one in the m^{th} station leaving the line and one from the queue moving into the first station (if there is no unit in the queue, all units wait until another unit arrives, when they all move).

Alkalmazott Matematikai Lapok 5 (1979) 123—140

SZÁMÍTÓGÉPHÁLÓZATOK BIZONYOS ÜZENETIRÁNYÍTÁSI ELJÁRÁSAINAK EGYFAJTA MATEMATIKAI MODELLJE

MANIGÁTI CSABA
Budapest

A cikk célja az operációkutatás eszközei felhasználásának bemutatása egy viszonylag új, de már kiterjedt irodalommal rendelkező területen. A cikk három részre tagolódik: az első a számítógéphálózatok vázlatos jellemzéséről szól, a másodikban a statikus, elosztott struktúrájú S/F üzemmódú hálózatok egy speciális üzenetirányítási eljárásához felhasznált matematikai eszközök ismertetése található, végül a harmadik részben egy ismert — a matematikai programozás módszerein alapuló — determinisztikus üzenetirányítási modell bemutatására kerül sor.

1. Számítógéphálózatok üzenetirányítási eljárásai

Számítógéphálózatok jellemzése

Napjainkban folytatódik a számítógépek, a számítástechnika rohamos térhódítása az élet legkülönfélébb területein, így lehetővé, sőt szükségessé vált nemcsak az egyes számítógépek és részeik fejlesztése, hanem az — esetleg egymástól nagy távolságra levő — számítógépek összekapcsolása s így együttműködése is.

A számítógéphálózatok (Sz. H.-ok) napjainkban vitathatatlanul olyan érdeklődést keltenek, mint a *time-sharing* kb. 10 évvel ezelőtt. Embrionális állapotukban (kb. 1960—1970) főként egyetemek és kutatóintézetek „céltáblái" voltak, de kb. 1970 óta a kereskedelmi-üzleti világ nagyfokú érdeklődését is felkeltették (bankok, szállító vállalatok, repülőgéptársaságok, ...).

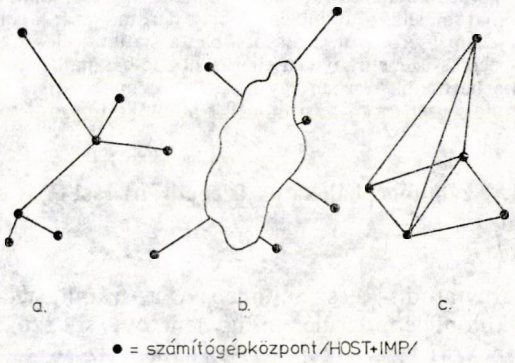
A továbbiakban számítógéphálózatoknak azokat a rendszereket nevezzük, melyek egymással kapcsolatban levő, egymással kölcsönösen kommunikáló és egymás erőforrásain osztozó független számítóközpontokból állnak (*resource sharing;* ahol az erőforrások *hardware* és *software* eszközök). Az egyes csomópontok dolgozhatnak tehát egyrészt helyi üzemmódban a saját operációs rendszerük irányítása, másrészt a hálózat elemeként egy magasabb szintű *supervisor program* vezérlése alatt. Természetesen a későbbiek során leírt modell számos más típusú, de feltételeinknek megfelelően működő hálózat (pl. különféle szállítási feladatok) esetében is használható. A számítóközpontok közötti üzenetváltás a kommunikációs hálózat révén történik, amelynek feladata az üzenetek hibamentes, gyors továbbítása az üzenetforrások és a rendeltetési helyek között az útképzési algoritmusnak megfelelő módon.

Számítógéphálózatokban az egyes számítóközpontok két fő részből: a csomóponti, ill. az erőforrás számítógépből állnak (továbbiakban: IMP, ill. HOST, ez utóbbi egy vagy több gépből álló rendszer). A csomóponti számítógépek az adatátviteli hálózat és az erőforrás számítógépek kapcsolatát vezérlik, az erőforrás szá-

mítógépek pedig a saját, ill. a hálózat más részéből jövő feladatokat oldják meg. Megjegyezzük, hogy az ebben a részben szereplő osztályozási, jellemzési szempontok főbb tájékozódási irányokként, fogódzkodóul szolgáltak csak, és — a téma rendkívüli összetettsége miatt is — nem törekedtünk különösebb (mérnöki) precizitásra, részletességre. Ilyen jellegű megközelítésre igényt tartók számára ajánlható pl. [18], [27], valamint a számítógéphálózatok témakörének egy egészen kitűnő bibliográfiája [26].

Csoportosítási lehetőségek

a) A számítógéphálózatok kommunikációs részét tekintve központosított (a), hurkolt (vagy gyűrűs) (b) és elosztott (c) struktúrát különböztetünk meg.



1. ábra

b) A felhasznált gépek szempontjából α) homogén: ha az erőforrás számítógépek azonosak vagy kompatibilisek, β) heterogén.

c) Az egymás közötti adatátviteli kapcsolat formája lehet:

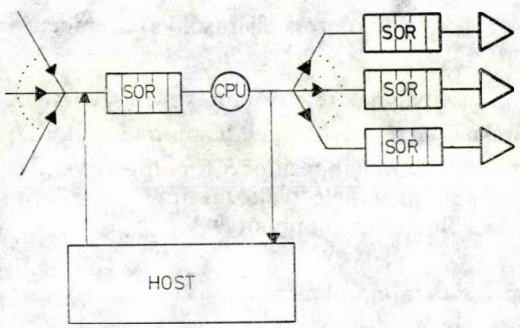
α) vonal kapcsolt *(circuit/line switching)*,

β) üzenet kapcsolt *(message switching)*,

γ) vegyes.

α) Telefonhívás jellegű módszer, a csomópontok a bemenő és kimenő vonalak között csak kapcsolást végeznek, s így a hívás után két csomópont akkor tud kommunikálni, ha a két számítógép között közvetlen fizikai kapcsolat létesül, s ez a „beszélgetés" ideje alatt is megmarad.

β) Ebben az esetben az üzenetek a feladási hely és a cél közötti állomásoknál a csomóponti számítógépben sorban állnak, majd pedig ha a megfelelő vonal szabad, az állomás továbbítja őket — az irányítási/útképzési algoritmusnak megfelelően — a következő csomópont felé, ilyenkor tehát a közbülső állomásokon tárolás-továbbítás *(Store-and-Forward)* folyamat megy végbe, így ezeket a hálózatokat a továbbiakban S/F hálózatoknak nevezzük. β) egy alesete, mikor a HOST üzenetet vevő IMP az üzenetet adott (állandó hosszúságú) ún. „csomagokra" osztja, és ezeket most már függetlenül kezelve továbbítja a célállomás felé, ahol újból üzenetté egyesítik *(packet-switching)*.



Csomóponti feldolgozás

2. ábra

Az üzenetek továbbításakor a sorbanállás az operatív memóriákban való tárolást jelenti *(core-switch)*. A számítógépes-kommunikációs hálózatok esetében a csomagkapcsolásos eljárás bizonyult a legjobbnak, a *Consultative Committee of International Telegraph and Telephon* X25 (1975. nov.) ajánlása β)-ra vonatkozik, így a továbbiakban ilyen típusú hálózatok matematikai úton megközelíthető problémáival foglalkozunk.

Útképzési technikák

A cikk csupán az elosztott struktúrájú S/F üzemmódú hálózatok üzenettovábbító szintjével foglalkozik. Mielőtt rátérnénk az útképzési eljárások ismertetésére, a későbbiek során használt terminológiákról adunk összefoglalót:

— üzenet: forrás és rendeltetési hely azonosítója, kibocsátási időpont, méret, esetleg prioritás által jellemzett információ-mennyiség,
— útképzési (irányítási) technika: az üzenetnek a forrástól a rendeltetési helyhez való továbbítási módjára vonatkozó előírás. (Ha N_i a forrás és N_j a célcsomópont, ezeket az üzeneteket a későbbiekben $i \to j$-vel jelöljük.)
— sor: kiszolgálásra várakozó üzenetsor,
— sorképzési elv: az üzenetsorok kezelésére vonatkozó előírás,
— üzenetkésés: az üzenet hálózatba kerülése, és a hálózatból való kilépése között eltelt idő. Három fő komponensből tevődik össze: a csomóponti lekezelésből, a várakozó sorokban eltöltött időkből, valamint az átviteli utakon való továbbítási időkből. T_i az i-edik csatornán, T a hálózatban töltött idő várható értéke,
— forgalom mátrix (γ, ill. γ_{ij}): (i, j) eleme megmutatja az $i \to j$ üzenetek másodpercenkénti átlagos számát,
— a hálózat topológiája/struktúrája: a számítógép-központok mint csúcsok, és a köztük levő kommunikációs összeköttetést biztosító vonalak, mint élek alapján kapott gráf.

Az S/F üzemmódú hálózatok tervezésének, üzemeltetésének egyik alapproblémája az üzenetek továbbításának/irányításának megszervezése, amely során lehetővé

válik a gyors kiszolgálás. Az útképzési eljárásokkal szemben az alábbi követelményeket támasztjuk:

— Az üzenetirányítási szabálynak biztosítani kell a gyors üzenettovábbítást,
— Az útképzési eljárásnak érzékenynek kell lenni a hálózat állapotváltozásaira:
  — a csomóponti és vonalhibából adódó topológiai változásokra,
  — a forrás—cél forgalmi terhelés változásaira,
  — a csomópontok telítettségi állapotaira.

Általános jellemzési szempontok:

— A hálózat állapotáról érkező információ (forgalmi terhelés és topológia), mint az idő függvénye:
  — idő invariáns (statikus),
  — változó (dinamikus).
— Az információ jellege:
  — az egész hálózatról kapjuk (globális),
  — helyi jellegű vagy nem teljes (lokális).
— Az útképzési feladat:
  — egy processzoré (centralizált),
  — egyidejű, a hálózat összes csomópontjában (elosztott).
— Az üzenet(ek) hálózatban eltöltött ideje:
  — az összes üzenetre vonatkozó idő minimalizálása (rendszer-optimalizálás),
  — az egyes üzenetekre vonatkozó idő minimalizálása (felhasználói optimum).

Az útképzési technikák két alapvető osztályba sorolhatók:

A) Determinisztikus irányítás: egy meghatározott és a továbbiakban változatlan döntési-irányítási szabályon alapszik, melynél az utak hurokmentesek, tehát az üzenet nem tér vissza az általa egyszer már érintett csomó pontba.

B) Adaptív technikák: figyelembe veszik a hálózat állapotváltozásait úgy, hogy vagy nem rendelkeznek információval a hálózat (aktuális) állapotáról (bizonyos meghatározott valószínűségek szerint továbbítják az üzeneteket a kimenő vonalakon), vagy megbecsülik a rendszer pillanatnyi állapotát. Ebben az esetben bizonyos rövid időperiódusokban zárt utas (hurkos) üzenetirányítás is előfordulhat.

Részletesebb osztályozás, valamint a különféle útképzési eljárások összehasonlítása található [17]-ben.

Egy lehetséges A típusú irányítási eljárás az ún. rögzített (fix) útképzési technika: ez az útképzési módszer egy megelőzően egyértelműen meghatározott — csak a kezdet—vég pártól függő — útvonalat ad meg. A fix irányítások megbízható vonalakkal és csomópontokkal dolgoznak. Ezek a technikák felhasználhatók az üzenetek hálózatban töltött átlagos idejének analitikus vizsgálatakor, valamint különféle tervezői optimumfeladatok megadásánál.

2. Matematikai eszközök statikus, eloszlott felépítésű S/F hálózatok vizsgálatánál

Számítógéphálózatok vizsgálatának három módszere terjedt el: az analízis (egzakt és közelítő eljárások), a mérés és a szimuláció (lásd pl. [2], [12—17], [21]). A továbbiakban problémáink vizsgálatára, ill. a kapott eredmények értékelésére a méréssel és szimulációval segített analízist hívjuk segítségül. Az S/F hálózat definíciójából látható, hogy egy ilyen rendszer lényegében sorbanállási modellek hálózataként kezelhető, így tehát nyilvánvaló, hogy a jellemzés egyik módja a tömegkiszolgáló rendszerek elméletének felhasználása lehet. Ilyen jellegű — a valószínűségszámítás segítségével történő — megközelítés található például [13]-ban, ill. [14]-ben.

A számítógép-hálózatok tervezésekor a számos lényeges probléma megoldása során a következő főbb szempontok lehetnek érdekesek (nem „erősorrendben"):

1. Átlagos üzenet-késés
2. Útképzési technika
3. Átlagos költség
4. A rendszer teljes költsége
5. A rendszer terhelhetősége
6. Sorképzési elv
7. A csomópontonkéni tároló kapacitás
8. A kommunikációs vonalak kapacitása
9. A rendszer struktúrája (topológiája)
10. Megbízhatósági követelmények.

A rendszer felhasználója számára például fontos információ, hogy milyen gyorsan (1), milyen megbízhatósággal (10), és mennyiért (3) szolgálják ki. Ezek a szempontok nyilvánvalóan igen szoros kapcsolatban vannak egymással. A matematikai tárgyalás egyszerűsítése céljából a továbbiakban néhány olyan feltevést teszünk, amelyek gyakorlati realitását meggyőző szimulációs eredmények jelzik.

A tömegkiszolgálási modellek felhasználásáról

Szemmel látható, hogy az S/F hálózatok analízisénél, valamint tervezésénél felhasznált különféle matematika diszciplinák (pl. gráfelmélet, matematikai programozás, megbízhatóságelmélet, ...) számára alapvető jelentőségű a sorbanállási modellek kezelése. A következőkben számítógépes-kommunikációs hálózatok jellemzése során felhasznált tömegkiszolgálási modellekről lesz szó.

A dolgozat további részében S/F hálózatok vizsgálatára súlyozott, véges, hurokmentes, többszörös éleket nem tartalmazó gráfot használunk:

$G=[N, A, C]$, ahol N a csúcsok, A az élek, C a hozzájuk tartozó kapacitások halmaza, n a csúcsok, m pedig az élek száma.

A rendszer felhasználója számára értékes információt nyújt az átlagos késés (T) ismerete. Ez egyetlen kommunikációs csatorna esetében a következő módon adható meg: vizsgáljuk meg ezt az esetet, tekintsünk egy „1" kiszolgálós rendszert. Ez KENDALL osztályozása szerint $A(\tau)/B(t)/1/K/m/z$-vel jellemezhető [1], ahol $A(\tau)$ a beérkezések között eltelt idő, $B(t)$ a kiszolgálási idő eloszlása, K a rendszer tárolókapacitása, m a források száma, z a sorképzési elv. Tegyük fel, hogy

$A(\tau)$, $B(t)$ sztochasztikus folyamatok függetlenek, valamint, hogy a csatornák és csomópontok tökéletesen megbízhatók.

A következőkben $K=\infty$, $m=\infty$, $z=$FIFO *(First In First Out)*. $K<\infty$ vizsgálata megtalálható pl. [28]-ban. Egy kiszolgálós rendszerre a következő eredmények ismertek:

M/M/1 (*Poisson*-érkezés, exp. kiszolgálás):
egyensúlyállapotban $T_i=\bar{t}/(1-\varrho_i)=1/(\mu C_i-\lambda_i)$,

ahol

λ_i az üzenetek beérkezésének várható értéke (üzenet/sec),

$\varrho_i=\lambda_i\bar{t}$ (<1) a csatorna kihasználtságára jellemző tényező, C_i az i-edik csatorna kapacitása (bit/sec), $1/\mu$ az üzenetek hosszának várható értéke (bit/üzenet), $\bar{t}$, $\bar{t}^2$ pedig $B(t)$ első és második momentuma.

M/G/1 esetén (*Poisson*-érkezés, $B(t)$ tetszőleges):

$$T_i=\lambda_i\bar{t}^2/(2(1-\varrho_i))+\bar{t}$$

a *Pollacsek—Hincsin-formulából* (Függelék 1. pont).

G/G/1 (általános érkezés, kiszolgálás) esetén nincs zárt formula, viszont KINGMAN igen jó felső korlátot adott $\varrho_i\to 1$ esetén [13]:

$$T_i\leqq\lambda_i(\sigma_i^2+\sigma_k^2)/(2(1-\varrho_i))+\bar{t},$$

ahol σ_i^2, ill. σ_k^2 $A(\tau)$, ill. $B(t)$ szórásnégyzete.

Sorbanállási *hálózat* esetén az átlagos késés megadása jóval bonyolultabb. Egzakt és közelítő technikák ismertek, a közelítő technikák lehetnek iteratív, diffúziós vagy izolációs eljárások [23].

Sztochasztikus rendszerek esetén alapcél lehet a hálózat felbontása, dekompozíciója könnyen jellemezhető egy kiszolgálós modellekre, természetesen figyelembe véve az eredeti struktúrát és forgalmat. M/M/1 esetén, ha $A(\tau)$ és $B(t)$ függetlenek, a felbontás elvégezhető, ui. JACKSON megmutatta [11], hogy egymástól független csomóponti rendszerek esetében egyensúlyi állapotban a kiszolgálás csak az adott csomópontbeli sortól függ, azaz a rendszer felbontható 1 kiszolgálós modellekre. Másik alapvető eredmény, ami a dekompozíciót lehetővé teszi BURKE nevéhez fűződik [4], aki megmutatta, hogy ha $A(\tau)$ és $B(t)$ exponenciálisak, akkor a távozások között eltelt idő eloszlása szintén exponenciális, s így konzerválható a csúcsok közötti forgalom *Poisson*-jellege (ui. egyébként más csúcsok befolyásolhatják egymás kiszolgálását, példa található erre „*tandem*" esetben [12]-ben).

Jelen esetben ez a felbontás azonnal nem tehető meg, hiszen az üzenetek hossza és a vonalsebesség nem változik, s így a különböző állomásokon a kiszolgálás időtartama ugyanarra az üzenetre ugyanakkora. Ha viszont az üzenetek hosszát valószínűségi változóként kezeljük — bár ez nem fedi a valós helyzetet — a szimulációs eredmények elég sokféle hálózati konfigurációra igen jó közelítést mutatnak; ennek az ún. függetlenségi feltevésnek a bevezetése KLEINROCK nevéhez fűződik [12].

Összefoglalva, modellünktől a következők teljesülését kívánjuk meg:

— $A(\tau)$, $B(t)$ exponenciális, stacionárius és független folyamatok,

— Az üzenetek hossza exponenciális eloszlású $A(\tau)$-tól független (függetlenségi feltevés),

— A különböző csomópontokba való érkezések egymástól függetlenek, hasonlóképpen a közbeeső állomások kiszolgálása is,
— Sorképzési elv: FIFO *(First In First Out)*,
— Útképzési technika: fix,
— Az üzenet 1 csomagból áll *(single packet;* az ún. *multi-packet* esetet lásd pl. [13]-ban).

Az általános jellemzési szempontok alapján elmondható: a vizsgált modellben a hálózat leírása statikus.

Dinamikus hálózati modellre példa [20].

2.1. Megjegyzés: A sorképzési elv nemcsak FIFO lehet, hanem helyettesíthető az ún. *work-conserving* kiszolgálási móddal [14], ilyen pl. a LIFO, *Random, Round Robin, Feedback, Processor Sharing* üzenetkezelési eljárás.

2.2. Megjegyzés: RUBIN modellje nem használja a függetlenségi feltevést, így a folyamok szerkezetére és a topológiára vonatkozó megszorításokat tartalmaz [24].

Legyen Z_{jk}: $=j \to K$ üzenetek késésének várható értéke,
γ_{jk}: $=j \to K$ üzenetek várható értéke (üzenet/sec) *Poisson*-érkezésnél,
$1/\mu$ az üzenet exponenciális eloszlású hosszának várható értéke.

Feltevéseinket felhasználva LITTLE tételével adódik az először KLEINROCK által [12] megadott T:

$$T \equiv \sum_{j,k} \frac{\gamma_{jk}}{\gamma} Z_{jk} = \sum_i \frac{\lambda_i}{\gamma} T_i \quad \text{(Függelék 2. pont)},$$

ahol T_i az átlagos késés az i-edik csatornán. Ha figyelembe vesszük még az üzenet i-edik csatornán való továbbításához szükséges időt (P_i) és feltételezzük, hogy a K feldolgozási idő minden csomópontra ugyanaz, akkor

$$T = K + \sum_i \frac{\lambda_i}{\gamma} [T_i + P_i + K].$$

Ha T_i-t felbontva felírjuk a sorbanállás és kiszolgálás idejét, kapjuk:

$$T = \sum_i \frac{\lambda_i}{\gamma} \left[\frac{1}{\mu' C_i} + \frac{\lambda_i/(\mu C_i)}{\mu C_i - \lambda_i} + P_i + K \right],$$

ahol $1/\mu'$ a HOST üzenetek, $1/\mu$ az összes üzenet (pl. a különféle visszajelzések) átlagos hossza. A T célfüggvény további vizsgálatának néhány ismert szempontja [15]-ben található.

2.3. Megjegyzés: A T-re legutóbb kapott eredményt az ARPANET modellezésen kívül a CAN*adian* UN*iversities computer* NET*work*-nél a gyakorlatban is felhasználták [9].

2.4. Megjegyzés: Látható, ha a folyam értéke a kapacitást közelíti, az egyes üzenetek késése rendkívül megnőhet, bár esetleg ez nem befolyásolja számottevően a T átlagot. Ennek elkerülésére javasolta MEISTER, MÜLLER, RUDIN a

$$T^{(k)} = \left[\Sigma \frac{\lambda_i}{\gamma} T_i^k \right]^{1/k}, \quad k > 0$$

célfüggvény alkalmazását ([19], lásd még 2.B.) feladatot), sőt figyelembe véve nemcsak az élek, hanem a csomópontok kapacitását, és azokkal is, mint M/M/1 rendszerrel számolva az előzőhöz hasonló típusú függvény adható meg.

A Multicommodity Network Flow problémaköre

A későbbiek során szó lesz egy matematikai modellről, a feladat a hálózati folyamok elméletéből ismert ún. *multicommodity*, azaz többtermékes folyam probléma:

Adott $G=[N, A, C]$ hálózat, valamint adottak bizonyos árucikkek (most: üzenetek), melyeket forrásuk és céljuk határoz meg, továbbá minden $i \to j$ árucikkre (üzenetre) bizonyos $r_{ij} \geqq 0$ igény áll fenn (most $r_{ij}=\gamma_{ij}$).

Feladat: megadni azt az irányítást, amelyet követve a szimultán árucikk (üzenet) folyamok a definiált célfüggvényt (pl. késés, költség) optimalizálják úgy, hogy esetleg kényszerfeltételeket is kielégítenek (pl. csatornakapacitás).

Az általános m.c.f. probléma a következőképpen írható le:

Adott a fenti hálózat és egy $R=[r_{ij}]$, $r_{ij} \geqq 0$ $n \times n$-es követelménymátrix, ekkor

$\min\limits_{\Phi} P(\Phi)$ v. $\max\limits_{\Phi} P(\Phi)$, ahol P a megadott célfüggvény, Φ pedig R-et kielégítő folyamkonfiguráció, azaz

$$\sum_{k=1}^{n} f_{kl}^{(i,j)} - \sum_{m=1}^{n} f_{lm}^{(i,j)} = \begin{cases} -r_{ij}, & l = i \\ +r_{ij}, & l = j \\ 0, & \text{különben} \end{cases}$$

a minden cikkre fennálló konzervációs törvény,

$$f_{kl}^{(i,j)} \geqq 0 \quad \text{minden} \quad i, j, k, l\text{-re}, \quad \text{ahol} \quad f_{kl}^{(i,j)}$$

az $i \to j$ cikk-folyam (k, l) élre eső mennyisége.

Φ-nek ki kell elégíteni bizonyos járulékos feltételeket (pl. csatornakapacitás és/vagy költségfeltétel). Ha ez nem áll fenn: az m.c.f. kényszermentes.

Optimum problémák, alapesetek

Az irányítási politikák két alapvető osztálya közül a továbbiakban determinisztikus, speciálisan bizonyos típusú fix irányítási eljárásokat fogunk felhasználni. Ezeket a technikákat könnyebb analizálni és a hálózattervezés fázisában általában ezek használhatók. Ennek indoklására: a determinisztikus és adaptív útképzési eljárások vizsgálata során kiderült, hogy az előzőekben különféle feltevések mellett analitikusan megadott idő-célfüggvény, ill. az adaptív esetben kapott szimulációs eredmények egyezése különféle valós hálózati terhelések mellett igen jó; hasonlóképpen az előbbi — néhány feltevéssel élő — determinisztikus, és a számos hálózati konfiguráció esetén szimulált modellre kapott eredmények egyezése kitűnő [13], [21]. Az általános jellemzési szempontok alapján látható, hogy az adaptív politika lokális, elosztott, felhasználói optimumot adó, az e részben leírt 2.B. feladatbeli fix útképzés pedig globális, centralizált és rendszeroptimalizáló. Ez utóbbi esetében érdekes lehet egy determinisztikus, de felhasználói optimumot adó eljárás megadása, és

a késésben mutatkozó eltérések összehasonlítása. A felhasználói optimumot adó politika ekvivalens a következő célfüggvényű rendszeroptimalizáló politikával [7]:

$$P(f) = \sum_{i=1}^{m} \int \frac{df_i}{C_i - f_i} = \ln \prod_{i=1}^{m} \frac{1}{C_i - f_i}.$$

Így a kétféle optimumot adó eljárás összehasonlítható, és GERLA megmutatta [7], hogy eltérésük kevesebb, mint 1%.

Az előzőek alapján ésszerűnek tűnik a hálózatanalízis és tervezés céljaira ezeket a determinisztikus modelleket felhasználni. A korábban felvetett szempontok vizsgálata során számos — a hálózati paraméterekkel kapcsolatos — optimum feladat fogalmazható meg. A továbbiakban célunk *egyetlen* paraméter kiválasztása és „viselkedésének" leírása a többi paraméter függvényében. Célszerűnek tűnik a hálózat analízise során elemzett T kiválasztása és a kérdés további vizsgálata, amelyhez a matematikai programozás eszközeit is felhasználjuk. A hálózattervezési probléma egy általános megfogalmazása a következő lehet:

Min T Feltéve:

topológia struktúra erőforrás elhelyezkedés útképzés folyamvezérlés csatornakapacitások tárolókapacitások csomóponti működési sebesség prioritás üzenethosszúság ⋮	költség (D), forgalmi (γ), megbízhatósági	kötöttségek

Bizonyos ésszerű feltételekkel a feladat „változói" a következőkre redukálódnak:
— vonalkapacitás,
— irányítási/útképzési technika,
— topológia.

A feladat — a drasztikus egyszerűsítések ellenére — még mindig igen bonyolult. Természetesen T bármelyik — a feltételek között szereplő — paraméterrel helyettesíthető (belátható, hogy a T-re és D-re vonatkozó feladatok egymás duálisai). A továbbiak során tegyük még fel, hogy a megbíthatóság a topológia függvénye, így végül a célfüggvény és a feltételek, valamint a különböző paraméterek megválasztásával három alapvető optimum feladatot kapunk:

2.A. Feladat: Kapacitásprobléma.
Adott γ, az irányítási technika, topológia, ekkor

a) min T, feltéve $D = \sum_{i=1}^{m} d_i(C_i) \leqq D_{\max}$, vagy

b) min D, feltéve $T \leqq T_{\max}$ úgy, hogy $f_i := \frac{\lambda_i}{\mu} \leqq C_i$.

A probléma megoldásának bonyolultsága a költség-kapacitás függvénytől függ.

Lineáris esetben $(d_i(C_i)=d_iC_i)$ pl. a $\sum_i \frac{\lambda_i}{\gamma} T_i$ függvényt minimalizáló optimális kapacitások a *Lagrange-multiplikátoros eljárással* kaphatók (Függelék 3. pont):

a) $$C_i = f_i + \frac{D_m}{d_i} \frac{(\lambda_i d_i)^{1/2}}{\sum_{j=1}^{m} (\lambda_j d_j)^{1/2}} \quad \text{és} \quad T = \frac{\bar{n}\left[\sum_{j=1}^{m} \sqrt{\lambda_j d_j/\lambda}\right]^2}{\mu D_m},$$

ahol $$D_m = D - \sum_{i=1}^{m} f_i d_i \quad \text{és} \quad \bar{n} = \frac{\lambda}{\gamma}, \quad \lambda = \sum_i \lambda_i,$$

b) $$C_i = f_i + \frac{\sum_{i=1}^{m} \sqrt{d_i f_i}}{T_{\max}} \quad \text{és} \quad D = \sum_{i=1}^{m} \left[d_i f_i + \frac{\left(\sum_{1}^{m} \sqrt{d_j f_j}\right)^2}{\gamma T_{\max}} \right].$$

Megjegyzés: D_m éppen a rendszer teljes költsége és a minimális kapacitásokhoz tartozó költség különbsége, s így nyilván $D_m \geqq 0$ kell legyen (ez ekvivalens az $f_i \leqq C_i$ és $\sum_{i=1}^{m} d_i C_i \leqq D$ feltételekkel). Konkáv költségfüggvénynél általában csak lokális minimum kapható. Egy lehetséges megoldás a függvény iteratív linearizálása, és az iterációs lépéseknél a linearizált feladat megoldása [7].

KLEINROCK megmutatta, hogy speciális, de az ARPANET-nél jól használható $d_i(C_i)=d_iC_i^{\alpha}$, $0 \leqq \alpha \leqq 1$ esetben ismét a *Lagrange-multiplikátoros módszerrel* az alábbi nemlineáris egyenlet szolgáltatja az optimális csatornakapacitást: $C_i - \lambda_i / /\mu - g_i C_i^{(1-\alpha)/2} = 0$, ahol $g_i = (\lambda_i/(\mu\gamma\beta\alpha d_i))^{1/2}$ és β a *Lagrange-multiplikátor*. Hasonlóan vizsgálható a $D = \sum_i d_i \ln \alpha C_i$ eset is [12].

Diszkrét kapacitások esetén (ami a reálisabb eset) a feladat egész típusú lesz, és dinamikus programozási módszerek használhatók, egy másik lehetséges eljárás a *Lagrange dekompozíciós módszer* felhasználása [7].

2.A'. Feladat: Ugyanaz, mint a 2.A. feladat *a)* része, csak az ottani T helyett a MEISTER, MÜLLER, RUDIN által javasolt $T^{(k)}$ célfüggvény felhasználásával. Lineáris költségfüggvény esetében adódik:

$$C_i^{(k)} = f_i + \frac{D_m}{d_i} \frac{(\lambda_i d_i^k)^{1/(k+1)}}{\sum_j (\lambda_j d_j^k)^{1/(k+1)}} \quad \text{és} \quad T^{(k)} = \frac{(\bar{n})^{1/k}}{\mu D_m} \left[\sum_i \left(\frac{\lambda_i d_i^k}{\lambda} \right)^{\frac{1}{1+k}} \right]^{\frac{1+k}{k}}.$$

2.B. Feladat: Útképzési/irányítási probléma. Adott γ, topológia, kapacitások, ekkor

$$\min_{\{f_i\}} T, \quad \text{feltéve} \quad f_i \leqq C_i.$$

Az optimális irányítási feladat, mint konvex célfüggvényes nem lineáris m.c.f. probléma adható meg. A hagyományos konvex programozási technikák lassúak

és sok számolást igényelnek. A feladat speciális szerkezetét felhasználó optimális és szuboptimális megoldások is ismertek [7], [13], [25].

2.C. Feladat: Kapacitás és útképzés probléma. Adott γ, topológia, ekkor

$$\operatorname*{Min}_{\{f_i, C_i\}} D, \quad \text{feltéve} \quad T \leqq T_{\max} \quad D = \sum_{i=1}^{m} d_i(C_i), \quad f_i \leqq C_i.$$

Lineáris költségfüggvény esetében a kapacitások a folyam függvényében zárt alakban megadhatók, így a probléma visszavezethető a folyamokra vonatkozó nem lineáris m.c.f. feladatra, azaz a 2.B. feladatra, ahol a 2.A. feladat b) részének az eredményét felhasználva $\min_{\{f_i\}} D(f)$ és mivel megmutatható, hogy $D(f)$ konkáv a megengedett m.c.f.-ek konvex poliéderén, bizonyos módosítással alkalmazhatók a 2.B. feladat megoldására szolgáló eljárások, de ebben az esetben csak lokális optimumot kapunk. Konkáv esetben a kapacitások nem adhatók meg zárt alakban a folyamok függvényében.

Diszkrét kapacitáshalmazra nincs egzakt eljárás, néhány közelítő technika ismert [7], [8].

2.D. Feladat: Vegyes feladat. Adott γ, ekkor

$$\operatorname*{Min}_{\{C_i, f_i, \text{topológia}\}} D, \quad \text{feltéve} \quad D = \sum_{i=1}^{m} d_i(C_i), \quad T \leqq T_{\max}, \quad f_i \leqq C_i,$$

topológiai korlátozások.

Nincs egzakt megoldás, csak heurisztikus eljárások ismertek.

Megjegyezzük, hogy csomagkapcsolt hálózatok topológiájának tervezésével először FRANK, FRISCH és CHOU foglalkoztak [6]-ban. Az általuk kifejlesztett *Branch X-Change* (BXC) eljárás valamilyen topológiai konfigurációból kiindulva bizonyos régi élek elhagyása, valamint új élek bekapcsolása során lokális minimumot ad. A módszer iteratív, és mindegyik iterációs lépésnél három fő lépésből áll:

a) Egy új él bekapcsolása, egy régi elhagyása úgy, hogy a rendszer kétszeresen összefüggő maradjon, azaz bármely két csúcs között legalább két lehetséges út legyen. Ez az ún. lokális transzformációs lépés.

b) Az új konfigurációhoz a 2.C. feladatban leírt éleket rendelve adjuk meg a folyamokat és kapacitásokat, majd számoljuk a költséget. Ha javulás mutatkozik az *a)*-beli transzformációt elfogadjuk, különben elvetjük.

c) Ha az összes lokális transzformációt megvizsgáltuk ÁLLJ, különben irány *a)*.

Az ún. *Concave Branch Elimination* (CBE) eljárás, amelyet GERLA és YAGED vizsgáltak, akkor alkalmazható, ha a diszkrét költségek konkáv függvénnyel közelíthetők. A CBE módszer kiindulásakor az összes lehetséges élet tartalmazó gráfot használja és feltéve, hogy D konkáv, lokális minimumot ad.

Megállapítható, hogy 20—30-nál több csomópont esetén mind a BXC, mind pedig a CBE eljárás igen időigényes, ezen túl a CBE-nek megvan az a hátránya, hogy bár hatásosan eliminálja a gazdaságtalan éleket, de nem kapcsol be új összeköttetéseket. A hátrányok elkerülésére BXC, CBE felhasználásával új eljárásokat fejlesztettek ki. Ezek egyike az ún. „vágástelítődés" (*cut saturation*, CS) módszer, amely a BXC kiterjesztésének tekinthető. Számos vizsgált eset kapcsán megállapítható, hogy CS legalább olyan eredményű, mint BXC, sőt a számolás hatásossága szempontjából lényegesen jobb annál [3].

3. Egy determinisztikus modell vázlatos ismertetése

Tekintsünk egy — a 2. pontban tett feltételeknek eleget tevő — S/F hálózatot, és vizsgáljuk meg ekkor a 2.B. feladatot ([7]).

A probléma megfogalmazása

Idézzük fel az átlagos késésre az előzőekben kapott — legegyszerűbb — eredményünket:

$$T=\frac{1}{\gamma}\sum_{i=1}^{m}\frac{f_i}{C_i-f_i},\quad f_i=\frac{\lambda_i}{\mu},\quad \gamma=\sum_{1}^{n}\sum_{1}^{n}\gamma_{ij},\quad \gamma_{ij}=r_{ij}. \tag{3.1}$$

Megjegyzés. A javasolt eljárás a valóságot jobban leíró célfüggvényre is alkalmazható.

Ekkor az irányítási probléma a következőképpen fogalmazható meg: $\min_{\{f\}} T(f)$, $f=(f_1, \ldots, f_m)$, ahol f_i az i-edik élen levő teljes folyam értéke, feltéve:

a) f az $R=[r_{ij}]$ $n\times n$-es mátrixot kielégítő m.c.f.

b)

$$f\leqq C,\quad \text{ahol}\quad C=(C_1, \ldots, C_m). \tag{3.2}$$

a)-t, ill. *b)*-t, valamint *a)*, *b)*-t kielégítő folyamok halmazát rendre F_a-val, ill. F_b-vel, valamint $F:=F_a\cap F_b$-vel jelöljük. Belátható: F_a konvex poliéder.

Minden $f\in F_a$ kifejezhető az F_a-beli extremális folyamok konvex kombinációjaként:

$$f=\sum_{i=1}^{r}\alpha_i\varphi^i,\quad \sum_{i=1}^{r}\alpha_i=1,\quad \alpha_i\geqq 0,\quad i=1, \ldots, r, \tag{3.3}$$

ahol $\{\varphi^i\}_1^r$ F_a extremális pontjainak halmaza. $F_b:=\{f|f\leqq C\}$ szintén konvex halmaz, így $F=F_a\cap F_b$ is az.

Megjegyezzük, hogy $\lim_{f_i\to C_i} T(f)=+\infty$, $i=1, \ldots, m$ s ez azt jelenti, hogy a célfüggvény belső büntetőfüggvényként tartalmazza a kapacitásfeltételt; ennek gyakorlati jelentősége az, hogyha már találtunk egy megengedett (kezdeti) f_0 folyamot, a szokásos nemlineáris optimalizáló technikák alkalmazása során automatikusan biztosított lesz a kapacitásfeltétel, azaz az m.c.f. feladat kényszermentessé válik.

A dekompozíciós közelítés

F zárt és konvex, a célfüggvény szigorúan konvex (konvexek összege), tehát ha F nem üres, pontosan egy lokális minimum van, ami egyben globális minimum is. A minimum meghatározása céljából a feladatot a *Dantzig—Wolfe-féle dekompozíciós eljárást* felhasználva átalakítjuk:

$$\min T=\min\frac{1}{\gamma}\sum_{i=1}^{m}\left[\sum_{k=1}^{r}\alpha_k\varphi_i^k\Big/\left(C_i-\sum_{k=1}^{r}\alpha_k\varphi_i^k\right)\right], \tag{3.4}$$

feltéve: $\sum_1^r\alpha_k=1$, $\alpha_k\geqq 0$ minden k esetén, $\alpha=(\alpha_1, \ldots, \alpha_r)$, ahol φ_i^k a φ^k extr. folyam i-edik élre eső része.

A fenti feltétel egyszerűbb, mint a (3.2)-beli, viszont sok a változó és elvileg ismerni kell az extremális pontokat.

A dekompozíciós eljárás hatékonysága a következő két tényen alapszik:

a) Minden f folyam legfeljebb $m+1$ extremális ponttal adható meg, így egyidejűleg legfeljebb $m+1$ db α_k-t kell figyelembe venni.

b) Létezik a φ^k-kat generáló eljárás.

a) bizonyítása egyszerű: (3.3) miatt minden f felírható az összes extremális pont konvex kombinációjaként:

$$\begin{aligned} &\varphi_1^1\alpha_1+\ldots+\varphi_1^r\alpha_r = f_1 \\ &\quad\vdots \qquad\qquad\qquad\quad \vdots \\ &\varphi_m^1\alpha_1+\ldots+\varphi_m^r\alpha_r = f_m \\ &1\cdot\alpha_1+\ldots+1\cdot\alpha_r = 1 \\ &\alpha_i \geqq 0 \quad \text{minden } i \text{ esetén,} \end{aligned} \tag{3.5}$$

ahol f_i az i-edik élen átmenő folyam.

Feltétel szerint tehát létezik megengedett α, s így megengedett bázismegoldás, azaz melynél legalább M komponens pozitív, ahol M a rendszer mátrixának rangja.

Tehát, mivel esetünkben $M \leqq m+1$, minden f legfeljebb $m+1$ extremális folyam konvex kombinációjaként adódik.

KÖVETKEZMÉNY: Ha $b > m+1$ extremális folyam segítségével írható fel, akkor a konvex kombináció $b-(m+1)$ komponense eliminálható; $b=m+2$ esetben ez az ún. *pivot lépés.*

Megjegyzés: Belátható, hogy már legfeljebb $(m-n+2)$ extremális folyam is elegendő.

A dekompozíciós közelítés optimum feltételei

A *Kuhn—Tucker-tétel* szerint α (3.4)-re vonatkozóan akkor és csak akkor optimális, ha

$$\sum_{k=1}^{r} \alpha_k = 1, \quad \alpha_k \geqq 0, \quad k = 1, \ldots, r \tag{3.6a}$$

$$\partial T/\partial\alpha_k = \beta_0, \quad \text{ha} \quad \alpha_k > 0, \quad \partial T/\partial\alpha_k \geqq \beta_0, \quad \text{ha} \quad \alpha_k = 0, \tag{3.6b}$$

ahol β_0 állandó (Függelék 4. pont). Legyen

$$\beta_k := \partial T/\partial\alpha_k = \sum_{i=1}^{m} l_i \varphi_i^k$$

$$l_i := \partial T/\partial f_i = \frac{C_i}{\gamma(C_i-f_i)^2}, \tag{3.7}$$

így β_k a φ^k folyam teljes költségének tekinthető, ha feltételezzük, hogy l_i az i-edik élen levő egységnyi folyam költsége (β_k a T késés növekedése α_k egységnyi növekedése esetén). (3.6b)-tiátírva:

$$\beta_k = \beta_0, \quad \alpha_k > 0 \quad \text{és} \quad \beta_k \geqq \beta_0, \quad \alpha_k = 0 \quad \text{esetben,} \tag{3.8}$$

azaz minden nem 0 extremális folyam azonos β_0 költségű, és minden extremális folyam, amelynek a költsége több, mint β_0 az 0 szinten van. (3.6a)-t és (3.8.)-at felhasználva:

$$\sum_{k=1}^{r} \alpha_k \beta_k = \min_k \beta_k \tag{3.9}$$

vagy (3.3) és (3.7) alapján

$$\sum_{i=1}^{m} l_i f_i = \min_k \left\{ \sum_{i=1}^{m} l_i \varphi_i^k \right\}. \tag{3.10}$$

Mivel pedig egy lineáris minimumköltséges folyamprobléma megoldása extremális folyam [10], így (3.10) segítségével:

$$\sum_{i=1}^{m} l_i f_i = \min_{v \in F_a} \sum_{i=1}^{m} l_i v_i, \tag{3.11}$$

s ez éppen azt jelenti: f pontosan akkor optimális, ha nincs olyan v, amelynél a késésnövekedés kisebb, mint f-é. Látható, hogy (3.11) az optimalitás tesztelése szempontjából jobb, mint (3.8), hiszen nem szükséges az összes extremális folyam ismerete, egy „legrövidebb út" számolás is elegendő.

A dekompozíciós közelítés alapproblémája

Tegyük fel, hogy adott már az extremális folyamok egy rendszere: φ^k, $k=1, \ldots, j<r$ és $\{\alpha_1, \ldots, \alpha_j\}$ részhalmaz szerint akarjuk (3.4)-et minimalizálni. Ez a probléma a projektív gradiens módszer [5], [7] segítségével oldható meg. A következő lépés annak eldöntése, hogy az alapprobléma optimális megoldása vajon nem globális optimum-e.

A dekompozíciós közelítés folyamgeneráló eljárása (alprobléma)

Tegyük fel, hogy az alapproblémát megoldottuk, azaz már van egy $\alpha=(\alpha_1, \ldots, \alpha_j)$ (3.8)-at kielégítő megoldásunk. Tekintsük a

$$\min_{v \in F_a} \sum_{i=1}^{m} l_i v_i \tag{3.12}$$

lin. min. költséges feladatot. Legyen φ' (3.12) megoldása és legyen $\beta' = \sum_{i=1}^{m} l_i \varphi_i'$. Ha $\beta' \geqq \beta_0 = \sum l_i f_i$: az α megoldás globális optimumot ad a (3.11) tulajdonság miatt.

Ha $\beta' < \beta_0$: α nem globális optimum. A következő lépés: φ'-t hozzávesszük az alapproblémához, s az így kapott új alapproblémát kezdjük megoldani.

Az „Extremális Folyam (E.F.) Módszer"

Az E.F. eljárás a dekompozíciós közelítés alapproblémájának és folyamgeneráló eljárásának ismételt felhasználásából áll, amíg a kívánt pontosságú T-t meg nem kapjuk. Az alprobléma minden iterációs lépésénél egy új extremális folyamot kapunk, ezek számát a sorozatos iterációs lépések folyamán a pivot operáció tartja $m+1$

alatt. Az algoritmus vázlata: Legyen b az éppen felhasznált extremális folyamok száma. Legyen $b=m+1$, jelölje most n az iterációs lépések számát és legyen $n=1$. Legyen $\Phi^0=\begin{pmatrix}\varphi^{1,0} & \dots & \varphi^{b,0}\\ 1 & \dots & 1\end{pmatrix}$ a kezdeti bázis, $[\Phi^0]^{-1}$ ennek inverze, $\alpha^0=(\alpha_1^0, \dots, \alpha_b^0)^T$ a kezdeti bázismegoldás, $f_0=\sum_{k=1}^{b}\varphi^{k,0}\alpha_k^0$ a kezdeti megengedett folyam.

Az E.F. eljárás alapproblémája

$$\operatorname*{Min}_{\{\varphi^{k,n-1},\,k=1,\dots,b\}} T.$$

Ezután legyen $\alpha^n=(\alpha_1^n, \dots, \alpha_b^n)$ az alapprobléma optimális megoldása, $f^n=\sum_{k=1}^{b}\varphi^{k,n-1}\alpha_k^n$ az alapprobléma optimális folyama,

$$l^n=(l_1^n, \dots, l_m^n)^T, \quad \text{ahol} \quad l_i=[\partial T/\partial f_i]_{f^n}$$

minden i esetén.

Az E.F. eljárás pivot lépése

Ha $b=m+1$: $\Phi^n \leftarrow \Phi^{n-1}$ és térjünk át az alprobléma megoldására, különben a pivot eljárással eliminálunk egy extremális folyamot. Ezután legyen $\alpha^n=(\alpha_1^n, \dots, \alpha_{m+1}^n)$ az új bázismegoldás, Φ^n az új bázis, $[\Phi^n]^{-1}$ ennek az inverze.

Az E.F. eljárás alproblémája

Adjuk meg az l^n metrikához tartozó φ^n („legrövidebb út") folyamot.

Az E.F. eljárás döntési problémája

Legyen $\beta_k^n=[l^n]^T\varphi^{k,n}$ a k-adik extremális folyam költsége, minden $k=1, \dots, m+1$; $\beta'=[l^n]^T\varphi'$ a legrövidebb folyam költsége. Ha $\beta' \geqq \min_k \beta_k^n$, akkor f^n optimális, különben legyen $b=m+2$, $\varphi^{m+2,n} \leftarrow \varphi'$, $\alpha_{m+2}^n=0$, $n=n+1$ és térjünk vissza az alapprobléma megoldására. Az algoritmus konvergenciatulajdonságai, valamint a kezdeti megengedett folyam előállítása [7]-ben található.

Az E.F. eljárás, amelyet itt csak Sz.H.-kra vizsgáltunk, különféle más hálózati folyamproblémák megoldásánál is jól használható (pl. szállítás, elosztás stb.).

Részletesebben: az E.F. módszer minden olyan minimum „költséges" folyamproblémára alkalmazható, ahol

1. T konvex.
2. T „költség" csak az éleken levő folyamtól függ.
3. $\partial T/\partial f_i$ folytonos és nem negatív.

Kis változtatással elhagyható (2), így az E.F. algoritmus s különböző vásárló/szállítóeszköz osztályra alkalmazható, ahol mindegyik osztály másként befolyásolja a teljes „költséget": $T=T(f^1, \dots, f^s)$, ahol f^i az i-edik osztályhoz tartozó folyam.

FÜGGELÉK

1. M/G/1 típusú sor esetén az átlagos várakozási időt a *Pollacsek—Hincsin-formula* adja meg [13]:

$$W = \lambda_i \bar{t}^2/(2(1-\varrho_i)).$$

Mivel T_i=átlagos kiszolgálási idő+átlagos várakozási idő$=\bar{t}+W$, így $T_i=\bar{t}+\lambda_i\bar{t}^2/(2(1-\varrho_i))$.

2. Jelölje Π_{jk} a $j\to k$ üzenetek által igénybevett élek halmazát, $i\in\Pi_{jk}$ ha $j\to k$ során az üzenetek áthaladtak az i-edik élen.

Nyilván:

$$\lambda_i = \sum_j \sum_k \gamma_{jk}, \quad Z_{jk} = \sum_{i\in\Pi_{jk}} T_i$$

így

$$T = \sum_{j=1}^{n}\sum_{k=1}^{n} \frac{\gamma_{j,k}}{\gamma} Z_{jk} = \sum_{j=1}^{n}\sum_{k=1}^{n} \frac{\gamma_{j,k}}{\gamma} \sum_{i\in\Pi_{jk}} T_i = \sum_{i=1}^{m} \frac{T_i}{\gamma} \sum_{\substack{j\\ i\in\Pi_{jk}}} \sum_k \gamma_{j,k} = \sum_{i=1}^{m} \frac{\lambda_i}{\gamma} T_i.$$

Ez az eredmény egyébként azonnal adódik LITTLE ismert tételéből [13]: $\bar{N}=\lambda T$, ahol $\bar{N}$ jelöli a rendszerben levő üzenetek/igények átlagos számát; ekkor ugyanis az i-edik csatornát használó üzenetek átlagos száma $\lambda_i T_i$, így az egész hálózatra vonatkozó γT nyilván ezek összege.

3. A feladat: Min T, $T=\sum_{i=1}^{m} \frac{\lambda_i}{\gamma} \frac{1}{\mu C_i-\lambda_i}$, feltéve $d_i(C_i)=d_i C_i$ és $D=\sum_i d_i C_i$. Felhasználva a célfüggvény konvexitását, a *Lagrange-módszer* [22] alkalmazásával:

$$L = T+\alpha\left(\sum_{i=1}^{m} C_i d_i - D\right).$$

A *Lagrange-függvényt* deriválva C_i szerint, és a $\partial L/\partial C_i=0$, $i=1, \ldots, m$ egyenletrendszert megoldva, kapjuk:

$$-\frac{\lambda_i}{\gamma}\frac{\mu}{(\mu C_i-\lambda_i)^2}+\alpha d_i = 0,$$

innen

$$C_i = \frac{\lambda_i}{\mu}+\frac{1}{\sqrt{\alpha\cdot\gamma}}\left(\frac{\lambda_i}{\mu d_i}\right)^{1/2},$$

$$\sum_{i=1}^{m} C_i d_i = D = \sum_{i=1}^{m} \frac{\lambda_i d_i}{\mu}+\frac{1}{\sqrt{\alpha\cdot\gamma}}\sum_{i=1}^{m}\left(\frac{\lambda_i}{\mu d_i}\right)^{1/2}, \quad \text{így} \quad \frac{1}{\sqrt{\alpha\cdot\gamma}}\text{-t}$$

kifejezve, és a

$$D_m := D-\sum_{i=1}^{m}\frac{\lambda_i d_i}{\mu}$$

jelöléssel adódik:

$$C_i = \frac{\lambda_i}{\mu}+\frac{D_m}{d_i}\frac{(\lambda_i d_i)^{1/2}}{\sum_{j=1}^{m}(\lambda_j d_j)^{1/2}},$$

majd pedig ezt T-be helyettesítve és felhasználva $\bar{n}=\frac{\lambda}{\gamma}$-t kapjuk:

$$T=\frac{\bar{n}\left[\sum_{j=1}^{m}\sqrt{\lambda_j d_j}\right]^2}{\mu D_m}.$$

Könnyen látható, hogy $\bar{n}$ éppen az átlagos úthossz. Legyen ui. n_{jk} a Π_{jk}-ban levő élek száma. Az átlagos úthosszt a következőképpen definiáljuk [12]:

$$\bar{n}=\sum_{j=1}^{n}\sum_{k=1}^{n}\frac{\gamma_{jk}}{\gamma}n_{jk}.$$

Ekkor:

$$\lambda=\sum_{i=1}^{m}\lambda_i=\sum_{j=1}^{n}\sum_{k=1}^{n}\gamma_{jk}n_{jk}$$

és innen $\bar{n}=\frac{\lambda}{\gamma}$, azaz $\bar{n}=n$.

4. A célfüggvény konvex és mivel ha a feltétel lineáris egyenlet(rendszer), akkor a *Kuhn—Tucker-regularitási feltétel* mindig teljesül, a *Kuhn—Tucker-tétel* ([22] 202. old.) direkt alkalmazásával:

$$\left[-\frac{\partial T}{\partial\alpha}+\beta_0(1,\ldots,1)^T\right]\cdot\alpha=0,\quad -\frac{\partial T}{\partial\alpha}+\beta_0\leqq 0'$$

$$\beta_0=\alpha_0^+-\alpha_0^-,\quad \alpha_0^+\geqq 0,\quad \alpha_0^-\geqq 0.$$

Koordinátákra bontva:

$$\left(-\frac{\partial T}{\partial\alpha_i}+\beta_0\right)\cdot\alpha_i=0\quad\text{és}\quad -\frac{\partial T}{\partial\alpha_i}+\beta_0\leqq 0$$

minden i-re, tehát $\partial T/\partial\alpha_i=\beta_0$, $\alpha_k\geqq 0$ esetén, és $\partial T/\partial\alpha_i\geqq\beta_0$, $\alpha_k=0$ esetén.

IRODALOM

[1] ALLEN, A. O., "Elements of queuing theory for system design" *IBM Syst. J.*, 2 (1975) 161—187.

[2] BENKŐ TIBORNÉ, TARNAY, K., ZÁRAY, É. és MANIGÁTI CS., „Számítógéprendszerek szimulálása CANDYS programcsomaggal." *Információ Elektronika*, 4 (1976) 295—302.

[3] BOORSTYN, R. R. and FRANK, H., "Large scale network topological optimization" *IEEE COM-25* (1977. jan.) 29—47.

[4] BURKE, P., "The output of a queuing system" *Oper. Res.* 4 (1956) 699—704.

[5] DEMYANOV, U. F. and RUBINOV, A. M., *Approximate Methods in Optimizaton Problems* (American Elsevier, 1970).

[6] FRANK, H., FRISCH, I. and CHOU, W., "Topological considerations in the design of ARPA network", *AFIPS SJCC* 36 (1970).

[7] GERLA, M., *The Design of Store-and-Forward (S/F) Networks for Computer Communications* (School of Engineering and Applied Science Report, University of California, Los Angeles, 1973. jan.).

[8] GERLA, M. and KLEINROCK, L., "On the topological design of distributed computer networks", *IEEE COM-25* (1977. jan.) 48—60.

[9] GUINDON, R. "Network design principles" *1st European Workshop: Computer Networks, Arles* (1973) 285—300.

[10] HU, T. C., *Integer Programming and Network Flows* (Addison—Wesley, Massachusetts, 1969).
[11] JACKSON, J., "Networks of waiting lines", *Oper. Res.* **5** (1957) 518—521.
[12] KLEINROCK, L., *Communication Nets: Stochastic Message Flow and Delay* (McGraw-Hill, New York, 1964).
[13] KLEINROCK, L., *Queuing Systems Vol.* **2**.: *Computer Applications* (John Wiley & Sons. Inc., New York, (1976).
[14] KOBAYASHI, H. KONHEIM, A. G., "Queing models for computer communications system analysis," *IEEE COM-***25** (1977. jan.) 2—28.
[15] MANIGÁTI, Cs., *Számítógépes hálózatok modellezésének néhány matematikai problémája* (7. szemináriumi füzet, KFKI, MSZKI, 1976).
[16] MANIGÁTI, Cs., "Operációkutatási eszközök a számítógéphálózatok matematikai modellezése során" előadás, *VII. Magyar Operációkutatási Konferencia, Pécs* (1977. okt. 11—14.).
[17] MANIGÁTI, Cs. és TALLÓCZY, I., „Számítógéphálózatok matematikai modellezésének kérdései", *Információ-Elektronika* **1** (1978) 52—58, és SZTAKI—SZÁMKI szeminárium (1978. ápr.).
[18] MARTIN, J., *Telecommunications and the Computer* (Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1976).
[19] MEISTER, B., MÜLLER, H. R. and RUDIN, H. R., JR., "New optimization criteria for message switching networks" *IEEE COM-***19** (1971. jun.) 256—260.
[20] MOLISZ, W., "Optimization of dynamic multicommodity flows in computer networks" *IFIP TC—6 COMNET '77 Budapest* (1977. okt. 3—7). Working Papers, I, 329—343.
[21] PRICE, W. L., "Simulation studies of flow control and routing in packet-switched computer networks" előadás, *5th Conference on the Theory of Operating Systems, Visegrád* (1979. febr.).
[22] PRÉKOPA, A., *Lineáris programozás I* (Bolyai János Matematikai Társulat, Budapest, 1968).
[23] PUJOLLE, G. and SOULA, C., "A study of flows in queueing networks and an approximate method for solution", *th International Symposium on Modelling and Performance Evaluation of Computer Systems, Vienna* (1979. febr.) Conf. Reprints vol. 2.
[24] RUBIN, I., "Communication networks: message path delays", *IEEE IT-***20** (1974. nov.) 738—745.
[25] Sz. TURCHÁNYI, P. *Optimalizálási feladatok csomagkapcsolt számítógéphálózatok tervezésénél* (MTA SZTAKI Tanulmányok, 77/1978).
[26] SZENTIVÁNYI, T. and TALLÓCZY, I., *Computer Networks (Bibliography)* (SZÁMKI Közlemények, **17** Budapest 1978).
[27] TARNAY, K., *Computer Networks* (report, Central Research Institut the for Phisics, Computer Department, Budapest, 1976).
[28] ZEIGLER, J. F., *Nodal Blocking in Large Networks* (School of Eng. and Appl. Science, Report, Univ. of California, Los Angeles, 1971. okt.).

(Beérkezett: 1978. november 14.)

MANIGÁTI CSABA
SZÁMÍTÓGÉPALKALMAZÁSI KUTATÓINTÉZET
1536 BUDAPEST, PF. 227.

A MATHEMATICAL MODEL FOR DESIGN OF COMPUTER NETWORKS

CS. MANIGÁTI

The purpose of this paper is to show some employment of operation research tools on a relatively new and an extremly complex area, namely the modelling, analysis and design of computer networks. This article consists of three parts. In the first one may find a short outlined characterization of computer-communication networks; the second part contains the mathematical tools for describing a special routing procedure of static, distributed S/F network.

At last in the third part a well-known deterministic routing policy is described, based on the theory of mathematical programming. In our case the optimal message routing can be formulated as a convex nonlinear multicommodity network flow problem.

Alkalmazott Matematikai Lapok 5 (1979) 141—156

AZ OPTIMÁLIS ELLENŐRZÉSI INTERVALLUMHOSSZRÓL

EROL GELENBE
Párizs

Jelen dolgozatban olyan tranzakciókat bonyolító (repülőgép helyfoglaló, file kezelő, adatbázis kezelő stb.) számítástechnikai rendszerek matematikai modellezésével foglalkozunk, amelyek időszakosan fellépő hibákkal működnek, és ellenőrzési pontok beiktatásával, valamint visszapergető és javító rendszer alkalmazásával őrzik meg az integritást. Az ellenőrzési intervallumhosszak eloszlását tetszőlegesnek feltételezve, kiszámítjuk a rendszer hozzáférhetőségét néhány tag feltétel mellett. Kimutatjuk, hogy a maximális hatékonyság akkor érhető el, ha a hasznos működési idő az egymást követő ellenőrzési pontok között determinisztikus, valamint kiszámítjuk ennek optimális értékét. Analizáljuk a rendszer kiszolgálásra váró tranzakcióinak sorbanállási folyamatát, szükséges és elégséges feltételeket adunk ennek ergodicitására. Kiszámítjuk a zérus sorhosszúság stacionárius valószínűségét. Ez elvezet a rendszer hatékonyságának finomabb analíziséhez, mely figyelembe veszi azt a tényt is, hogy egy meghibásodás kijavításához szükséges idő függvénye lehet azoknak az aktuális feladatoknak, amelyeket a kiszolgáló elvégzett az utolsó ellenőrzési pont óta, és így az optimális ellenőrzési intervallumhosszúság pontosabb képletét adjuk meg.

A modell és a kapott eredmények megbízhatóságelméleti összefüggéseikben is újak. Olyan sorbanállási folyamattal foglalkozunk ugyanis, ahol a hibák kijavítását megszakíthatja a karbantartás és a hibák kijavítása. Feltesszük, hogy egy meghibásodás kijavításának ideje függ attól az időtől, amely a meghibásodás bekövetkezése és az utolsó karbantartás között eltelt.

1. Bevezetés

A cikkben elemzett modellt általában véve a megbízhatóságelmélet keretein belül tárgyaljuk. Indítékunk erre az, hogy a modell megjelenik mind a megbízható adatbázis kezelő, mind a tranzakciókat bonyolító számítástechnikai rendszerek tanulmányozása során.

Tekintsünk egy adatbáziskezelő-rendszert (kiszolgáló), mely tranzakciókat bonyolít (fogyasztók) és a kiszolgálás FCFS (*first-come-first-served;* az elsőként jött az elsőként kiszolgált) rendszerben történik. Fontos, hogy képesek legyünk a központi memóriaegység tartalmának rekonstruálására hiba bekövetkezése és így egyes memóriatartalmak érvénytelenné válása esetén. E célból minden egyes tranzakciót felülvizsgáló sávban *(audit trail)* tárolunk. Időközönként ellenőrzőpontot létesítünk, és hiba észlelésekor a felülvizsgáló sávban tárolt, az utolsó ellenőrzés óta lebonyolított minden egyes tranzakciót ismételten végrehajtunk. Ellenőrzőpont létesítésekor a központi memóriaegység tartalmát — melyet esetleges hibák veszélyeztethetnek — átmásoljuk más memóriaegységekbe (lemez vagy mágnesszalag). Hiba észlelésekor a tranzakciók ismételt elvégzését megelőzően az ellenőrző memóriatartalmat visszaíratjuk a központi memóriaegységbe. GELENBE [8] dolgozatában ezt a rendszert sorbanállási problémaként modellezi. A kiszolgáló három lehetséges állapot egyikében lehet, és az állapotok egymás utáni bekövetkezései *Markov-*

láncot alkotnak. GELENBE [9] dolgozatában figyelembe veszi a több különböző típusú hiba és ellenőrzési pont esetét. YOUNG [15], KOVALENKO [10], CHANDY [2, 3] olyan, a hibákat utólagosan javító rendszereket mutatnak be, melyeknél nincs figyelembe véve a sorbanállás és a tranzakciók esetleges késleltetése. E szerzők állandó ellenőrzési intervallumhosszúságot vizsgálnak és megadják az optimális értéket, amelynél maximális a rendszer hatékonysága. CHANDY [2, 3] feltételezi, hogy a hibavalószínűség kicsi. ROBIN [13] az optimális ellenőrzés egy kérdését a jelen cikkben is tanulmányozott modell alapján vizsgálja.

Hasonló modell fellép megbízhatóságelméleti problémák tanulmányozása során is. Tekintsünk egy olyan kiszolgálóegységet, amely előre meghatározott időpontokban karbantartáson megy keresztül, és amely meghibásodásoknak van kitéve. Tegyük fel, hogy a karbantartás és a hibák kijavítása alatt nincs kiszolgálás. Feltételezzük, hogy a hibák kijavításának ideje függvénye lehet annak az időnek is, amely a meghibásodás és a legutolsó karbantartás között eltelt, valamint azt, hogy a fogyasztók kiszolgálása érkezési sorrendben történik. A dolgozatban pontosan ilyen feltételezésekkel élünk.

Problémánk megfogalmazását a 2-es és a 3-as fejezet tartalmazza. A modellben szereplő sorbanállási folyamatban a fogyasztók érkezése *Poisson-folyamatot* alkot, a kiszolgálási idők *exponenciális eloszlásúak*. A meghibásodás időpontjait *Poisson-folyamatnak* feltételezzük, amely független a fogyasztók érkezésétől, a kiszolgálástól és a karbantartástól. A javítási idő függvénye a meghibásodás bekövetkezése és a legutolsó karbantartás között eltelt időnek. A kiszolgálást megszakítja a karbantartás és a javítás.

2. Visszapergető, javító rendszer

A fogyasztóknak nyújtott szolgálat függ a kiszolgáló állapotától, X_t-től ($t \geqq 0$), amelyre:

$$X_t = \begin{cases} 2, & \text{ha a rendszer ellenőrzőpontot létesít,} \\ 1, & \text{ha meghibásodás javítása történik,} \\ 0, & \text{ha szabályosan működik.} \end{cases} \tag{2.1}$$

Tranzakciók végrehajtása csak a 0 állapotban történik. $\{X_t, t \geqq 0\}$ sztochasztikus folyamat a következő tulajdonságokkal:

I. Két egymás utáni 2-es állapotba való átmenet között a 0 állapotban töltött teljes idő olyan $F(y)$ eloszlás- és $f(y)$ sűrűségfüggvényű valószínűségi változó, mely független a folyamat múltjától. Ez alatt az idő alatt érhető el a kiszolgáló a sorban álló fogyasztók számára. Legyen:

$$E(Y) = \int_0^\infty y dF(y) < \infty.$$

II. A 2-es állapotban a folyamat véletlenszerűnek tekintett ideig marad a belépés után, feltesszük, hogy ez az idő független a folyamat múltjától, eloszlásfüggvénye: $C(y)$. Ennyi idő szükséges az ellenőrzőpont létesítéséhez, majd a folyamat visszatér a 0 állapotba. Feltesszük, hogy:

$$E_c = \int_0^\infty y dC(y) < \infty.$$

III. A 0 állapotból az 1-es állapotba való átmenet γ paraméterű *Poisson-folyamat,* ahol γ a kiszolgálás meghibásodásának gyakoriságát leíró paraméter.

IV. Az 1-es állapotba való átmenet után az ott tartózkodás idejét a következőképpen definiáljuk: legyen $h: R^+ \to R^+$ mérhető függvény. A t időpontban bekövetkező meghibásodásra (0-ból 1-be való átmenetre) legyen:

$$t' = \sup\{\sigma:\ \sigma < t,\ \chi_2(\sigma) = 1\}^1.$$

[1] $\chi_i(t)$ az $X(t)$ folyamat i állapotban való tartózkodásának az indikátorfüggvénye, azaz $\chi_i(t) = \begin{cases} 1, & \text{ha } X(t)=i, \\ 0, & \text{ha } X(t)\neq i. \end{cases}$ (A lektor megjegyzése.)

Legyen Y_t valószínűségi változó, melyre

$$Y_t = \begin{cases} \int_{t'}^{t} \chi_0(\tau)\,d\tau, & \text{ha} \quad X_t = 0 \quad \text{vagy} \quad X_t = 1 \\ 0, & \text{máskor,} \end{cases} \tag{2.2}$$

vagyis Y_t a teljes idő, amit a kiszolgáló a 0 állapotban tölt a $[t', t]$ időintervallumon, ahol t' az utolsó t előtti időpont, amikor a kiszolgáló a 2-es állapotban volt. A t időpontban bekövetkezett meghibásodás után a kiszolgáló az 1-es állapotban $h(Y_t)$ ideig marad.

Tehát a kiszolgáló helyreállításának ideje függvénye Y_t-nek, vagyis annak, hogy mennyi idő telt el a hiba bekövetkezése és észlelése között.

Fordítsuk figyelmünket a (2.1) és I—IV. által definiált $\{X_t, t \geqq 0\}$ sztochasztikus folyamat stacionárius valószínűségeire.

$$\Pi_j \triangleq \lim_{t\to\infty} P(X_t = j) \quad (j = 0, 1, 2, \ldots). \tag{2.3}$$

Mivel a folyamat nem *Markov-típusú,* illetve nem *Markov-felújítási folyamat* CINLAR [4], ezért Π_j-t nem számíthatjuk ki az érvényes klasszikus eredményekből. Viszont egy elemi gondolatmenettel célt érünk:

Legyen: $$Z_t = \begin{cases} 1, & \text{ha} \quad X_t \in \{0, 1\}, \quad t \geqq 0 \\ 0, & \text{máskor.} \end{cases}$$

Vegyük észre, hogy:

$$P(X_t = 1) = P(X_t = 1 | Z_t = 1) P(Z_t = 1).$$

$\{Z_t, t \geqq 0\}$ alternáló felújítási folyamat, így kihasználhatjuk azok tulajdonságait Π_0 kiszámítására.

Válasszunk ki egy olyan időpontot két ellenőrzési pont között, melyre a teljes $X_t=0$-ban töltött idő y. A meghibásodások *Poisson-folyamat* szerint történnek, annak jól ismert tulajdonságaiból következik (vö. Cox [5] 27—28. old.), hogy n meghibásodást feltételezve, a meghibásodások függetlenek és az időintervallumon egyenletes eloszlásúak $f(x)=\frac{1}{y}$ eloszlásfüggvénnyel.

Így a várható értéke annak a teljes időnek, amit $\{Z_t\}$ az 1 állapotban tölt adott y-ra:

$$y + \sum_{n=0}^{\infty} n \left(\frac{\gamma y}{n!}\right)^n e^{-\gamma y} \int_0^y \frac{h(x)}{y}\,dx = y\left(1 + \gamma \int_0^y \frac{h(x)}{y}\,dx\right),$$

és ennek várható értéke y összes lehetséges értékére:

$$E(Y)+\gamma \int_0^\infty dF(y) \int_0^y h(x)dx. \tag{2.4}$$

Így:

$$\lim_{t\to\infty} P(Z_t = 1) = \frac{E(Y)+\gamma \int_0^\infty dF(y) \int_0^y h(x)\,dx}{E_c+E(Y)+\gamma \int_0^\infty dF(y) \int_0^y h(x)dx} \tag{2.5}$$

felhasználva egy, az alternáló felújítási folyamatoknál általánosan használt tételt (Cox [5]). Megjegyezzük, hogy

$$\Pi_0 = \lim_{t\to\infty} P(X_t = 0 | X_t = 0 \text{ vagy } 1) \lim_{t\to\infty} P(Z_t = 1).$$

A bizonyításhoz felhasználunk egy I. MITRANI-tól származó gondolatot, amelyért köszönettel tartozom. Legyen $a(x)$ két egymást követő ellenőrzési pont közötti teljes idő eloszlásának sűrűségfüggvénye, valamint $b(x)$ várható értéke a teljes időnek, ami alatt a kiszolgáló a 0 állapotban van két ellenőrzési pont között ama feltétel mellett, hogy ennek az időtartamnak hossza, $\mathscr{L}$, egyenlő x-szel. Adott x esetén vegyük észre (használva a „véletlen megfigyelő" sajátosságait), l. pl. (TAKÁCS [14] 10—11. old.), hogy

$$\lim_{t\to\infty} P(X_t = 0 | X_t = 0 \text{ vagy } X_t = 1 \text{ és } \mathscr{L} = x) = \frac{b(x)}{x},$$

mivel a t időpont egyenletesen oszlik el a $[0, x]$ intervallumon.

$$\Pi_0 = \int_0^\infty \frac{b(x)}{x} \frac{xa(x)dx}{\int_0^\infty xa(x)\,dx} = \frac{\int_0^\infty b(x)a(x)\,dx}{\int_0^\infty xa(x)\,dx},$$

mivel $xa(x)\,dx\left(\int_0^\infty xa(x)\,dx\right)^{-1}$ annak a valószínűsége, hogy a véletlen megfigyelés egy $\mathscr{L}=x$ intervallumba esik, (ismét TAKÁCS [14] 10—11. old.-t felhasználva).

De

$$\int_0^\infty b(x)a(x)dx = E(Y),$$

és a 0 állapotban két egymást követő ellenőrzési pont között töltött időt, $\int_0^\infty xa(x)\,dx$-et (2.4)-ből kaphatjuk meg.

Így érvényes a

2.1. TÉTEL. Az $\{X_t, t \geqq 0\}$ folyamathoz tartozó stacionárius valószínűségekre:

$$\Pi_0 = E(Y)\Big(E_c + E(Y) + \gamma \int_0^\infty dF(y) \int_0^y h(x)\,dx\Big)^{-1}, \tag{2.6}$$

$$\Pi_1 = \frac{\gamma \int_0^\infty dF(y) \int_0^y h(x)\,dx}{E_c + E(Y) + \gamma \int_0^\infty dF(y) \int_0^y h(x)\,dx}, \tag{2.7}$$

és

$$\Pi_2 = 1 - \Pi_0 - \Pi_1.$$

Mivel Π_0 annak a stacionárius valószínűsége, hogy a kiszolgáló elérhető a fogyasztók számára, ezért Π_0-t a kiszolgáló hozzáférhetőségének fogjuk nevezni.

A következő fejezetben Π_0 ismét szerepelni fog a sorhosszúság ergodicitásának feltételeiben. Π_0-t felírhatjuk a következő, könnyebben kezelhető alakba:

$$\Pi_0 = \left[1 + \frac{E_c}{E(Y)} + \frac{\gamma}{E(Y)} \int_0^\infty h(y)\big(1 - F(y)\,dy\big)\right]^{-1}. \tag{2.8}$$

Az optimális ellenőrzési intervallumhossz

Több szerző (YOUNG [15], CHANDY [2, 3], GELENBE [8, 9], KOVALENKO [10] vizsgálta az optimális ellenőrzési szakaszok megválasztásának problémáját. Mielőtt röviden áttekintenénk az eddigi eredményeket, bevezetünk néhány definíciót.

Az *ellenőrzési szakasz* (ESZ) a (2.2)-ben definiált Y valószínűségi változó $F(y)$ eloszlásfüggvénnyel, ez az a teljes időtartam két egymást követő ellenőrzési pont között, ami alatt a rendszer alkalmas még le nem bonyolított tranzakciók bonyolítására.

A *teljes tényleges idő ellenőrzési pontok között* (TTI) az az idő, amely két ellenőrzési pont között eltelik, ami az ESZ-ből és olyan időszakaszokból áll, amikor meghibásodás javítása történik. Ezt az időt — mely egy valószínűségi változó — ξ-vel, eloszlásfüggvényét $\hat{F}(y)$-nal fogjuk jelölni.

YOUNG [15] és CHANDY [2, 3] felteszik, hogy a meghibásodás gyakorisága igen kicsi, $(\gamma E(\xi) \ll 1)$. Eme feltétel mellett kiszámítják $E(\xi)$-nek azt az értékét, amely mellett maximális a hatékonyság. CHANDY [2, 3] felteszi azt is, hogy ξ=konstans, azaz determinisztikus. GELENBE [8] felteszi, hogy Y exponenciális eloszlású és kiszámítja $E(Y)$-nak azt az értékét, amely mellett maximális a hatékonyság. KOVALENKO [10] felteszi, hogy Y konstans.

Most megvizsgáljuk (l. (2.6)) $F(y)$-nak azt a választását, amely mellett a hozzáférhetőség, Π_0, maximális: ez az ellenőrzési szakasz optimalizálásának általános problémája. Az eddigi eredmények korlátozó feltételezések mellett érvényesek, mivel vagy γ-t tételezték fel igen kicsinek, vagy $F(y)$-nak csak bizonyos speciális eseteit vizsgálták, vagy mindkét korlátozással éltek.

Legyen $H(y) \triangleq \int_0^y h(x)\,dx$ és jelöljük J_a-val azon eloszlásfüggvények összességét, melyek várható értéke a rögzített $a \geqq 0$-val egyenlő. Legyen $H(Y)$ várható értéke E_H, ha eloszlásfüggvénye valami $F \in J_a$.

2.2. Megjegyzés: Ha $E_H \geqq H(a)$ minden $F \in J_a$-ra, akkor J_a-nak az az eleme, amelyre Π_0 maximális, rögzített $a \geqq 0$ mellett:

$$U_a(y) = \begin{cases} 1, & \text{ha } y \geqq a, \\ 0, & \text{ha } y < a. \end{cases}$$

Megjegyezzük, hogy $E_H \geqq H(a)$, ha $H(y)$ konvex (azaz $h(y)$ monoton növekvő).

Bizonyítás: Mivel rögzített a esetén

$$\Pi_0 = \left[1 + \frac{E_c}{a} + \frac{\gamma}{a} \int_0^\infty dF_a(y)\, H(y)\right]^{-1} = \left[1 + \frac{E_c}{a} + \frac{\gamma}{a} E_H\right]^{-1},$$

így $\int_0^\infty dU_a(y) H(y) = H(a)$ adódik a feltételekből.

Eddig a javítási időt, $h(y)$-t általánosan kezeltük. Most feltesszük, hasonlóan az eddigi vizsgálatokhoz, hogy

$$h(y) = \alpha y + \beta, \tag{2.9}$$

ahol $\alpha, \beta > 0$ konstansok. Ez a feltételezés kézenfekvő az adatbázisokon végrehajtott műveletekre: β egy rögzített időtartam, mely ahhoz szükséges, hogy az ellenőrzés pillanatában tárolt információt áttöltsük a központi memóriába, és αy annak az időnek felel meg, ami alatt ismételten végrehajtjuk azokat a tranzakciókat, amelyeket a rendszer y-nyi idő alatt szolgált ki. α választására még vissza fogunk térni a dolgozat végén. Megjegyezzük, hogy:

$$H(y) = \frac{\alpha}{2} y^2 + \beta y \tag{2.10}$$

és

$$E_H = \frac{\alpha}{2} E(Y^2) + \beta E(Y) \geqq H(E(Y)) = \frac{\alpha}{2} (E(Y))^2 + \beta E(Y).$$

Tegyük fel, hogy az optimális ellenőrzési intervallumhossz $F^*(y)$ eloszlásfüggvénye J_a-ban van (azaz $a = \int_0^\infty y dF^*(y)$ valamilyen $a \geqq 0$-ra). Ekkor a fenti megjegyzésből $F^*(y) = U_0(y)$ és:

$$\Pi_0(F^*(y)) = \left[1 + \frac{E_c}{a} + \alpha\gamma\frac{a}{2}\right]^{-1}. \tag{2.11}$$

Vizsgáljuk meg α és β választásának kérdését (2.9)-ben. Jelöljük $p^*(0, 0)$-val annak stacionárius valószínűségét, hogy a rendszer üres, feltéve, hogy a 0 állapotban van (normál működés). Ekkor egy y hosszúságú intervallumon mutatott normális működés alatt átlagosan $y(1 - p^*(0, 0))$ ideig lesz aktív a rendszer. Tegyük fel,

hogy tranzakciók a rendszerbe λ intenzitással érkeznek és a kiszolgálás intenzitása μ. Ekkor egy y hosszúságú intervallumon a lebonyolított tranzakciók átlagos száma $\mu y(1-p^*(0,0))$. Legyen k azoknak a tranzakcióknak az aránya, amelyeket meghibásodás miatt ismételten futtatni kell; ha mindegyik újbóli futtatás ugyanolyan hosszú, mint az eredeti, akkor ésszerű feltenni, hogy y időegységnyi normál működés után a megismételt futtatások összideje átlagosan

$$\frac{1}{\mu}(\mu k y(1-p^*(0,0))),$$

így:

$$\alpha \approx k(1-p^*(0,0))$$

és

$$h(y) = ky(1-p^*(0,0))+\beta. \tag{2.12}$$

A 3. fejezetben be fogjuk bizonyítani, hogy:

$$p^*(0,0) = 1-\frac{\lambda}{\mu \Pi_0}.$$

Így:

$$\alpha = \frac{k\lambda}{\mu \Pi_0}. \tag{2.13}$$

(2.11) és (2.13) felhasználásával:

$$\Pi_0\left(1+\frac{E_c}{a}+\gamma\beta\right) = 1-\frac{k\gamma a\lambda}{2\mu}. \tag{2.14}$$

(2.14)-ből most már lehetséges az optimális ellenőrzési intervallumhossz, $\hat{a}$ meghatározása.

2.3. *Eredmény.*

A (2.14) képletben Π_0-t a-ban

$$\hat{a} = \frac{E_c}{1+\beta\gamma}\left(\sqrt{1+\frac{2(1+\beta\gamma)}{\varrho\gamma k E_c}}-1\right) \tag{2.15}$$

maximalizálja, itt:

$$\varrho = \frac{\lambda}{\mu}.$$

(2.15) egy általunk levezetett új formula az optimális ellenőrzési intervallumhosszra, nyilvánvaló, hogy a képletben szerepel ϱ, a tranzakciót bonyolító rendszer leterheltségi faktora. $\hat{a}$-tól lényegessen különböző érték $(2Ec/\alpha\gamma)^{1/2}$ adódik az optimális ellenőrzési intervallumhoszra YOUNG [15] és CHANDY [3] eredményeiből, ami nem alkalmazható α-nak (2.13) alapján történő választása esetén.

Tekintsük azt az esetet, amikor $p^*(0,0)\cong 1$ (a rendszer erősen leterhelt) és

$$\frac{2(1+\beta\gamma)}{\varrho\gamma k E_c} \gg 1.$$

Ekkor:

$$\hat{a} \approx \sqrt{\frac{E_c}{\varrho\gamma k(1+\beta\gamma)}},$$

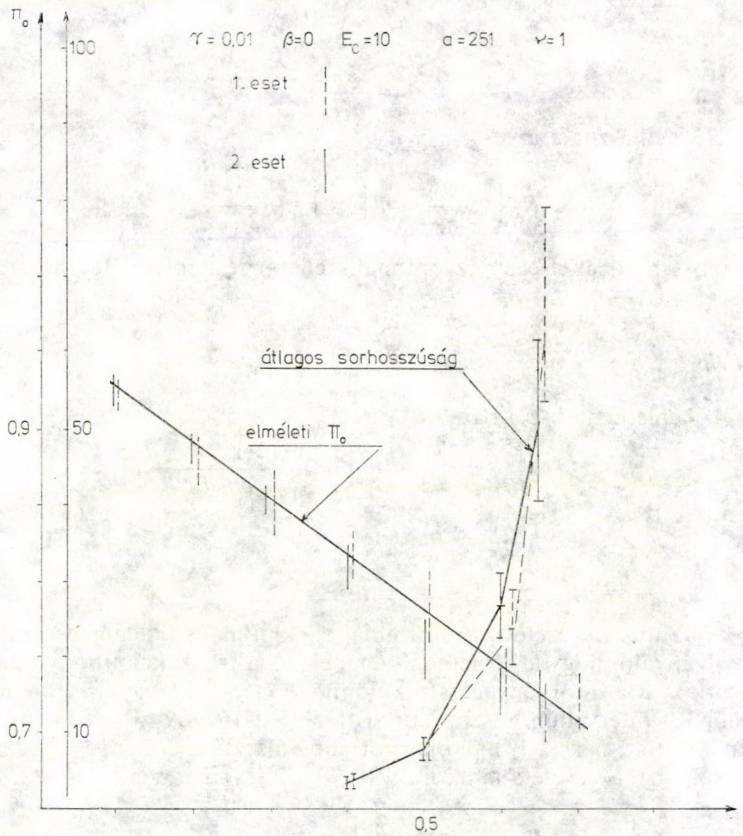
amely emlékeztet YOUNG [15] és CHANDY [3] formuláira. Egy másik érdekes eset, amikor:

$$2(1+\beta\gamma) \ll \varrho\gamma k E_c.$$

Ekkor:

$$\hat{a} \approx \frac{1}{\varrho\gamma k}.$$

A javítási időre adott (2.12) képlet, valamint (2.13) használatának jogosultságát FLAMAND [7] egy sor szimulációval erősítette meg. Az 1. ábrán mintát láthatunk az eredményekből. Két kísérlet lett elvégezve.



1. ábra

Az 1. esetben a lebonyolított tranzakciók számát (a felülvizsgáló sáv tartalmát) rögzítették, ugyanennyi tranzakció kerül ismételten feldolgozásra meghibásodás esetén. A 2. esetben $p^*(0, 0)$-t becsülték azzal, hogy mérték az üres állapot időarányát a rendszer 0 állapotában egy meghibásodásig, ez az, amit (2.12)-ben használtunk. A szimuláció igen jó egyezést mutat az 1-es és a 2-es esetek között Π_0-ra, és elfogadható egyezést (a konfidencia intervallumok átfedik egymást) a tranzakciós igények átlagos sorhosszúságára. Mindkét esetben 95%-os konfidencia-intervallumokat vettünk fel. Π_0-ra (2.14) felhasználásával adódó elméleti eredmény szintén fel van tüntetve. Az átlagos sorhosszúság képletét nem tudtuk egzaktul kiszámítani (l. 3. fejezet).

A teljes tényleges idő optimális ellenőrzési intervallumhossz esetén

Számítsuk ki $\hat{F}(y)$-t, az ellenőrzési pontok közötti teljes tényleges időt azok optimális választása esetén.

(2.4)-ből

$$E(\xi) = (1+\gamma\beta)\left(\sqrt{\frac{2E_c}{\alpha\gamma}} + E_c\right).$$

Célunk, hogy megkapjuk $\hat{f}^*(s) \triangleq \int_0^\infty e^{-sy} d\hat{F}(y)$-t. Megjegyezzük, hogy konstans $\hat{a}$ esetén, ha adott n időpont a $[0, \hat{a}]$ intervallumban *Poisson-pontfolyamat* szerinti eloszlással, akkor az n időpont független, és a $[0, \hat{a}]$ intervallumon egyenletes eloszlású. Így az n javítási idő is független, közös sűrűségfüggvényük:

$$f_h(y) = \begin{cases} (\alpha\hat{a})^{-1}, & \text{ha } y \in [\beta, \beta+\hat{a}\alpha], \\ 0, & \text{máskor.} \end{cases}$$

Ezért, ha

$$f_h^*(s) \triangleq \int_0^\infty f_h(y) e^{-sy} dy,$$

akkor

$$\hat{f}^*(s) = [e^{-s\hat{a}}]\left[\sum_{n=0}^{\infty} \frac{(\gamma\hat{a})^n}{n!} e^{-\gamma\hat{a}} (f_h^*(s))^n\right] = \exp\left[-s\hat{a} + \gamma\hat{a}(f_h^*(s) - 1)\right],$$

ahol:

$$f_h^*(s) = (\hat{a}\alpha s)^{-1} e^{-\beta s}(1 - e^{-\alpha s}).$$

ξ szórásnégyzetére:

$$E[(\xi - E(\xi))^2] = \gamma\hat{a}\left[\beta^2 + \hat{a}\alpha\beta + \frac{(\alpha\hat{a})^2}{3}\right]. \tag{2.16}$$

Látjuk, a kifejezés az optimális intervallumhossz választása esetén elég nagy lehet, még γ kis értéke esetén is.

3. A várakozási sor analízise

Ebben a fejezetben a kiszolgálásra váró tranzakciók sorbanállási folyamatát, valamint a kiszolgáló állapotát analizáljuk. Célunk az, hogy meghatározzuk a rendszer teljesítményének mutatóit, pl. a kiszolgáló *kihasználtságát* (azaz annak stacionárius, ergodikus valószínűségét, hogy a sorhosszúság nem zéró), a *telítettségi feltételt* (azaz a tranzakciós igények érkezési intenzitásának ama λ értékét, ami felett a stacionárius sorhosszúság végtelen).

Az ilyen analízis új a sorbanállási folyamatok körében. Számunkra az $(N, X, Y) \triangleq \{N_t, X_t, Y_t, t \geqq 0\}$ sztochasztikus folyamat érdekes, ahol N_t és X_t az igénylők száma, illetve a rendszer állapota a t időpontban; Y_t-t pedig (2.2)-ben definiáltuk.

Szemléletesen, az Y_t valószínűségi változó lehetővé teszi a $h(Y_t)$ időtartam kiszámítását, amely a t időpontban bekövetkezett meghibásodás kijavításához szükséges a rendszer 0 állapota esetén. A sorbanállási folyamat analízisét a következő feltételek mellett végezzük. A tranzakciós igények λ paraméterű *Poisson-folyamat* szerint érkeznek be, és FCFS rendszerben lesznek kiszolgálva, a rendszer 0 állapota esetén. Nem történik kiszolgálás a kiszolgáló 1-es, ill. 2-es állapotában. A kiszolgálási idő μ paraméterű exponenciális eloszlású

Vizsgálatunkat az $(N, Y|X=0) \triangleq \{N_t, Y_t|X_t=0, t \geqq 0\}$ feltételes folyamattal kezdjük.

Az $(N, Y|X=0)$ *feltételes folyamat vizsgálata*

Legyen $p(n, y, t)=P(N_t=n, Y_t=y|X_t=0)$[2], és vezessük be a következő jelöléseket:

$$\eta(y) = \frac{f(y)}{1-F(y)}, \tag{3.1}$$

$$r_y(j) = \frac{(\lambda h(y))^j}{j!} \exp(-\lambda h(y)), \quad j \geqq 0.$$

Nyilván $\eta(y)=\dfrac{f(y)}{1-F(y)}$ annak valószínűsége, hogy ellenőrzőpontot létesítünk a $[t, t+dy)$ balról zárt, jobbról nyílt intervallumon, feltéve, hogy $X_t=0$ és $Y_t=yr_y(j)$ annak a valószínűsége, hogy j igény érkezése történik egy t időpontban kezdődött javítási időszak alatt, ugyanakkor $X_t=0$ és $Y_t=y$.

A következő egyenleteket a szokásos módon kapjuk, $y \geqq 0$ és $n>0$-ra:

$$\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial y}\right) p(n, y, t) = -(\lambda+\mu+\eta(y)+\gamma)\, p(n, y, t)+ \tag{3.2}$$

$$+\lambda p(n-1, y, t)+\mu p(n+1, y, t)+\gamma \sum_{j=0}^{n} r_y(j)\, p(n-j, y, t).$$

[2] Vagyis $p(n, y, t)\Delta y = P(N_t=n, y \leqq Y_t \leqq y+\Delta y|X_t=0)$ $\Delta y>0$-ra.

Ha $n=0$, $y\geqq 0$:

$$\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial y}\right)p(0,y,t)=-(\lambda+\eta(y)+\gamma)p(0,y,t)+ \tag{3.3}$$
$$+\mu p(1,y,t)+\gamma r_y(0)p(0,y,t).$$

$y=0$-ra és $n\geqq 0$-ra:

$$p(n,0,t)=\int_0^\infty \sum_{j=0}^{n} c(j)p(n-j,y,t)\eta(y)dy, \tag{3.4}$$

ahol

$$c(j)=\int_0^\infty \frac{(\lambda z)^j}{j!}e^{-\lambda z}dC(z), \quad j\geqq 0 \tag{3.5}$$

annak a valószínűsége, hogy j igény érkezése történik ellenőrzőpont létesítése alatt (l. 2. fejezet).

A következőkben az $(N, Y|X=0)$ folyamat stacionárius eloszlását fogjuk megvizsgálni.

3.1. TÉTEL. (3.2), (3.3) és (3.4)-nek akkor és csak akkor létezik stacionárius megoldása, ha

$$\frac{\lambda}{\mu}<\left[1+\frac{E_c}{E(Y)}+\frac{\gamma}{E(Y)}\int_0^\infty h(y)(1-F(y))dy\right]^{-1}.$$

A „csak akkor" rész bizonyítása (szükséges feltétel):

Legyen (3.2)-ben és (3.3)-ban $\dfrac{\partial p(\cdot,\cdot,\cdot)}{\partial t}=0$.

Legyen:

$$p^*(n,s)=\int_0^\infty e^{-sy}p(n,y)dy,$$

ahol elhagytuk a t-től való függőséget.

Legyen:

$$G_y(x)=\sum_{n=0}^{\infty}x^n p(n,y), \quad G^*(s,x)=\int_0^\infty G_y(x)e^{-sy}dy.$$

Innen $n>0$-ra

$$sp^*(n,s)-p(n,0)=-(\lambda+\mu+\gamma)p^*(n,s)+\mu p^*(n+1,s)+ \tag{3.6}$$
$$+\lambda p^*(n-1,s)-\int_0^\infty e^{-sy}\eta(y)p(n,y)dy+\sum_{j=0}^{n}\gamma\int_0^\infty r_y(j)p(n-j,y)e^{-sy}dy,$$

és:

$$sp^*(0,s)-p(0,0)=-(\lambda+\gamma)p^*(0,s)+\mu p^*(1,s)-$$
$$-\int_0^\infty \eta(y)e^{-sy}p(0,y)dy+\gamma\int_0^\infty r_y(0)p(0,y)e^{-sy}dy,$$

ahonnan:

$$G^*(s,x)[s+\lambda(1-x)+\mu(1-1/x)+\gamma]-G_0(x) = \tag{3.7}$$

$$= p^*(0,s)\mu(1-1/x)+\gamma H^*(s,x)-\int_0^\infty \eta(y)G_y(x)e^{-sy}dy,$$

ahol:

$$H^*(s,x) = \int_0^\infty e^{-sy}e^{-\lambda(1-x)h(y)}G_y(x)dy.$$

(3.4) felhasználásával

$$G_0(x) = C^*(\lambda(1-x))\int_0^\infty \eta(y)G_y(x)dy, \tag{3.8}$$

ahol

$$C^*(\lambda(1-x)) = \int_0^\infty e^{-\lambda(1-x)z}dC(z).$$

Így (3.7)-ből

$$G^*(s,x) = \frac{p^*(0,s)\mu(1-1/x)-\int_0^\infty \eta(y)G_y(x)e^{-sy}dy+G_0(x)}{s+\lambda(1-x)+\mu(1-1/x)}+ \tag{3.9}$$

$$+\frac{\gamma[H^*(s,x)-G^*(s,x)]}{s+\lambda(1-x)+\mu(1-1/x)},$$

ahol $G_0(x)$ (3.8)-ból adódik. Világos, hogy ha a stacionárius eloszlás $\{p(n,y), n\geqq 0, y\geqq 0\}$ létezik, akkor szükségképpen:

$$\lim_{x\to 1}\lim_{s\to 0} G^*(s,x) = 1.$$

(3.9)-ben a határátmenetet elvégezve, a jobb oldalon mindkét tagban határozatlan alakot kapunk.

Ezért legyen először:

$$\lim_{s\to 0} G^*(s,x) = \frac{p^*(0,0)\mu(1-1/x)+[C^*(\lambda(x-1))-1]\int_0^\infty \eta(y)G_y(x)dy}{\lambda(1-x)+\mu(1-1/x)}+ \tag{3.10}$$

$$+\frac{\gamma[H^*(0,x)-G^*(0,x)]}{\lambda(1-x)+\mu(1-1/x)},$$

és most alkalmazzuk a *l'Hôspital-szabályt:*

$$\lim_{x\to 1}\frac{d}{dx}C^*(\lambda(1-x)) = \lambda E_c,$$

$$\lim_{x\to 1} C^*(\lambda(1-x)) = 1,$$

$$\lim_{x\to 1}\frac{d}{dx}H^*(0,x) = \lim_{x\to 1}\frac{d}{dx}G^*(0,x)+\int_0^\infty \lambda h(y)G_y(1)dy,$$

úgyhogy:

$$\lim_{x\to 1}\lim_{s\to 0} G^*(s,x) = \frac{\mu p^*(0,0)+\lambda E_c\int_0^\infty \eta(y)G_y(1)dy+\lambda\gamma\int_0^\infty h(y)G_y(1)dy}{-\lambda+\mu}. \tag{3.11}$$

A bizonyítás további menetéhez szükségünk lesz a következő lemmára.

3.2 LEMMA. $G_y(1)=\lim_{t\to\infty}\sum_{n=0}^{\infty}p(n,z,t)$-re igaz, hogy

$$G_y(1)=\frac{1-F(y)}{E(Y)}.$$

Bizonyítás. A lemma ismét TAKÁCS [14] 10—11. old. felújítási folyamatokra vonatkozó jól ismert eredményének következménye.

Legyen:

$$K(y,t)=\int_0^\infty p(n,z,t)dz$$

úgy, hogy

$$\lim_{t\to\infty}K(y,t)=\int_0^y G_z(1)dz.$$

Ekkor:

$$K(y,t)=\sum_{k=1}^{\infty}P[t<\sigma_k\leqq t+y<\sigma_{k+1}], \tag{3.12}$$

ahol (mivel az $(N, Y|X=0)$ feltételes folyamattal foglalkozunk) minden t időpillanat olyan időtartamra vonatkozik, ami alatt a kiszolgáló a 0 állapotban van, vagyis t a „valós" t' időnek felel meg (az (N, Y, X) folyamat idejében), ahol:

$$t=\int_0^{t'}\chi_0(\tau)d\tau$$

(3.12)-ben $\sigma_1<\ldots<\sigma_k<\sigma_{k+1}<\ldots$-k az $Y_{\sigma_i}=0$ feltétellel vannak definiálva. Nyilván $(\sigma_{k+1}-\sigma_k)$-k függetlenek és közös eloszlásfüggvényük $F(y)$.

Így

$$\lim_{t\to\infty}K(y,t)=\int_0^y\frac{1-F(x)}{E(Y)}dx, \tag{3.13}$$

ami igazolja a lemma állítását.

Visszatérve (3.11)-re és a 3.1. tételre, látjuk, hogy

$$\int_0^\infty\eta(y)G_y(1)dy=\int_0^\infty\frac{dF(y)}{1-F(y)}\frac{1-F(y)}{E(Y)}=\frac{1}{E(Y)}.$$

Ezért

$$p^*(0,0)=1-\frac{\lambda}{\mu}\left[1+\frac{E_c}{E(Y)}+\gamma\int_0^\infty\frac{1-F(y)}{E(Y)}h(y)dy\right]=1-\frac{\lambda}{\mu\Pi_0}, \tag{3.14}$$

ha (3.2), (3.3) és (3.4) stacionárius megoldása létezik. De szükségképpen

$$p^*(0, 0) = \int_0^\infty p(0, y)dy > 0,$$

amiből következik az állítás első része.

Az „akkor" rész bizonyítása (elégséges feltétel):

Az elégséges feltétel bizonyításához tekintsük a σ_i, $(Y_{\sigma_t}=0)$ után közvetlenül következő σ_i^+ pillanatokat az $(N, Y|X=0)$ folyamat idejében.

Az

$$(\hat{N}|X = 0) \equiv \{N_{\sigma_t^+},\ i \geqq 1 | X_t = 0,\ t \geqq 0\}$$

folyamat diszkrét idejű *Markov-lánc,* könnyen látható, hogy aperiodikus és irreducibilis. Először ennek ergodicitását bizonyítjuk a tétel feltételei mellett felhasználva PAKES most következő lemmáját.

3.3. LEMMA. Legyen $\{B_i, i \geqq 1\}$ aperiodikus, irreducíbilis *Markov-lánc.* B_i ergodikus, ha teljesülnek a következő feltételek:

$$|E(B_{i+1} - B_i | B_i = j)| < \infty$$

minden j-re,

és

$$\limsup_{j \to \infty} E(B_{i+1} - B_i | B_i = j) < 0.$$

Legyen $\alpha = \sigma_{i+1} - \sigma_i$; a következőkben minden eseményt az $N_{\sigma_i^+} = j$ feltétel mellett tekintünk. Legyen T a $[\sigma_i^+, \sigma_{i+1}^+)$ intervallumon az a teljes időtartam, ami alatt a sorhosszúság nem zéró. Nyilván $T \leqq \alpha$. Legyen $D(\alpha)$ a $[\sigma_i^+, \sigma_{i+1}^+)$ intervallumon kiszolgált, a sorból kilépő fogyasztók száma

$$1 \geqq P(T = \alpha) \geqq P(D(\alpha) < j),$$

de

$$P(D(\alpha) < j) \geqq \int_0^\infty dF(\tau) \sum_{l=0}^{j-1} \left(\frac{\mu\tau}{l!}\right)^l e^{-\mu\tau},$$

mivel a $[D(\alpha) < j]$ eseményből következik, hogy α hosszúságú időintervallumon, ahol a sorhosszúság végig nagyobb zérusnál, a sorból kilépők száma kisebb, mint j. Ezért:

$$\lim_{j \to \infty} P[T = \alpha] = 1. \tag{3.15}$$

Most tekintsük az

$$E[N_{\sigma_{i+1}^+} - N_{\sigma_i^+} | N_{\sigma_i^+} = j] = E[D(\alpha) - A(\alpha)]$$

feltételes várható értéket, ahol $A(\alpha)$ a beérkezések száma a $[\sigma_i^+, \alpha_{i+1}^+)$ intervallumon.

Nyilván:

$$E[A(\alpha)] = \lambda\Big[E(Y)+E_c+\gamma\int_0^\infty h(y)(1-F(y))dy\Big],$$

$$E[D(\alpha)] = \mu E(T),$$

$$\lim_{j\to\infty} E(T) = E(Y).$$

Tehát, ha a tétel feltételei teljesülnek, akkor egyben teljesül a *Pakes-lemma* mindkét feltétele $(\hat{N}|X=0)$-ra, így ez a *Markov-lánc* ergodikus. Könnyen látható, hogy $p(n, y)$ szintén létezik.

Egy közvetlen következmény a

3.4. TÉTEL. A 3.1. tétel feltételei mellett, az $(\hat{N}|X=0)$ sztochasztikus folyamatnak létezik stacionárius eloszlása, és ennek generátorfüggvénye $G(x)\triangleq\lim_{s\to 0} G^*(s, x)$ kielégíti a

$$G(x) = \frac{\gamma H^*(0, x)+p^*(0, 0)\mu(1-1/x)+\big(C^*(\lambda(x-1))-1\big)\int_0^\infty \eta(y)G_y(x)dy}{\gamma+\lambda(1-x)+\mu(1-1/x)}$$

egyenletet.

Az (N) folyamat vizsgálata

Legyen az (eredeti) sorhosszúság pontfolyamata $(N)\triangleq\{N_t, t\geqq 0\}$. A következőkben kiterjesztjük a 2. fejezet és a 3. fejezet eddigi eredményeit az (N) folyamatra.

Munkánk most már jórészt technikai. Vessük be a következő jelöléseket $t\geqq 0$-ra:

$$a(n, u, t) \triangleq P[N_t = n,\ U_t = u,\ X_t = 2],$$

ahol $X_t=2$-re:

$$t-U_t \triangleq \sup\{\tau:\ \tau < t \text{ és } X_\tau = 0\}$$

és

$$b(n, y, v, t) = \begin{cases} 0, & \text{ha } v > h(y), \\ P[N_t = n,\ Y_t = y,\ V_t = v,\ X_t = 1], & \text{máskor,} \end{cases}$$

ahol $X_t=1$-re:

$$t-V_t \triangleq \sup\{\tau:\ \tau < t \text{ és } X_\tau = 0\}.$$

Legyen $a(n, u)$ és $b(n, y, v)$ az ezeknek megfelelő határérték (feltéve, hogy létezik), ha a $t\to\infty$ határátmenetet elvégezzük.

3.5. Megjegyzés:

$$a(n, u) = \Pi_0(1-C(u))\sum_{j=0}^{n}\frac{(\lambda\mu)^j}{j!}e^{-\lambda m}\int_0^\infty p(n-j, y)\eta(y)dy.$$

3.6. Megjegyzés:

$$b(n, y, v) = \Pi_0 \gamma \sum_{j=0}^{n} \frac{(\lambda v)^j}{j!} e^{-\lambda v} p(n-j, y).$$

3.7. TÉTEL. $a(n, u)$ akkor és csak akkor létezik minden $n \geqq 0$, $y \geqq 0$-ra, és $b(n, y, v)$ akkor és csak akkor létezik minden $y \geqq 0$-ra és $0 \leqq v \leqq h(y)$-ra, ha $p(n, y)$ létezik minden $n \geqq 0$, $y \geqq 0$-ra.

IRODALOM

[1] BOROVKOV, A. A., *Stochastic Processes in Queueing Theory* (Springer, New York, 1976).
[2] CHANDY, K. M., "A survey of analytic models of roll-back and recovery strategies", *IEEE Computer* **8** (1975) No. 5, 40—47.
[3] CHANDY, K. M., BROWNE, J. C., DISSLY, C. W. and UHRING, W. R., "Analytical models for roll back and recovery strategies in data base systems", *IEEE Transactions on Software Engineering* **1** (1975) 100—110.
[4] CINLAR, E., *Introduction to Stochastic Processes* (Prentice Hall, New York, 1975).
[5] COX, D. R., *Renewal Theory* (Methuen, London, 1962).
[6] COX, D. R. and LEWIS, P. A. W., *The Statistical Analysis of Series of Events* (Methuen, London, 1966).
[7] FLAMAND, J. and GELENBE, E., "Simulation of roll-back recovery in a data base system", megjelenés alatt.
[8] GELENBE, E. and DEROCHETTE, D., "On the stochastic behaviour of a computer system under intermittent failures", *Modelling and Performance Evaluation of Computer Systems, H. Beilner and E. Gelenbe eds.*, North-Holland, 1976.
[9] GELENBE, E., "On roll-back recovery with multiple checkpoints", *Proceedings of the and International Symposium on Software Engineering* (IEEE Press, San Francisco, 1976).
[10] KOVALENKO, I. N. and STOIKOVA, L. S., "On the productivity of a system and the problem solving time in the presence of random failures and aperiodic memorization of the results" *Kibernetic* **5** (1974) 73—75. (Cybernetics Plenum Publishing Corporation New York, 1976).
[11] LOYNES, R. M., "The stability of a queue with non-independent interarrival and service times", *Proc. Cambridge Philos. Soc.* **58** (1962) 497—520.
[12] PAKES, A. G., "Some conditions for ergodicity and recurrence of Markov chains", *Operations Research* **17** (1969) 1058—1061.
[13] ROBIN, M., megjelenés alatt.
[14] TAKÁCS, L., *Introduction to the Theory of Queues* (Oxford University Press, New York, 1962).
[15] YOUNG, J. W., "A first-order approximation to the optimum checkpoint interval", *CACM* **17** (1974) 530—531.

(Beérkezett: 1979. április 19.)

EROL GELENBE
DEPARTEMENT DE MATHEMATIQUES UNIVERSITE PARIS-NORD
AVENUE J. B. CLÉMENT, 93 VILLETANEUSE

ON THE OPTIMUM CHECKPOINT INTERVAL

E. GELENBE

In this paper we give a detailed theoretical treatment of the problem of determining the optimum checkpoint interval, i.e. the total time between successive checkpoints during which the database system is not recovering from failures, which maximizes system availability.

We first obtain, under some general assumptions, the expression for the availability. We then show that the optimum checkpoint interval must be deterministic and that it is a function of the system load. An explicit expression for its value is given. These are the main practical contributions of the paper.

The results of practical interest are obtained by means of a theoretical analysis of a queueing system representing system behaviour. This queue has failures and repair times which are a function of the age of the failure with respect to the most recent maintenance (checkpoint) epoch. The queueing model appears to be novel and may have applications to other reliability studies.

Alkalmazott Matematikai Lapok **5** (1979) 157—164

A NEMLINEÁRIS FOLYAMPROBLÉMA EGY MEGOLDÁSI MÓDJÁRÓL

KAS PÉTER
Budapest

MAYER JÁNOS
Budapest

A dolgozatban nemlineáris célfüggvényű hálózati folyamproblémák egy megoldási módjával foglalkozunk, a feladatot konkáv célfüggvény maximalizálására fogalmazva. A dolgozat 5 részre tagozódik. A bevezetés után a második részben megfogalmazzuk a feladatot és rögzítjük a jelöléseket. A harmadik rész az alakalmazott algoritmus, a redukált gradiens módszer áttekintését tartalmazza. A negyedik részben megadjuk az algoritmus elemeinek hálózati megfelelőit. Végül az ötödik rész a javasolt algoritmus leírását adja.

1. Bevezetés

A nemlineáris hálózati folyamprobléma megoldási módszereiről az utóbbi időben számos dolgozat született. A szokásos hálózati feltételeken kívül további feltételek bevezetése jelentősen megnöveli a modellek alkalmazhatóságát. Ilyen például a *„multi-commodity"* probléma, melynek széles körű alkalmazásai vannak a kommunikációs és közlekedési hálózatoknál. A javasolt iteratív megoldási módszerek egy része az eredeti, standard nemlineáris folyamprobléma megoldását igényli minden iterációban. Ezért fontos kérdés e feladat hatékony kezelése. Ebben a dolgozatban egy, a redukált gradiens módszerre épülő algoritmust ismertetünk. Célunk az algoritmus olyan megfogalmazása, amely hálózatos fogalmakkal dolgozik és standard hálózati technikát használ.

Bár a redukált gradiens módszer alkalmazásainál általában a degeneráció lexikografikus kezelésére nincs szükség, a folyamprobléma speciális feltételrendszere indokolttá teszi lexikográfia bevezetését. Például a hálózatot telítő folyam esetén minden bázis degenerált.

A dolgozatban tárgyalt módszerhez hasonló eljárást ismertet S. NGUYEN [5], [6] dolgozataiban. Az általunk javasolt algoritmus a használt normálást tekintve több változatot tárgyal, eltérő a degeneráció kezelése, valamint kapacitáskorlátokat is figyelembe vesz.

A módszer hatékonyságát a felhasznált hálózati technikára vonatkozó eddigi tapasztalatok valószínűsítik, valamint alátámasztja az is, hogy NGUYEN a saját módszerével kapcsolatban nagyon jó számítógépes tapasztalatokról számol be. A számítógépes program elkészítése folyamatban van. Mivel kisméretű hálózatokra vonatkozó tapasztalatok semmitmondóak, a programnak többszáz csúcs kezelését kell megoldani.

2. A probléma megfogalmazása, jelölések

Tekintsünk egy $[\mathcal{N}, \mathcal{A}, \mathbf{q}]$ hálózatot, ahol $\mathcal{N}$ a csúcsok halmaza $\{p_1, \ldots, p_n\}$, $\mathcal{A}$ az irányított élek halmaza $\{e_1, \ldots, e_m\}$, $\mathbf{q} \in R^m$ komponensei az élfolyamokra kirótt kapacitások, $\mathbf{q} \geqq \mathbf{0}$. Legyen $\mathbf{b} \in R^n$ a folyamértéket definiáló vektor, $\sum_{i=1}^{m} b_i = 0$.

Tegyük fel, hogy $[\mathcal{N}, \mathcal{A}, \mathbf{q}]$ hálózatban létezik $\mathbf{b}$ folyamértékhez tartozó folyam.

A feladat olyan $\mathbf{x} \in R^m$ folyam meghatározása, amely egy $f: R^m \to R^1$ konkáv célfüggvényt maximalizál. A fenti feltevések mellett a maximum mindig eléretik. A probléma nemlineáris programozási megfogalmazása a következő:

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

$$\mathbf{A}\mathbf{x} = \mathbf{b} \tag{2.1}$$

$$\mathbf{0} \leqq \mathbf{x} \leqq \mathbf{q}$$

ahol $\mathbf{A}$ az $[\mathcal{N}, \mathcal{A}, \mathbf{q}]$ hálózat által definiált $n \times m$ méretű pont-él incidenciamátrix.

A (2.1) probléma lineáris feltételekkel korlátozott konvex programozási feladat. Mivel a szimplex tábla mennyiségei hálózatos fogalmakkal (fa, vágás, hurok) könnyen reprezentálhatók, kézenfekvő, hogy (2.1) megoldására a redukált gradiens módszert [8] alkalmazzuk.

3. A redukált gradiens (RG) módszer rövid összefoglalása

E részben célunk az RG-módszer összefoglalása, az iránykereső feladat olyan tárgyalása, mely a hálózati fogalmakkal történő reprezentációt elősegíti. Lineáris feltételekkel korlátozott esetben a redukált gradiens algoritmus a megengedett irányokkal dolgozó módszerek egyikének tekinthető. Az RG-módszer specialitása az explicite megoldható iránykereső feladat. Ezért az előnyért általában a szimplex tábla (teljes rangú $\mathbf{A}$ mátrix esetén a bázisinverz) tárolása és iterációnkénti módosítása az ár. Az RG-módszer egy iterációja lényegében három fő lépésből áll:

(i) Írjuk fel az iránykereső feladatot jelenlegi megengedett pontunkban, majd transzformáljuk „könnyen" megoldhatóvá. (Ez utóbbi mondat jelentését részletesen tárgyaljuk.)
(ii) Oldjuk meg az iránykereső feladatot, más szóval határozzunk meg egy megengedett irányt.
(iii) Maximalizáljuk a célfüggvényt a megfelelő félegyenes és a megengedett tartomány metszetén.

A kapott pont szintén megengedett lesz, a hozzá tartozó célfüggvényérték nagyobb lesz, mint az előző pontban. A dolgozatban nem foglalkozunk az egyenes menti maximalizálás kérdésével, mert ebben a lépésben a lineáris feltételek speciális struktúrája nem használható ki.

Tekintsük a megengedett irány meghatározásának problémáját. E részben, bár a 2. pont jelöléseit használjuk, nem térünk ki az **A** mátrix struktúrájából adódó előnyökre.

Legyen $\mathbf{x} \in R^m$ megengedett pont és **B** az **A** oszlopainak egy bázisa. Particionáljuk **A** oszlopait és az iránykereső feladatban fellépő vektorok komponenseit a **B** bázisnak megfelelően. Ez azt jelenti, hogy megfelelő átrendezés után:

$\mathbf{A}=(\mathbf{B}, \mathbf{N})$, $\mathbf{x}=(\mathbf{y}, \mathbf{z})$, $\mathbf{q}=(\mathbf{s}, \mathbf{t})$ és a megengedett irányokra $\mathbf{w}=(\mathbf{u}, \mathbf{v})$.

A (2.1) feladat feltételei ilyen jelöléssel a következőképpen fogalmazhatók át:

$$\begin{aligned} \mathbf{By}+\mathbf{Nz} &= \mathbf{b} \\ \mathbf{0} \leqq \mathbf{y} &\leqq \mathbf{s} \\ \mathbf{0} \leqq \mathbf{z} &\leqq \mathbf{t}. \end{aligned} \tag{3.1}$$

Ekkor az **x** megengedett pontban a **B** bázisra vonatkozó iránykereső feladat az alábbi:

$$\begin{aligned} &\max (\nabla_y f \cdot \mathbf{u} + \nabla_z f \cdot \mathbf{v}) \\ &\mathbf{Bu} + \mathbf{Nv} = \mathbf{0} \\ &u_k \geqq 0, \quad \text{ha} \quad y_k = 0, \\ &\qquad\qquad\qquad\qquad k = 1, \ldots, \varrho(\mathbf{A}) \\ &u_k \leqq 0, \quad \text{ha} \quad y_k = s_k, \\ &v_l \geqq 0, \quad \text{ha} \quad z_l = 0, \\ &\qquad\qquad\qquad\qquad l = 1, \ldots, m-\varrho(\mathbf{A}) \\ &v_l \leqq 0, \quad \text{ha} \quad z_l = t_l, \\ &\qquad \|\mathbf{v}\| \leqq 1, \end{aligned} \tag{3.2}$$

ahol $\varrho(\mathbf{A})$ az **A** mátrix rangja, $\|\mathbf{v}\|$ tetszőleges normát jelent $R^{m-\varrho(\mathbf{A})}$-ban.

Megjegyezzük, hogy a (3.2) feladat kitűzésében a **B** bázis megválasztásáról még semmit sem mondunk. Célunk olyan **B** bázis megkeresése, melyre vonatkozó (3.2) feladat egyszerűen megoldható. Tekintsük az alábbi (3.3) feladatot, melyet a (3.2) iránykereső feladat redukáltjának fogunk nevezni:

$$\begin{aligned} &\max (\nabla_y f \cdot \mathbf{u} \qquad + \nabla_z f \cdot \mathbf{v}) \\ &\mathbf{Bu} + \qquad\qquad \mathbf{Nv} = \mathbf{0} \\ &v_k \geqq 0, \quad \text{ha} \quad z_k = 0, \\ &\qquad\qquad\qquad\qquad k = 1, \ldots, m-\varrho(\mathbf{A}) \\ &v_k \leqq 0, \quad \text{ha} \quad z_k = t_k, \\ &\qquad \|\mathbf{v}\| \leqq 1. \end{aligned} \tag{3.3}$$

Feladatunk olyan **B** bázis keresése, amellyel felírt (3.3) redukált iránykereső feladat optimális megoldása egyben optimális megoldása a (3.2) feladatnak is. Ha létezik nemdegenerált **B** bázis, azaz $\mathbf{0}<\mathbf{y}<\mathbf{s}$, akkor a (3.2) és (3.3) feladatok nyilván ekvivalensek. Mielőtt rátérnénk a **B** bázis alkalmas megválasztására, írjuk fel (3.3) optimális megoldását. A (3.3) feladat — a $\|\mathbf{v}\| \leqq 1$ feltételtől eltekintve —

lineáris programozási feladat, melyből **u** eliminálható. Az elimináció végrehajtása után kapott feladat:

$$\begin{aligned} &\max \mathbf{r}'\mathbf{v} \\ &v_k \geqq 0, \quad \text{ha} \quad z_k = 0, \\ &v_k \leqq 0, \quad \text{ha} \quad z_k = t_k, \qquad k = 1, \ldots, m-\varrho(\mathbf{A}) \\ &\|\mathbf{v}\| \leqq 1, \end{aligned} \tag{3.4}$$

ahol $r_k = \dfrac{\partial f}{\partial z_k} - \nabla_y f \cdot \mathbf{d}_k$, $k=1, \ldots, m-\varrho(\mathbf{A})$ és $\mathbf{d}_k$ a **B** bázissal felírt szimplex tábla k-adik oszlopa. Az **r** vektort redukált gradiensnek nevezzük. A (3.4) feladat optimális megoldását három különböző normában vizsgáljuk.

Legyen $I=\{i: z_i=0 \text{ és } r_i<0 \text{ vagy } z_i=t_i \text{ és } r_i>0\}$.

a) $$\|\mathbf{v}\|_1 = \max_{1 \leqq i \leqq m-\varrho(\mathbf{A})} |v_i|.$$

(3.4) optimális megoldása:

$$v_i = \begin{cases} \operatorname{sign} r_i, & i \notin I \\ 0, & i \in I \end{cases}$$

b) $$\|\mathbf{v}\|_2 = \sum_{i=1}^{m-\varrho(\mathbf{A})} |v_i|.$$

(3.4) optimális megoldása:

$$v_i = \begin{cases} \operatorname{sign} r_i, & i = i_0 \\ 0, & \text{máskor.} \end{cases}$$

Az i_0 indexet $|r_{i_0}| = \max\limits_{i \notin I} |r_i|$ definiálja

c) $$\|\mathbf{v}\|_3 = \left(\sum_{i=1}^{m-\varrho(\mathbf{A})} v_i^2 \right)^{\frac{1}{2}}.$$

(3.4) optimális megoldása

$$\hat{v}_i = \begin{cases} r_i, & i \notin I \\ 0, & \text{máskor} \end{cases}$$

és

$$\mathbf{v} = \frac{1}{\|\hat{\mathbf{v}}\|_3} \hat{\mathbf{v}}.$$

3.1. Megjegyzés. Vegyük észre, hogy a (3.4) feladat optimális megoldásaként adódó **v** komponenseire $\operatorname{sign} r_i = \operatorname{sign} v_i$, $i=1, \ldots, m-\varrho(\mathbf{A})$ és $v_i=0$, ha $i \in I$ mindhárom normában teljesül.

A (3.4) feladat optimális megoldásának meghatározása után két esetet különböztetünk meg.

1. eset. (3.4) optimális megoldásában a célfüggvény értéke 0. Ebben az esetben **x** optimális megoldása a (2.1) feladatnak.

2. eset. (3.4) optimális megoldásában a célfüggvény értéke pozitív. Ebben az esetben a $\mathbf{Bu}+\mathbf{Nv}=\mathbf{0}$ egyenletrendszerből $\mathbf{u}$ komponenseit kiszámítva $\mathbf{w}=(\mathbf{u}, \mathbf{v})$ megengedett irány, az $f(\mathbf{x})$ célfüggvény $\mathbf{w}$ irányban lokálisan növelhető.

Az egyetlen fennmaradó kérdés a $\mathbf{B}$ bázis alkalmas megválasztása a degenerált esetben. Az RG-módszer lexikografikus technikával való kiegészítését először KLEINMICHEL és SADOWSKI javasolta [3] dolgozatában.

Az egyszerű tárgyalásmód kedvéért tegyük fel a továbbiakban, hogy $\mathbf{q}=+\infty$ a (2.1) problémában, azaz az élfolyamok nincsenek felülről korlátozva. Definiáljuk a következő indexhalmazokat:

$$I_R=\{i\colon x_i>0\},\quad I_S=\{i\colon x_i=0\}$$

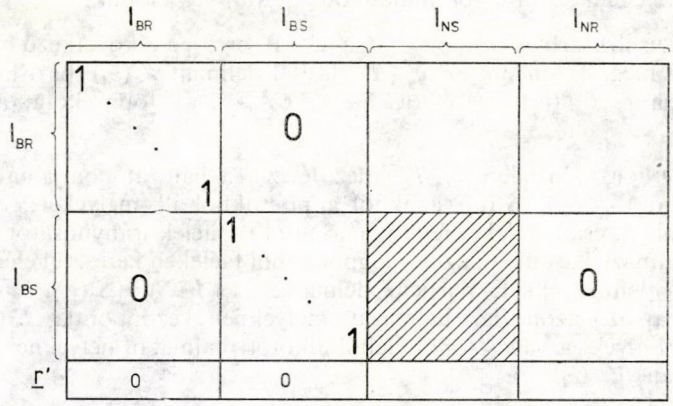
$$I_B=\{i\colon \mathbf{a}_i\in\mathbf{B},\ \mathbf{a}_i \text{ az } \mathbf{A} \text{ mátrix } i\text{-edik oszlopa}\}$$

$$I_N=\{1,2,\ldots,m\}\setminus I_B$$

$$I_{BR}=I_B\cap I_R\quad I_{BS}=I_B\cap I_S$$

$$I_{NR}=I_N\cap I_R\quad I_{NS}=I_N\cap I_S$$

A $\mathbf{B}$ bázis degeneráltsága azt jelenti, hogy $I_{BS}\neq\emptyset$. Első lépésként végezzünk bázis-cseréket mindaddig, míg $|I_{BR}|$ tovább már nem növelhető. Írjuk fel a szimplex táblát arra a $\mathbf{B}$ bázisra vonatkozóan, melynél $\max\limits_{(\mathbf{B})}|I_{BR}|$ eléretik.



1. ábra

Ezen a táblán végezzünk lexikografikus pivot lépéseket I_{BS} és I_{NS} elemei között, míg $r_k>0$, $k\in I_{NS}$ feltétel fennállásából $d_{ik}\leqq 0$, $i\in I_{BS}$ feltétel fennállása következik. (A d_{ik} számok a szimplex tábla megfelelő elemeit jelentik.) A lexikografikus pivot eljárás részletes tárgyalását az olvasó megtalálja PRÉKOPA [7] könyvében.

A lexikografikus pivotálás eredményeképpen kapott $\mathbf{B}$ bázissal írjuk fel a megfelelő (3.2) iránykereső feladatot és belőle származtatott (3.3), illetve (3.4) feladatokat. A (3.4) feladat optimális megoldásaként adódó $\mathbf{v}$ vektor segítségével megoldva a $\mathbf{Bu}+\mathbf{Nv}=\mathbf{0}$ egyenletrendszert a kapott $\mathbf{u}$ vektor komponensei — a 3.1. megjegyzés szerint — automatikusan kielégítik az $u_i\geqq 0$, $i\in I_{BS}$ feltételeket. Ezért ezen $\mathbf{w}=(\mathbf{u}, \mathbf{v})$ irány egyúttal a (3.2) feladat optimális megoldása is.

A redukált gradiens módszer a most vázolt formában általában nem konvergens, felléphet az ún. „cikcakk" jelenség. A konvergenciát az aktív feltételek körültekintőbb kezelésével, pl. ε_k-technika használatával lehet biztosítani. A konvergencia bizonyítása a [3] és [4] dolgozatokban található. Az ε_k-technika bevezetése az algoritmus hálózati interpretációját nem érinti.

4. A hálózati adaptációt előkészítő megjegyzések

E részben bizonyítás nélkül felsoroljuk az LP és a hálózati fogalmak közötti alapvető kapcsolatokat.

(i) Az **A** pont-él incidencia-mátrix rangja $n-1$.
(ii) Az **A** mátrix oszlopvektorainak bázisai és $[\mathcal{N}, \mathcal{A}, \mathbf{q}]$ hálózat feszítő fái egy-egyértelműen megfeleltethetők egymásnak.

Rögzítsünk egy **B** bázist és a megfelelő T feszítő fát.

4.1. DEFINÍCIÓ. Minden $e_j \notin T$ élhez létezik egy és csak egy hurok az $\{e_j\} \cup T$ élhalmazban, melynek éleit $C(e_j)$-vel jelöljük és az e_j élhez tartozó báziskörnek nevezzük. Minden báziskörhöz rendelhetünk egy irányítást az $e_j \notin T$ él irányításának megfelelően. A báziskör éleinek e_j-vel egyező irányba mutató éleinek halmazát $C^+(e_j)$-vel, az ellenkező irányba mutatókat $C^-(e_j)$-vel jelöljük.

(iii) A **B** bázishoz tartozó szimplex tábla $\mathbf{a}_j \notin \mathbf{B}$ oszlopa a következő módon konstruálható meg. Tekintsük az $e_j \notin T$ él által definiált $C(e_j)$ báziskört és legyen $d_{ij}=1$, ha $e_i \in C^-(e_j)$, $d_{ij}=-1$, ha $e_i \in C^+(e_j)$, $\mathbf{a}_j$ többi komponense legyen zéró.

4.2. DEFINÍCIÓ: Minden $e_i \in T$ élhez létezik a hálózat pontjainak egyértelmű partíciója $S'(e_i)$, $S''(e_i)$. $S'(e_i)$ azokból a pontokból áll, melyeket e_i kezdőpontjából érhetünk el, csak faéleket felhasználó úttal (a faélek irányításától eltekintünk). Az $S''(e_i)$ halmazt hasonlóan az e_i végpontjából faéleken keresztül elérhető pontok halmazaként definiáljuk. E partíció definiálja a $V(e_i)=(S'(e_i), S''(e_i))$ vágást. A vágásba tartozó azon élek halmazát, melyeknek kezdőpontja $S'(e_i)$-ben van, jelöljük $V^+(e_i)$-vel, a vágás többi éléből alkotott halmazt (melyeknek kezdőpontja $S''(e_i)$-ben van) $V^-(e_i)$-vel.

(iv) A **B** bázisra vonatkozó szimplex tábla i-edik sorát — mely az $\mathbf{a}_i \in \mathbf{B}$ vektornak felel meg — a következőképpen konstruálhatjuk meg: Tekintsük az $e_i \in T$ él által definiált $V(e_i)$ vágást, legyen $d_{ij}=+1$, ha $e_j \in V^+(e_i)$ és $d_{ij}=-1$, ha $e_j \in V^-(e_i)$, a sor többi eleme legyen zéró.
(v) Az **r** vektor komponensei a következőképpen számolhatók:
Legyen $e_j \notin T$ és tekintsük a $C(e_j)$ báziskört, akkor

$$r_j = \frac{\partial f}{\partial x_j} - \sum_{e_k \in C^-(e_j)} \frac{\partial f}{\partial x_k} + \sum_{e_k \in C^+(e_j)} \frac{\partial f}{\partial x_k}.$$

5. Az RG-módszer hálózati interpretációja

Legyen $\mathbf{x}$ adott megengedett megoldása (2.1) feladatnak és T rögzített feszítő fa. Az iránykereső feladatban az élfolyamokra kirótt nem-pozitivitási feltételeket elkerülendő, cseréljük meg azon e_j élek irányítását, ahol $x_j=q_j$. (Ez formálisan az $x_j:=-x_j$ helyettesítés végrehajtását jelenti.) Az alábbiakban megfogalmazzuk az algoritmus egy iterációját.

1. lépés: Ha T nem-degenerált, folytassuk a 4. lépésnél.

2. lépés: Degenerált esetben keressünk olyan T' feszítőfát, mely a lehető legtöbb nem-degenerált élet tartalmazza. Az eljárás a degenerált faélek által generált vágások megkeresését követeli. Ha ilyen vágás tartalmaz nem-degenerált élt, akkor vegyük be a feszítőfába és hagyjuk el a vágást definiáló faélt. Ha az eljárás végén kapott T' fa nem-degenerált, folytassuk a 4. lépésnél.

3. lépés: Alkalmazzunk lexikografikus technikát megfelelő feszítőfa megtalálására a $T=T'$ fából kiindulva. Számozzuk meg a hálózat éleit 1-től m-ig, a feszítőfa éleit előre véve. Számítsuk ki az $\mathbf{r}$ vektor komponenseit. Ha $r_j>0$ és $C^-(e_j)\neq\emptyset$, akkor generáljuk a $V(e_i)$ vágásokat minden $e_i\in C^-(e_j)$ esetén. Mindegyik $V(e_i)$ vágás az élekhez rendelt számok egy nagyság szerint rendezett sorozatával jellemezhető. Ha ezen sorozatok lexikografikus minimuma $V(e_i)$-nél éretik el, akkor cseréljük ki az $e_i\in T$ élt az $e_j\notin T$ éllel. Így természetesen újra feszítőfát kapunk. Ismételjük az eljárást, amíg $r_j>0$ feltétel fennállásából következik $C^-(e_j)=\emptyset$ minden $e_j\notin T$ esetén.

4. lépés: Az iránykereső-feladat megoldása. Induljunk ki az azonosan zéró cirkulációból, határozzuk meg $\mathbf{r}$ komponenseit. Az iránykereső feladatot optimalizáló cirkulációt a három különböző normában a következőképpen oldjuk meg.

A leírásban 3. rész jelöléseit használjuk.

a) $\|\cdot\|_1$ norma.
Tekintsük a nem-fa éleket és a redukált gradiens komponenseit sorra egymás után. Ha $e_j\notin T$, akkor az optimális v_j előjelétől függően e_j irányában vagy ellenkező irányban küldjünk egységnyi folyamot a $C(e_j)$ báziskör mentén, illetve ne változtassunk, ha $j\in I$.

b) $\|\cdot\|_2$ norma.
Ez esetben az iránykereső feladat optimális megoldását szolgáltató cirkulációban egyetlen $C(e_{j_0})$ báziskör mentén lesz csak nem-zéró a folyam. E báziskört a 3. fejezetben leírt módon választjuk ki és egységnyi folyamot küldünk $C(e_{j_0})$ báziskör mentén r_{j_0} előjele által megszabott irányban.

c) $\|\cdot\|_3$ norma.
Az *a)* ponthoz képest a különbség csak annyi, hogy $\mathbf{r}$ komponenseinek aránya határozza meg a báziskörök mentén körbeküldött folyamok nagyságát.

Megjegyezzük, hogy eljárásunk egyúttal az optimális **u** vektor komponenseit is szolgáltatja, ezek éppen a feszítőfa élein folyó, a fenti eljárásból adódó élfolyamok lesznek.

Természetesen a kapott irány bizonyos komponenseit (-1)-gyel kell szorozni, ha a megfelelő él irányát az iránykereső feladat megoldása előtt megfordítottuk.

5. lépés: A (2.1) feladat célfüggvényét a kapott irány mentén maximalizáljuk az élek kapacitásainak figyelembevételével. Itt bármelyik nem-lineáris programozási módszer alkalmazható pl. az aranymetszés. Az $\|\cdot\|_1$ és $\|\cdot\|_3$ norma esetében valamennyi élfolyam módosulhat, míg $\|\cdot\|_2$ normában csak egy báziskör mentén módosul a megengedett folyam.

IRODALOM

[1] ASSAD, A. A., "Multicommodity network flows — A survey", *Networks* **8** (1978) 37—91.
[2] FORD, L. R. and FULKERSON, D. R., *Flows in Networks* (Princeton University Press, Princeton, New Jersey, 1963).
[3] KLEINMICHEL, H. und SADOWSKI, H., "Der verallgemeinerte RG-Algorithmus bei linearen Restriktionen, die Behandlung des Entartungsfalls und die Konvergenz des Verfahrens", *Beiträge zur Numerische Mathematik* **3** (1975) 37—55.
[4] MAYER, J., "A nonlinear programming algorithm for the solution of a stochastic programming model of A. Prékopa", *Survey of mathematical programming, ed. A. Prékopa*, Vol. 2, Akadémiai Kiadó, Budapest, 1979, 129—139.
[5] NGUYEN, S., "An algorithm for the traffic assignment problem", *Transp. Sci.* **8** (1974) 203—216.
[6] NGUYEN, S., "A unified approach to equilibrium methods for traffic assignments", *Traffic Equilibrium Methods ed. M. Florian* (Springer Verlag, New York, 1976) 148—182.
[7] PRÉKOPA, A., *Lineáris programozás, I.* (Bolyai János Matematikai Társulat, Budapest, 1968).
[8] WOLFE, P., *Methods of Nonlinear Programming, Recent advances in mathematical programming, ed. R. Graves and P. Wolfe* (New York, 1963) 67—86.

(Beérkezett: 1979. április 12.)

KAS PÉTER ÉS MAYER JÁNOS
MTA SZÁMÍTÁSTECHNIKAI ÉS AUTOMATIZÁLÁSI KUTATÓ INTÉZET
1502 BUDAPEST XI., KENDE U. 13—17.

AN ALGORITHM FOR THE SOLUTION OF THE NONLINEAR NETWORK FLOW PROBLEM

P. KAS and J. MAYER

In the present paper we outline the basic ideas of the application of the reduced gradient method to the single-commodity nonlinear network flow problem. In other words an interpretation of the method in graph-terms is given. The proposed algorithm differs from Nguyen's method in the way of handling degeneracy it works with 3 different norms, and capacity constraints are also incorporated.

Alkalmazott Matematikai Lapok 5 (1979) 165—175

A BIMÁTRIX JÁTÉK EGYENSÚLYI PONTJAINAK MEGKERESÉSÉRŐL

SOÓS ZSOLT
Budapest

Ebben a dolgozatban a bimátrix játék több egyensúlyi pontjának megkeresési lehetőségével foglalkozunk. Ezen cél érdekében az egyensúlyi pontok geometriai elhelyezkedését vizsgáljuk, majd MAJTHAY A. módszerét kiterjesztjük több egyensúlyi pontot megtaláló algoritmussá. A kapott módszer eredményeit V. AGGARWAL módszerével is összehasonlítjuk.

1. Bevezetés

Ebben a dolgozatban a bimátrix játék egyensúlyi pontjainak tulajdonságaival és megkeresésükkel foglalkozunk.

A 2. fejezetben az egyensúlyi pont és a komplementaritási probléma kapcsolatát mutatjuk meg. A következő fejezetben az egyensúlyi pontok elhelyezkedésének és egymáshoz való kapcsolódásának geometriai vizsgálatát adjuk. Ez a vizsgálat lesz az alapja a 4. fejezetben ismertetett *Majthay-féle lexikografikus módszer* [6], [8] végességére adott igen egyszerű bizonyításunknak. Az ezt követő fejezetben ezt a módszert felhasználva javasolunk egy eljárást több egyensúlyi pont megkeresésére, melyhez hasonlót M. J. TODD is javasolt [9]. Az utolsó fejezetben V. AGGARWAL egy módszerét [1] hasonlítjuk össze számítási eredmények alapján az általunk javasolt eljárással.

Ezúton szeretnék köszönetet mondani BERNAU HEINZnek és STRAZICKY BEÁTÁnak a dolgozat írása közben nyújtott értékes megjegyzéseikért.

2. Egyensúlyi pont megfeleltetése egy komplementaritási probléma megoldásának

Jelölje $\mathbf{e}_i$, $i=1, 2, \ldots, m$ az i-edik egységvektort E_m-ben, az m dimenziós euklideszi térben, továbbá $\hat{\mathbf{e}}_i$, $i=1, 2, \ldots, n$ az i-edik egységvektort E_n-ben. Legyenek az $\mathbf{e}$ és $\hat{\mathbf{e}}$ m, ill. n dimenziós vektorok a következőképpen definiálva

$$\mathbf{e} = \mathbf{e}_1 + \mathbf{e}_2 + \ldots + \mathbf{e}_m$$

és

$$\hat{\mathbf{e}} = \hat{\mathbf{e}}_1 + \hat{\mathbf{e}}_2 + \ldots + \hat{\mathbf{e}}_n.$$

A bimátrix játékot két játékos, A és B játssza. A-nak m darab, B-nek n darab tiszta stratégiája van. Az A játékos költségét jelölje g_{ij}, ha A az i-edik tiszta stratégiáját, B pedig a j-edik tiszta stratégiáját játssza. Ugyanezen játék esetén B költ-

ségét $\hat{g}_{ji}$ jelöli. Ezek az értékek két valós mátrixot, $\mathbf{G}$-t és $\hat{\mathbf{G}}$-t határoznak meg, amik egyben egyértelműen definiálják a bimátrix játékot. A játékosok a tiszta játékok keverékét játsszák, így a két játékos stratégiahalmaza a következő:

$$S = \{\mathbf{p} | \mathbf{p} \in E_m,\ \mathbf{p} \geqq \mathbf{0},\ \mathbf{e}^T\mathbf{p} = 1\},$$

ill.

$$\hat{S} = \{\hat{\mathbf{p}} | \hat{\mathbf{p}} \in E_n,\ \hat{\mathbf{p}} \geqq \mathbf{0},\ \hat{\mathbf{e}}^T\hat{\mathbf{p}} = 1\}.$$

Egy $\mathbf{p}_0$ stratégia esetén $\mathbf{p}_0$ i-edik komponense azt jelenti, hogy A milyen valószínűséggel játssza az i-edik tiszta játékot. A B játékos esetén $\hat{\mathbf{p}}_0$ i-edik komponensének hasonló a jelentése.

DEFINÍCIÓ. A $\mathbf{G}$ és $\hat{\mathbf{G}}$ mátrixok által meghatározott bimátrix játék *Nash-féle egyensúlyi pontja* a $(\mathbf{p}_0, \hat{\mathbf{p}}_0)$ pár, ha

$$\mathbf{p}_0^T \mathbf{G} \hat{\mathbf{p}}_0 \leqq \mathbf{p}^T \mathbf{G} \hat{\mathbf{p}}_0, \quad \text{minden } \mathbf{p} \in S \text{ esetén} \tag{2.1}$$

és

$$\hat{\mathbf{p}}_0^T \hat{\mathbf{G}} \mathbf{p}_0 \leqq \hat{\mathbf{p}}^T \hat{\mathbf{G}} \mathbf{p}_0, \quad \text{minden } \hat{\mathbf{p}} \in \hat{S} \text{ esetén.} \tag{2.2}$$

A továbbiakban az általánosság megszorítása nélkül feltehetjük, hogy $\mathbf{G}$ és $\hat{\mathbf{G}}$ elemei pozitívak. Ugyanis amennyiben ez nem áll fenn, akkor alkalmas nagy K konstans esetén a $\mathbf{G} + K\mathbf{e}\hat{\mathbf{e}}^T$ és $\hat{\mathbf{G}} + K\hat{\mathbf{e}}\mathbf{e}^T$ mátrixok elemei már pozitívak, és könnyen látható, hogy a két utóbbi mátrix által meghatározott mátrixjáték egyensúlyi pontjai megegyeznek az eredeti $\mathbf{G}$ és $\hat{\mathbf{G}}$ által meghatározott mátrixjáték egyensúlyi pontjaival.

Megmutatjuk, hogy minden egyensúlyi pontnak megfelel egy komplementaritási probléma megoldása. Ennek érdekében ekvivalens átalakításokat hajtunk végre az (2.1), (2.2) egyenlőtlenségeken.

Legyenek t és $\hat{t}$ a következőképpen definiálva:

$$t = \mathbf{p}_0^T \mathbf{G} \hat{\mathbf{p}}_0$$

és

$$\hat{t} = \hat{\mathbf{p}}_0^T \hat{\mathbf{G}} \mathbf{p}_0.$$

A $\mathbf{G}$ és $\hat{\mathbf{G}}$ elemeire tett feltevés miatt $t > 0$ és $\hat{t} > 0$. (2.1)-et és (2.2)-t $\mathbf{e}$-vel, illetve $\hat{\mathbf{e}}$-val szorozva balról és átrendezve kapjuk

$$\mathbf{G}\hat{\mathbf{p}}_0 - t\mathbf{e} \geqq \mathbf{0}, \tag{2.3}$$

$$\hat{\mathbf{G}}\mathbf{p}_0 - \hat{t}\hat{\mathbf{e}} \geqq \mathbf{0}. \tag{2.4}$$

Továbbá legyenek

$$\mathbf{z}_0 = \frac{1}{\hat{t}}\mathbf{p}_0 \quad \text{és} \quad \hat{\mathbf{z}}_0 = \frac{1}{t}\hat{\mathbf{p}}_0.$$

Ezeket behelyettesítve (2.3), (2.4)-be, azt kapjuk, hogy

$$t\mathbf{G}\hat{\mathbf{z}}_0 - t\mathbf{e} \geqq \mathbf{0},$$

$$\hat{t}\hat{\mathbf{G}}\mathbf{z}_0 - \hat{t}\hat{\mathbf{e}} \geqq \mathbf{0}.$$

Ebből pedig

$$\mathbf{G}\hat{\mathbf{z}}_0 - \mathbf{e} \geqq \mathbf{0}, \tag{2.5}$$

$$\hat{\mathbf{G}}\mathbf{z}_0 - \hat{\mathbf{e}} \geqq \mathbf{0}. \tag{2.6}$$

Egyszerű számolással adódnak az ún. komplementaritási feltételek:

$$\mathbf{z}_0^T(\mathbf{G}\hat{\mathbf{z}}_0 - \mathbf{e}) = 0, \tag{2.7}$$

$$\hat{\mathbf{z}}_0^T(\hat{\mathbf{G}}\mathbf{z}_0 - \hat{\mathbf{e}}) = 0, \tag{2.8}$$

és természetesen $\mathbf{z}_0 \geqq 0$, $\hat{\mathbf{z}}_0 \geqq \mathbf{0}$. Bebizonyítottuk a következő tételt:

TÉTEL: Egy-egy értelmű megfeleltetés létesíthető a bimátrix játék (2.1), (2.2) által definiált *Nash-féle egyensúlyi pontja* és a (2.5), (2.6), (2.7) és (2.8) által definiált komplementaritási probléma megoldása között.

Ha a komplementaritási probléma megoldható, akkor a

$$t = \frac{1}{\sum_{j=1}^{n} \hat{z}_j}, \quad \hat{t} = \frac{1}{\sum_{j=1}^{m} z_j}$$

választással a

$$\mathbf{p}_0 = \hat{t}\mathbf{z}_0, \quad \hat{\mathbf{p}}_0 = t\hat{\mathbf{z}}_0$$

pontpár a bimátrix játék egyensúlyi pontja lesz.

3. Geometriai vizsgálatok

Ebben a fejezetben C. E. LEMKE és J. T. HOWSON cikke [5] alapján megvizsgáljuk az egyensúlyi pontok halmazát, azok egymással való összefüggését és számosságukat. A 4. fejezetben ezek ismeretében adunk egy egészen egyszerű bizonyítást a *Majthay-féle módszer* végességére.

Kiindulunk az előző fejezet

$$\mathbf{G}\hat{\mathbf{z}} - \mathbf{e} \geqq \mathbf{0} \tag{3.1a}$$

$$\hat{\mathbf{z}} \geqq \mathbf{0} \tag{3.1b}$$

$$\hat{\mathbf{G}}\mathbf{z} - \hat{\mathbf{e}} \geqq \mathbf{0} \tag{3.2a}$$

$$\mathbf{z} \geqq \mathbf{0} \tag{3.2b}$$

egyenlőtlenség-rendszeréből. Látható, hogy a (3.1a), (3.1b) rendszer a következő $(m+n)$ darab lineáris egyenlőtlenségből áll:

$$\mathbf{g}_{i.}\hat{\mathbf{z}} - 1 \geqq 0, \quad i = 1, 2, \ldots, m, \tag{3.3}$$

$$\hat{\mathbf{e}}_i^T\hat{\mathbf{z}} \geqq 0, \quad i = 1, 2, \ldots, n, \tag{3.4}$$

ahol $\mathbf{g}_{i.}$ a $\mathbf{G}$ mátrix i-edik sora.

(3.1a) és (3.1b)-nek eleget tevő $\hat{\mathbf{z}}$ vektorok halmazát jelöljük $\hat{L}$-val. $\hat{L}$ határa a

$$\mathbf{g}_i.\hat{\mathbf{z}}-1=0, \quad i=1, \ldots, m, \tag{3.5}$$

$$\hat{\mathbf{e}}_i^T\hat{\mathbf{z}}=0, \quad i=1, \ldots, n \tag{3.6}$$

egyenlőségrendszer legalább egy egyenletének eleget tevő $\hat{\mathbf{z}}$ pontok. Minden $\hat{\mathbf{z}}\in\hat{L}$ ponthoz hozzárendelhetünk egy $\mathbf{M}(\hat{\mathbf{z}})$ mátrixot a következő módon. A $\mathbf{G}^T$ i-edik oszlopát bevesszük $\mathbf{M}(\hat{\mathbf{z}})$ oszlopai közé, ha (3.3) i-edik egyenlőtlensége egyenlőség formájában teljesül, hasonlóan $\hat{\mathbf{e}}_j$-t bevesszük $\mathbf{M}(\hat{\mathbf{z}})$ oszlopai közé, ha (3.4) j-edik egyenlőtlensége egyenlőség formájában teljesül. Ha $\mathbf{B}$ a $(\mathbf{G}^T, \hat{\mathbf{E}})$ mátrix bizonyos oszlopaiból összeállított mátrix, akkor legfeljebb egy olyan $\tilde{\mathbf{z}}\in\hat{L}$ létezik, hogy $\mathbf{B}=\mathbf{M}(\tilde{\mathbf{z}})$ és rang $\mathbf{B}=n$ és ezt a pontot $\hat{L}$ extremális pontjának fogjuk nevezni. A továbbiakban a következő nemdegeneráltsági feltétellel fogunk élni $\hat{L}$-ra vonatkozóan. Legyen $\overline{\mathbf{B}}$ egy $n\times r$-es mátrix, melynek oszlopai $(\mathbf{G}^T, \hat{\mathbf{E}})$ oszlopai közül valók. Ha létezik olyan $\hat{\mathbf{z}}_0$, hogy $\overline{\mathbf{B}}=\mathbf{M}(\hat{\mathbf{z}}_0)$, akkor rang $\overline{\mathbf{B}}=r$. A $\mathbf{G}^T$ a $\mathbf{G}$ mátrix transzponáltját jelöli.

Ezen nemdegeneráltsági feltétel mellett látható, hogy egy extremális pont n darab egyenlőségnek fog eleget tenni a (3.3), (3.4) rendszerben.

Adott $\hat{\mathbf{z}}_0$ extremális pont esetén $\mathbf{M}(\hat{\mathbf{z}}_0)$ oszlopait jelölje rendre $\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_n$. Egy lemmát mondunk ki a

$$\hat{\mathbf{z}}=\hat{\mathbf{z}}_0+\sum_{i=1}^{n} t_i\mathbf{d}^i$$

módon előállítható pontokra, aholy $\mathbf{d}^i$ az $\mathbf{M}^T$ mátrix inverzének i-edik oszlopvektora.

3.1. LEMMA: Ha $\hat{\mathbf{z}}_0$ extremális pontja $\hat{L}$-nak, akkor létezik $k>0$ szám úgy, hogy minden $\mathbf{t}\in\left\{\mathbf{t}\in E_n | t_i\geqq 0, \sum_{i=1}^{n} t_i^2\leqq k\right\}$ esetén a $\hat{\mathbf{z}}=\hat{\mathbf{z}}_0+\sum_{i=1}^{n} t_i\mathbf{d}^i$ pont eleme $\hat{L}$-nak.

Ezen lemma következményei a következők:

1. Ha $\hat{\mathbf{z}}_0$ egy extremális pont és $\hat{\mathbf{z}}=\hat{\mathbf{z}}_0+t_i\mathbf{d}^i$, $t_i\geqq 0$ és eléggé kicsi, akkor $\mathbf{M}(\hat{\mathbf{z}})$ oszlopait $\mathbf{M}(\hat{\mathbf{z}}_0)$ oszlopaiból az i-edik törlésével kapjuk. Ezt $\hat{L}$ egy élének nevezzük, melynek $\hat{\mathbf{z}}_0$ a végpontja.

2. Pontosan n darab nemkorlátos éle van $\hat{L}$-nak. Ezek a $\hat{\mathbf{z}}=k\hat{\mathbf{e}}_i$ alakúak, ahol $k>0$ és eléggé nagy. Ez abból következik, hogy $\mathbf{G}$ elemei pozitívak.

3. Ha $\hat{\mathbf{z}}_1$ olyan pont, hogy $\mathbf{M}(\hat{\mathbf{z}}_1)$ rangja $(n-1)$, akkor $\hat{\mathbf{z}}_1$ $\hat{L}$ egy élén fekszik. Ezen az élen fekvő pontok, melyek $\hat{\mathbf{z}}_1+t_i\mathbf{d}^i$ alakúak, $t_i\geqq 0$ $|t_i|\leqq K$, ugyanazon $\mathbf{M}(\hat{\mathbf{z}})$ mátrixszal rendelkeznek ($t_i\in\{1, 2, \ldots, n\}$).

Így két extremális pont szomszédos, ha a nekik megfelelő $\mathbf{M}$ mátrixok csak egy oszlopban különböznek.

Jelölje L a (3.2a), (3.2b) $(m+n)$ darab egyenlőtlenségnek eleget tevő $\mathbf{z}$ pontok halmazát. L és $\hat{L}$ struktúrája teljesen hasonló, így L esetén is hasonló definíciókat, állításokat és megjegyzéseket tehetünk, mint $\hat{L}$ esetén.

Legyen X az $\hat{L}$ és L direkt szorzata. Egy $\mathbf{x}=(\hat{\mathbf{z}}, \mathbf{z})$ pontot extremálisnak nevezünk, akkor és csak akkor, ha $\hat{\mathbf{z}}$ extremális pontja $\hat{L}$-nak és $\hat{\mathbf{z}}$ extremális pontja L-nek, továbbá azt mondjuk, hogy $\mathbf{x}$ egy élen fekszik, akkor és csak akkor, ha $\hat{\mathbf{z}}$ extremális pontja $\hat{L}$-nak és $\mathbf{z}$ egy élen fekszik, vagy fordítva.

Egy $\mathbf{x}=(\hat{\mathbf{z}}, \mathbf{z})\in X$ pontot egyensúlyi pontnak nevezünk, ha

$$(\mathbf{e}_i^T\mathbf{z})(\mathbf{g}_{i.}\hat{\mathbf{z}}-1)=0, \quad i=1, 2, \ldots, m,$$

$$(\hat{\mathbf{e}}_i^T\hat{\mathbf{z}})(\hat{\mathbf{g}}_{i.}\mathbf{z}-1)=0, \quad i=1, 2, \ldots, n.$$

Ezeket egyensúlyi vagy komplementaritási feltételeknek nevezzük.

3.1. TÉTEL: Egy nemdegenerált probléma egyensúlyi pontja extremális pontja X-nek.

Bizonyítás. Megtalálható [5]-ben.

3.1. DEFINÍCIÓ. Rögzített r valós szám esetén S_r jelölje az X azon pontjait, amelyek az $(m+n)$ komplementaritási feltétel közül legfeljebb az r-ediket nem elégítik ki.

3.2. TÉTEL. S_r pontja vagy extremális pontja X-nek, vagy X egy élén fekszik.

Bizonyítás. Megtalálható [5]-ben.

3.3. TÉTEL. Létezik X-nek pontosan egy nemkorlátos éle, amely S_r pontjaiból áll elő.

Ennek a tételnek a bizonyítása szintén megtalálható [5]-ben, de a későbbi vizsgálatainkban a bizonyítás konstrukciójára szükségünk lesz, így ezt itt is közöljük.

Bizonyítás: Legyen $\mathbf{z}=k\mathbf{e}_r$. Elég nagy k-ra $\mathbf{z}\in L$. Ha k_0 jelöli a legkisebb ilyen k-t, akkor $\mathbf{z}_0=k_0\mathbf{e}_r$ L extremális pontja, és létezik egy s index, hogy $\hat{\mathbf{g}}_s\mathbf{z}-1=0$. Legyen továbbá $\hat{\mathbf{z}}=k\hat{\mathbf{e}}_s$. Elég nagy k-ra $\hat{\mathbf{z}}\in\hat{L}$. Ha k_1 jelöli a legkisebb ilyen k-t, akkor $\hat{\mathbf{z}}_0=k_1\hat{\mathbf{e}}_s$ $\hat{L}$ egy extremális pontja. A $(\hat{\mathbf{z}}, \mathbf{z}_0)$ pontok $k>k_1$-re X nemkorlátos élét adják, és ezen él végpontja a $(\hat{\mathbf{z}}_0, \mathbf{z}_0)\in X$ pont. Látható, hogy a nemkorlátos él pontjai S_r-beliek.

A 3.1. lemma után tett 2. megjegyzés miatt X-nek pontosan $m+n$ darab nemkorlátos éle van. Az előbbiek miatt minden S_r, $r=1, 2, \ldots, m+n$ tartalmaz egy ilyen élet. Mivel X egy éle nem fekhet egyszerre S_{r_1}-ben és S_{r_2}-ben, $r_1\neq r_2$, így minden S_r pontosan egy nemkorlátos élet tartalmaz.

A következő tételt, mely a *Majthay-féle módszernek* is alapja, újra bizonyítás nélkül közöljük.

3.4. TÉTEL. Legyen $\mathbf{x}$ extremális pontja X-nek, és pontja S_r-nek is. Ekkor egy vagy két S_r-beli éle van X-nek, amely(ek)nek $\mathbf{x}$ a végpontja(uk). $\mathbf{x}$ egyensúlyi pont akkor és csak akkor, ha csak egy ilyen él létezik.

Két élet X-ben szomszédosnak nevezünk, ha közös a végpontjuk. S_r-beli szomszédos élek rendszerét r-útnak nevezzük.

Az utolsó tételből következik, hogy S_r-beli extremális pontból indulva r-úton vagy a kiinduló extremális ponthoz jutunk (zárt r-út), vagy egy egyensúlyi ponthoz, vagy egy nemkorlátos élre. Ezek után a következő tételt mondhatjuk ki.

3.5. TÉTEL. S_r nem üres halmaz. S véges sok diszjunkt r-út uniója. Minden egyes r-út vagy zárt r-út, és ekkor nem tartalmaz egyensúlyi pontot; vagy egy egyensúlyi pontot tartalmaz, vagy kettőt. Az egyensúlyi pontok száma páratlan.

4. A Majthay-féle algoritmus

Ebben a fejezetben ismertetünk egy módszert [6], a bimátrix játék egyensúlyi pontjának megtalálására. Először röviden vázoljuk az algoritmust, majd egy geometriai megvilágítását közöljük. Közben egy egyszerű bizonyítást is közlünk a módszer végességére.

Ez a módszer lesz az alapja annak a módszernek, mely több egyensúlyi pont megtalálására alkalmas. Ezt az algoritmust a következő fejezetben ismertetjük.

Könnyen látható, hogy a (2.5), (2.6) egyenletrendszerek az $\mathbf{u} \in E_m$, és $\hat{\mathbf{u}} \in E_n$ nemnegatív vektorok bevezetésével a következő alakban írhatók fel:

$$\mathbf{u} - \mathbf{G}\hat{\mathbf{z}} = -\mathbf{e} \tag{4.1}$$

$$\hat{\mathbf{u}} - \hat{\mathbf{G}}\mathbf{z} = -\hat{\mathbf{e}} \tag{4.2}$$

$$\mathbf{u} \geqq \mathbf{0}, \quad \hat{\mathbf{u}} \geqq \mathbf{0}, \quad \mathbf{z} \geqq \mathbf{0}, \quad \hat{\mathbf{z}} \geqq \mathbf{0}. \tag{4.3}$$

Vezessük be a következő jelöléseket:

$$\mathbf{A} = (\mathbf{E}, -\mathbf{G}), \quad \hat{\mathbf{A}} = (\hat{\mathbf{E}}, -\hat{\mathbf{G}})$$

$$\mathbf{x} = \begin{pmatrix} \mathbf{u} \\ \hat{\mathbf{z}} \end{pmatrix}; \quad \hat{\mathbf{x}} = \begin{pmatrix} \hat{\mathbf{u}} \\ \mathbf{z} \end{pmatrix}.$$

Így a (4.1), (4.2), (4.3) rendszer az

$$\mathbf{A}\mathbf{x} = -\mathbf{e}$$

$$\mathbf{x} \geqq \mathbf{0}$$

$$\hat{\mathbf{A}}\hat{\mathbf{x}} = -\hat{\mathbf{e}}$$

$$\hat{\mathbf{x}} \geqq \mathbf{0}$$

alakot ölti. Az $\mathbf{u}, \mathbf{z}$ és $\hat{\mathbf{u}}, \hat{\mathbf{z}}$ azonos indexű elemeit komplementer változóknak, míg a nekik megfelelő $\mathbf{A}$, illetve $\hat{\mathbf{A}}$-beli oszlopokat egymás komplementer vektorainak nevezzük.

Mivel $\mathbf{A}$ rangja m, $\hat{\mathbf{A}}$ rangja pedig n, így léteznek $\mathbf{B}$, illetve $\hat{\mathbf{B}}$ mátrixok, melyek $\mathbf{A}$ m darab, illetve $\hat{\mathbf{A}}$ n darab oszlopából állnak elő. Ezen $\mathbf{B}$ és $\hat{\mathbf{B}}$ mátrixokhoz léteznek egyértelműen $\mathbf{D}$, illetve $\hat{\mathbf{D}}$ mátrixok úgy, hogy igaz a következő két egyenlet:

$$\mathbf{B}\mathbf{D} = (-\mathbf{e}, \mathbf{A}),$$

$$\hat{\mathbf{B}}\hat{\mathbf{D}} = (-\hat{\mathbf{e}}, \hat{\mathbf{A}}).$$

A $(\mathbf{B}, \hat{\mathbf{B}})$ bázispárt komplementer párnak nevezzük, ha nem tartalmaznak komplementer vektorokat. Egy $\mathbf{x}$ vektort lexikografikusan pozitívnak nevezünk, ha $\mathbf{x}$ első nemnulla komponense pozitív. Az $\mathbf{x}_1$ vektor lexikográfikusan pozitívabb, mint $\mathbf{x}_2$, ha $\mathbf{x}_1 - \mathbf{x}_2$ vektor lexikografikusan pozitív, továbbiakban l-pozitív (jelben $\mathbf{x}_1 \succ \mathbf{x}_2$). A $(\mathbf{B}, \hat{\mathbf{B}})$ bázispárt l-megengedettnek nevezzük, ha a nekik megfelelő $\mathbf{D}$ és $\hat{\mathbf{D}}$ mátrix sorvektorai l-pozitívak. Nyilvánvalóan egy ilyen bázispárhoz tartozó bázismegoldás megengedett, hiszen $\mathbf{D}$ és $\hat{\mathbf{D}}$ első oszlopai nemnegatívak.

A továbbiakban ismertetjük az algoritmust.

1a. Legyenek $\mathbf{B}^{(-2)}=(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_m)$, $\hat{\mathbf{B}}^{(-2)}=(\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \ldots, \hat{\mathbf{e}}_n)$, $\mathbf{D}^{(-2)}=(-\mathbf{e}, \mathbf{A})$, $\hat{\mathbf{D}}^{(-2)}=(-\hat{\mathbf{e}}, \hat{\mathbf{A}})$. Látható, hogy a $(\mathbf{B}^{(-2)}, \hat{\mathbf{B}}^{(-2)})$ bázispár nem l-megengedett.

1b. Definiáljuk a $(\mathbf{B}^{(-1)}, \hat{\mathbf{B}}^{(-1)})$ párt a következőképpen. Legyen $k=m+1$ és $\mathbf{A}$ $(m+1)$-edik oszlopvektorát vonjuk be a $\mathbf{B}^{(-2)}$ bázisba. A kilépő vektor indexét a

$$\frac{1}{d_{h,k}^{(-2)}}\mathbf{d}_{h.}^{(-2)} = 1-\max_{i\in I^{(-2)}}\frac{1}{d_{i,k}^{(-2)}}\mathbf{d}_{i.}^{(-2)},$$

ahol $I^{(-2)}$ a $\mathbf{B}^{(-2)}$ bázis bázisindexeinek a halmaza.

A $\mathbf{D}$ transzformációs formulái a szimplex módszer transzformációs formulái. Mi PRÉKOPA A. [7] könyvének transzformációs formuláit használjuk.

Mivel $\mathbf{D}^{(-2)}$ sorai lineárisan függetlenek, így a kilépő vektor indexe egyértelmű. Vegyük észre, hogy $\mathbf{D}^{(-1)}$ sorai már lexikografikusan pozitívak. Legyen továbbá $\hat{\mathbf{B}}^{(-1)}=\hat{\mathbf{B}}^{(-2)}$.

1c. Legyen $\mathbf{B}^{(0)}=\mathbf{B}^{(-1)}$, $k=n+h$, és vonjuk be a $\hat{\mathbf{B}}^{(-1)}$ bázisba az $\hat{\mathbf{A}}$ k-adik oszlopát. A kilépő vektor j indexét az

$$\frac{1}{\hat{d}_{j,k}^{(-1)}}\hat{\mathbf{d}}_{j.}^{(-1)} = 1-\max_{i\in I^{(-1)}}\frac{1}{\hat{d}_{i,k}^{(-1)}}\hat{\mathbf{d}}_{i.}^{(-1)}$$

egyenlet határozza meg egyértelműen.

Láthatjuk, hogy a $(\mathbf{B}^{(0)}, \hat{\mathbf{B}}^{(0)})$ bázispár már megengedett bázispár.

2a. Ha $\hat{\mathbf{B}}^{(q)}=\hat{\mathbf{B}}^{(q-1)}$ és $\mathbf{D}^{(q-1)}$ j-edik oszlopa hagyta el a $\mathbf{B}^{(q-1)}$ bázist, akkor legyen $\mathbf{B}^{(q+1)}=\mathbf{B}^{(q)}$ és $\mathbf{d}_j^{(q)}$ komplementerpárját $\hat{\mathbf{d}}_k^{(q)}$-t bevonjuk a $\hat{\mathbf{B}}^{(q)}$ bázisba, míg a kilépő vektor j indexét az

$$\frac{1}{\hat{d}_{jk}^{(q)}}\hat{\mathbf{d}}_{j.}^{(q)} = 1-\min_{i\in I_+^{(q)}}\frac{1}{\hat{d}_{i,k}^{(q)}}\hat{\mathbf{d}}_{i.}^{(q)},$$

$\hat{I}_+^{(q)}=\{i\in\hat{I}^{(q)}, \hat{d}_{ik}^{(q)}>0\}$ egyenlet határozza meg.

2b. Ha $\mathbf{B}^{(q)}=\mathbf{B}^{(q-1)}$ és $\hat{\mathbf{D}}^{(q-1)}$ j-edik oszlopa hagyta el a $\hat{\mathbf{B}}^{(q-1)}$ bázist, akkor legyen $\hat{\mathbf{B}}^{(q+1)}=\hat{\mathbf{B}}^{(q)}$ és $\hat{\mathbf{d}}_j^{(q)}$ komplementerpárját a $\mathbf{d}_k^{(q)}$-t bevonjuk a $\mathbf{B}^{(q)}$ bázisba, míg a kilépő vektor j indexét az

$$\frac{1}{d_{jk}^{(q)}}\mathbf{d}_{j.}^{(q)} = 1-\min_{i\in I_+^{(q)}}\frac{1}{d_{ik}^{(q)}}\mathbf{d}_{i.}^{(q)},$$

$I_+^{(q)}=\{i\in I^{(q)}, d_{ik}^{(q)}>0\}$, egyenlet definiálja egyértelműen.

3. Ha $\mathbf{D}^{(q)}$ $(m+2)$-edik oszlopvektora vagy $\hat{\mathbf{D}}^{(q)}$ 2. oszlopvektora hagyja el az aktuális $(\mathbf{B}^{(q)}, \hat{\mathbf{B}}^{(q)})$ bázispárt, akkor olyan megengedett megoldást kaptunk, amely bázispárja komplementer bázispár. Ekkor módszerünket befejezzük. Ellenkező esetben a 2. lépés megfelelő részére lépünk.

Ezzel meghatároztuk a módszert. Vizsgáljuk meg egy kicsit más szempontból is.

Látható, hogy az első lépés egy nem megengedett pontból egy megengedett pontba vitt. Ez a pont ún. majdnem komplementaritási pont, mert a neki megfelelő $(\mathbf{B}^{(0)}, \hat{\mathbf{B}}^{(0)})$ bázispárban egy komplementer pár szerepel csak, nevezetesen az $(\mathbf{a}_{m+1}, \hat{\mathbf{a}}_1)$. A többi komplementerpárból csak az egyik szerepel, és létezik egy komplementer pár, melynek egyik tagja sincs a bázisban. Ez a pont az előző fejezetben definiált S_1-nek eleme.

Az előző fejezet 3.4. tételének értelmében S_1-ben maradva két irányba léphetünk. Ez a két irány a nembázisbeli komplementerpár elemeinek megfelelő irány. A módszer a továbblépőirányról egyértelműen gondoskodik, és így mindvégig S_1-beli pontban maradunk.

Az előző fejezet és az előzőkben tett vizsgálatok alapján MAJTHAY [6] cikkében szereplő 4., 5. tételt igen egyszerűen egybefoglalva bizonyíthatjuk.

4.1. TÉTEL. A fentiekben definiált módszer véges sok lépésben véget ér egy egyensúlyi pontban.

Bizonyítás. Az algoritmus első lépése a 3.3. tételt figyelembe véve, az S_1 egyetlen nemkorlátos élének végpontjába vezet. Az 3.5. tétel értelmében S_1 olyan útjára léptünk, melynek egyik befejezése S_1 egyetlen nemkorlátos éle, a másik egy egyensúlyi pont.

Az első lépés után nem a nemkorlátos élre lépünk, hanem a másik lehetséges irányba, ugyanis $\hat{\mathbf{d}}_j^{(0)} < \mathbf{0}$, így ebbe az irányba nem tudunk lépni az algoritmus megkötése miatt. Így tételünket bizonyítottuk.

5. Több egyensúlyi pont megtalálása

Ebben a fejezetben először V. AGGARWAL egy ötletét említjük [1], majd egy olyan algoritmust részletezünk, amelyhez hasonlót M. J. TODD is javasolt. Végül egy egyensúlyi pontunk „jó"-ságára vonatkozó egyszerű vizsgálati módot ismertetünk.

V. AGGARWAL a 3. fejezet 3.3. tételének ismeretében — mely szerint minden S_r tartalmaz nemkorlátos élű r-utat, melyen található pontosan egy egyensúlyi pont —, azt javasolta, hogy induljunk el minden S_r nemkorlátos élén, mert ekkor minden esetben egyensúlyi ponthoz jutunk. Az algoritmus problémái, hogy egy egyensúlyi pontot többször is megkaphatunk, és létezhet olyan egyensúlyi pont is, amelyhez így sem tudunk eljutni. Az utóbbira ő maga is mutatott példát.

Az általam javasolt módszernek az előbb felsorolt fogyatékosságai megvannak, viszont minél több pontot vizsgál, olyan egyensúlyi pontot is megtalálhat, melyet az első algoritmus nem. A továbbiakban ezt az algoritmust részletezzük.

Legyen $I = \{1, 2, \ldots, m+n\}$.

1. Legyenek $s = r = k = 1$.

2. Az S_r-en lépjünk az $\mathbf{E}_1^r$ egyensúlyi pontra a *Majthay-féle lexikográfikus pivot módszerrel.* Legyen $I_1 = I \setminus \{r\}$.

3. Ha $I_k \neq \emptyset$, akkor $j \in I_k$-ra a lexikográfikus pivotálással induljunk el az $\mathbf{E}_k^r$ egyensúlyi pontból induló j-úton, és $I_k = I_k \setminus \{j\}$. Ellenkező esetben lépjünk az 5. lépésre.

4. Ha a pivotálás egyensúlyi ponton fejeződik be, akkor ha a kapott egyensúlyi pont jelzett, akkor lépjünk a 3. lépésre, ha nem jelzett, akkor jelöljük meg az $s=s+1$ indexszel és rendeljük hozzá az $I_s=I\setminus\{j\}$ indexhalmazt. Lépjünk a 3. lépésre.

Ha a pivotálás nemkorlátos élen fejeződik be, akkor lépjünk a 3. lépésre.

5. Legyen $k=k+1$. Ha létezik ilyen indexű egyensúlyi pont, akkor lépjünk a 3. lépésre, ellenkező esetben, ha $r<m+n$, akkor legyen $r=r+1$, és lépjünk a 2. lépésre; ha $r=m+n$, akkor eljárásunk befejeződik.

Néhány megjegyzést teszünk a fentiekben vázolt algoritmusra vonatkozóan.

Az s index a már elért, míg a k index a már leszámolt, illetve leszámlálás alatt levő egyensúlyi pontok számát jelöli. Az eljárás így nyilvánvalóan akkor fejeződik be, ha az $s-k$ különbség pozitívvá akar válni.

A 4. lépésben csak két eset lehet, holott a 3. fejezet 3.5. tétele olyan r-út létezését is kimondja, mely nem tartalmaz egyensúlyi pontot. Vegyük észre, hogy ez a mi esetünkben nem állhat fenn, mivel egyensúlyi pontból indulunk.

Ez a módszer az első lépéssorozatban ($r=1$) csak olyan egyensúlyi pontokat talál meg, melyek az $\mathbf{E}_1^r$ (S_1 nemkorlátos éléhez tartozó) egyensúlyi pontból majdnem komplementaritási éleken elérhetőek. Nevezzük ezt SZ_1 szigetnek. Amikor a 2. lépésben nem az S_1-ben indulunk el, akkor újabb szigeteket kaphatunk. Az így kapott SZ_i szigetek vagy azonosak, vagy diszjunktak.

A több, illetve összes egyensúlyi pont megtalálásának problémája akkor merül fel, amikor az egyensúlyi pontok közül valamilyen cél (f célfüggvény) szerint a legjobbat szeretnénk kiválasztani. Ha EGY jelöli a $\mathbf{G}$ és $\hat{\mathbf{G}}$ által meghatározott bimátrix játék egyensúlyi pontjait, akkor ez a feladat a következő:

$$\max f(\mathbf{p}, \hat{\mathbf{p}}).$$
$$(\mathbf{p}, \hat{\mathbf{p}}) \in EGY$$

Láttuk, hogy az egyensúlyi pontra vonatkozó feltételek megfelelő átalakítások után a következő alakot öltik:

$$(\mathbf{E}, -\mathbf{G})\begin{pmatrix}\mathbf{u}\\ \hat{\mathbf{z}}\end{pmatrix} = -\mathbf{e}, \tag{5.1a}$$

$$\begin{pmatrix}\mathbf{u}\\ \hat{\mathbf{z}}\end{pmatrix} \geqq \mathbf{0}, \tag{5.1b}$$

$$(\hat{\mathbf{E}}, -\hat{\mathbf{G}})\begin{pmatrix}\hat{\mathbf{u}}\\ \mathbf{z}\end{pmatrix} = -\hat{\mathbf{e}}, \tag{5.2a}$$

$$\begin{pmatrix}\hat{\mathbf{u}}\\ \mathbf{z}\end{pmatrix} \geqq \mathbf{0}. \tag{5.2b}$$

Ezen lineáris rendszer megoldása nem feltétlenül elégíti ki a komplementaritási feltételt is. Így ha a

$$\max \tilde{f}(\mathbf{z}, \hat{\mathbf{z}})$$

$(\mathbf{z}, \hat{\mathbf{z}})$ eleget tesz az (5.1a), (5.1b), (5.2a), (5.2b) lineáris rendszernek; matematikai programozási feladatot megoldjuk, egy felső becslést kapunk az elérhető legjobb egyensúlyi pont célfüggvényértékére. Ha a megoldás kielégíti a komplementaritási feltételeket is, akkor ez az optimum egybeesik a legjobb egyensúlyi ponttal.

Itt feltételeztük, hogy az $\tilde{f}$ függvény az (5.1a), (5.1b), (5.2a), (5.2b) feltételek által meghatározott megengedett tartományon felveszi a maximumát.

6. Számítógépes tapasztalatok

Az általunk javasolt módszer gyakorlati alkalmazhatóságának vizsgálata érdekében számítógépes programot készítettünk. Hogy összehasonlítsuk módszerünk bizonyos értelemben vett „jobb" voltát, AGGARWAL módszerére is készítettünk programot.

A programok FORTRAN nyelven íródtak az MTA CDC 3300-as számítógépére. Pozitív kifizetőmátrixú feladatokban próbáltuk ki, ahol a mátrix elemei 1 és 100 közé estek.

Már kis méret esetén is módszerünk új egyensúlyi pontokat szolgáltatott, ami hatékonyságának igen jó fokmérője. A következő táblázatban ismertetjük az egyes feladatok esetén kapott egyensúlyi pontok számát:

1. algoritmus: *Aggarwal módszere.*
2. algoritmus: az általunk javasolt módszer.

	Az *A* játékos tiszta strat. száma	A *B* játékos tiszta strat. száma	Az egyensúlyi pontok száma	
			1.	2.
1.	2	2	2	3(+1)
2.	3	3	2	3(+1)
3.	4	3	2	4(+2)
4.	4	4	2	5(+3)
5.	5	3	2	5(+3)
6.	6	5	4	10(+6)
7.	6	6	3	5(+2)
8.	8	6	4	6(+2)

A futási idők a feladatok méretének növekedésével együtt nőttek, de a feladatok típusától is igen nagy mértékben függtek.

Az algoritmus egy egyensúlyi pontra többször is léphet. Kisebb méretű feladatok esetén az ilyen többszöri ismétlődés ritka volt, nagyobb méret esetén a feladat struktúrájától függően változott. Egy egyensúlyi pont ismétlődése egy adott bimátrix játék esetén is igen változott, voltak egyszer elért, de voltak többször, 4—5-ször is ismétlődő pontok.

IRODALOM

[1] AGGARWAL, V., "On the generation of all equilibrium points for bimatrix games through the Lemke—Howson algorithm", *Mathematical Programming* **4** (1973) 233—234.

[2] COTTLE, R. W. and DANTZIG, G. B., "Complementary pivot theory of mathematical programming", *Mathematics of the Decision Sciences, Part I, G. B. Dantzig and A. F. Veinott, Jr. Eds. American Mathematical Society, Providence, R. I.* 1968, 115—136.

[3] LEMKE, C. E., "On complementary pivot theory", [2] 95—114.

[4] LEMKE, C. E., "Bimatrix equilibrium points and mathematical programming", *Management Science* **11** (1965) 681—689.
[5] LEMKE, C. E. and HOWSON, J. T., "Equilibrium points of bimatrix games", *J. Soc. Indust. Appl. Math.* **12** (1964) 413—423.
[6] MAJTHAY, A., "A lexicographic complementary pivot algorithm for the solution of bimatrix games", *Studia Sci. Math. Hungarica* **7** (1972) 181—188.
[7] PRÉKOPA, A., *Lineáris programozás* (Bolyai János Matematikai Társulat, Budapest, 1968).
[8] Soós Zs., „A lineáris komplementaritási probléma és megoldási módszerei", ELTE szakdolgozat, Budapest, 1977.
[9] TODD, M. J., "Comments on a note by Aggarwal", *Mathematical Programming* **10** (1976) 130—133.

(Beérkezett: 1979. január 2.)

SOÓS ZSOLT
MTA SZÁMÍTÁSTECHNIKAI ÉS AUTOMATIZÁLÁSI KUTATÓ INTÉZET
1111 BUDAPEST, KENDE U. 13—17.

ON THE FINDING OF BIMATRIX GAME'S EQUILIBRIUM POINTS

Zs. Soós

This paper deals with the possibility of finding more equilibrium points of bimatrix game For this purpose we consider the geometric localization of these points, and then we extend the method which was proposed by A. MAJTHAY over another algorithm finding more equilibrium points. We make a comparison between the solutions of the new and the *Aggarwal-algorithm*.

Alkalmazott Matematikai Lapok 5 (1979) 177—190

MEGENGEDETT MEGOLDÁSOK VIZSGÁLATÁVAL BŐVÍTETT BENDERS DEKOMPOZÍCIÓS ALGORITMUS

HOFFER JÁNOS
Budapest

A dolgozat a *Benders-féle dekompozíciós algoritmus* egy módosításával foglalkozik. Az eredeti algoritmust kibővítve, az optimum értékét egyre jobban megközelítő célfüggvényértékű megengedett megoldásokat nyerhetünk. A módosított algoritmus illusztrálására két számpéldát közlünk, és számítógépes tapasztalatainkat ismertetjük.

1. A Benders-algoritmus rövid ismertetése

BENDERS [1] 1962-ben tette közzé algoritmusát, amely az alábbi típusú optimalizálási feladatok dekompozícióval való megoldására szolgál:

$$\mathbf{A}\mathbf{x} + F(\mathbf{y}) \leqq \mathbf{b} \tag{1.1}$$

$$\mathbf{x} \geqq \mathbf{0} \tag{1.2}$$

$$\mathbf{y} \in S \tag{1.3}$$

$$\max\left(\mathbf{c}^T\mathbf{x} + f(\mathbf{y})\right), \tag{1.4}$$

ahol $\mathbf{A}$: $m \times n_1$ méretű mátrix,
F: $R^{n_2} \to R^m$ folytonos leképezés,
$\mathbf{b}$: m dimenziós vektor,
$\mathbf{x}$: n_1 dimenziós vektor,
$\mathbf{y}$: n_2 dimenziós vektor,
f: $R^{n_2} \to R$ folytonos függvény,
$\mathbf{c}$: n_1 dimenziós vektor,
S: az n_2 dimenziós tér korlátos és zárt részhalmaza (az alkalmazásokban leggyakrabban diszkrét halmaz).

Jelöljük az m dimenziós térben elhelyezkedő

$$\mathbf{A}^T\mathbf{u} \geqq \mathbf{c}$$

$$\mathbf{u} \geqq \mathbf{0}$$

P poliéder extremális pontjait $\mathbf{u}^1, \mathbf{u}^2, \ldots, \mathbf{u}^k$-val, extremális irányait $\mathbf{v}^1, \mathbf{v}^2, \ldots, \mathbf{v}^l$-lel.

Tekintsük az alábbi programozási feladatot a most bevezetett jelölésekkel:

$$(\mathbf{v}^i)^T F(\mathbf{y}) \leqq (\mathbf{v}^i)^T \mathbf{b}, \quad i = 1, \ldots, l \tag{1.5}$$

$$x_0 + (\mathbf{u}^j)^T F(\mathbf{y}) - f(\mathbf{y}) \leqq (\mathbf{u}^j)^T \mathbf{b}, \quad j = 1, \ldots, k \tag{1.6}$$

$$\mathbf{y} \in S \tag{1.7}$$

$$\max x_0 \tag{1.8}$$

Ekvivalencia-tétel [4]

1.1. TÉTEL. Az (1.1)—(1.4) és az (1.5)—(1.8) feladatok ekvivalensek a következő értelemben:

1. Ha a két feladat közül egyiknek nincs megengedett megoldása, akkor a másiknak sincs.
2. A két feladat célfüggvénye egyszerre nem korlátos.
3. *a)* Az (1.1)—(1.4) feladat optimális megoldását $\begin{pmatrix}\mathbf{x}^*\\ \mathbf{y}^*\end{pmatrix}$-gal jelölve, $\begin{pmatrix}\mathbf{c}^T\mathbf{x}^* + f(\mathbf{y}^*)\\ \mathbf{y}^*\end{pmatrix}$ optimális megoldása az (1.5)—(1.8) feladatnak.

3. *b)* Ha az (1.5)—(1.8) feladat optimális megoldása: $\begin{pmatrix}x_0^*\\ \mathbf{y}^*\end{pmatrix}$,

és ha ezen jelölések mellett $\mathbf{x}^*$ optimális megoldása az

$$\mathbf{A}\mathbf{x} \leqq \mathbf{b} - F(\mathbf{y}^*) \tag{1.9}$$

$$\mathbf{x} \geqq \mathbf{0} \tag{1.10}$$

$$\max \mathbf{c}^T\mathbf{x} \tag{1.11}$$

lineáris programozási feladatnak, akkor $\begin{pmatrix}\mathbf{x}^*\\ \mathbf{y}^*\end{pmatrix}$ optimális megoldása az (1.1)—(1.4) feladatnak.

Az ekvivalenciára vonatkozó megállapításaink szerint tehát az (1.1)—(1.4) feladat megoldása helyett elegendő az (1.5)—(1.8) optimális megoldását megkeresnünk, majd az (1.9)—(1.11) lineáris programozási feladat megoldásával megkaphatjuk az optimális megoldás első n_1 komponensét.

Azonban az (1.5)—(1.8) feladat explicit előállításához (és így megoldásához is) a P poliéder összes extremális pontjaira és irányaira szükségünk van. Ezek előállítása még viszonylag kisméretű feladatok esetében is rendkívül munkaigényes, sokszor megoldhatatlan. Ezért — BENDERS javaslata alapján — az algoritmusban az (1.5)—(1.8) feladat helyett annak — feltételek elhagyásával keletkező — relaxáltjait kell megoldani, majd egy-egy lineáris programozási feladattal ellenőrizni, hogy a relaxált feladat optimális megoldása megengedett megoldása-e az (1.5)—(1.8) feladatnak. Nemleges válasz esetén a lineáris feladatok egyben új és új extrémálisokat szolgáltatnak az (1.5)—(1.8) relaxáltjaihoz. Az ezekből adódó feltételeket hozzávéve a megelőző relaxált feladathoz újabb iterációt kell végrehajtani.

Az (1.5)—(1.8) feladatnak az i-edik iterációban megoldandó relaxáltját BSP(i)-vel jelöljük. Ha az említett lineáris programozási feladatok megoldásai során nyert extrémálisokat

$$\mathbf{v}^1, \mathbf{v}^2, \ldots, \mathbf{v}^{l_i},$$

illetve

$$\mathbf{u}^1, \mathbf{u}^2, \ldots, \mathbf{u}^{k_i}$$

jelöli, akkor BSP(i) a következő:

$$(\mathbf{v}^m)^T F(\mathbf{y}) \leqq (\mathbf{v}^m)^T \mathbf{b}, \qquad m = 1, \ldots, l_i$$

$$x_0 + (\mathbf{u}^j)^T F(\mathbf{y}) - f(\mathbf{y}) \leqq (\mathbf{u}^j)^T \mathbf{b}, \qquad j = 1, \ldots, k_i$$

$$\mathbf{y} \in S$$

$$\max x_0.$$

Megjegyzések:

1. A BSP(i) feladat optimális megoldását, ha az létezik, $\begin{pmatrix} x_0^i \\ \mathbf{y}^i \end{pmatrix}$-vel jelöljük. Az x_0^i érték bármely i esetén, a relaxáció miatt az (1.5)—(1.8) feladat optimális célfüggvényértékének felső becslése. Az ekvivalencia-tételből következően az x_0^i értékek az (1.1)—(1.4) feladat optimális célfüggvényértékének is felső becslései.

2. A BSP(i) feladatok konstrukciójából az

$$x_0^1 \geqq x_0^2 \geqq x_0^3 \geqq \ldots$$

reláció következik.

Az algoritmus lépései a következők:

1. lépés: Legyen $i=0$, $\mathbf{y}^0 \in S$ tetszőleges vektor, $x_0^0 = +\infty$ és térjünk rá a 2. lépésre.

2. lépés: Oldjuk meg az

$$\mathbf{A}^T \mathbf{u} \geqq \mathbf{c}$$

$$\mathbf{u} \geqq \mathbf{0}$$

$$\min (\mathbf{b} - F(\mathbf{y}^i))^T \mathbf{u}$$

— a továbbiakban LD($\mathbf{y}^i$)-nak nevezett — lineáris programozási feladatot.

a) Ha nincs megengedett megoldása, akkor az (1.1)—(1.4) feladatnak vagy nincs megengedett megoldása, vagy a célfüggvénye nem korlátos.

Az eljárás itt befejeződik.

b) Ha a célfüggvény nem korlátos, akkor a BSP(i) feltételeihez csatoljuk a kapott $\mathbf{v}$ extremális irányból adódó

$$\mathbf{v}^T F(\mathbf{y}) \leqq \mathbf{v}^T \mathbf{b}$$

feltételt (a szimplex módszer által előállított $\mathbf{u}$ extremális pontból adódó

$$x_0 + \mathbf{u}^T F(\mathbf{y}) - f(\mathbf{y}) \leqq \mathbf{u}^T \mathbf{b}$$

feltétel is csatolható!), és térjünk rá a 4. lépésre.

c) Ha optimális megoldás van, akkor jelöljük z^i-vel az optimum értékét, $\mathbf{u}$-val az optimális megoldást, és térjünk rá a 3. lépésre.

3. lépés: Vizsgáljuk meg, hogy teljesül-e az úgynevezett optimalitási kritérium

$$x_0^i \leqq z^i + f(\mathbf{y}^i).$$

Ha teljesül, akkor térjünk rá az 5. lépésre. Egyébként csatoljuk a BSP(i) feladat feltételeihez az

$$x_0 + \mathbf{u}^T F(\mathbf{y}) - f(\mathbf{y}) \leqq \mathbf{u}^T \mathbf{b}$$

feltételt, és térjünk rá a 4. lépésre.

4. lépés: Oldjuk meg az új feltételek hozzáadásával keletkezett BSP($i+1$) feladatot. $i := i+1$.

a) Ha optimális megoldása van, jelöljük azt $\begin{pmatrix} x_0^i \\ \mathbf{y}^i \end{pmatrix}$-vel, és térjünk rá a 2. lépésre.

b) Ha nincs megengedett megoldása, akkor az (1.1)—(1.4) feladatnak sincs megengedett megoldása.

Az eljárás itt befejeződik.

5. lépés: $\mathbf{y}^* := \mathbf{y}^i$, $x_0^* := x_0^i$, és oldjuk meg az (1.9)—(1.11) — a továbbiakban LP($\mathbf{y}^*$)-nak nevezett — lineáris programozási feladatot. Ennek optimális megoldását $\mathbf{x}^*$-gal jelölve, $\begin{pmatrix} \mathbf{x}^* \\ \mathbf{y}^* \end{pmatrix}$ az (1.1)—(1.4) feladat optimális megoldása, x_0^* az optimális célfüggvényérték.

Az algoritmus részletes leírása, a szükséges tételek és bizonyításuk megtalálhatók BENDERS [1]-ben, KOVÁCS LÁSZLÓ BÉLA [4]-ben és LASDON [5]-ben.

Ezen a ponton kell az olvasó figyelmét arra felhívni, hogy az 1. szakaszban leírtakra sokkal inkább illik az algoritmuscsalád elnevezés, mert:

— a fellépő BSP(i) feladatok jellege és így a megoldásukhoz felhasználandó módszer az F leképezéstől, az f függvénytől és az S halmaztól függ,

— nem részleteztük azt, hogy az LD($\mathbf{y}^i$), LP($\mathbf{y}^*$) feladatokat milyen módszerrel oldjuk meg.

2. Az algoritmus vizsgálata, a megengedett megoldások előállítása és vizsgálata

Ebben a fejezetben megmutatjuk, hogy az eredeti feladat néhány megengedett megoldását elő tudjuk állítani az algoritmus egyes lépéseiben, annak ellenére, hogy relaxációt (metszéseket) alkalmazunk az (1.5)—(1.8) feladat megoldására. Ily módon az algoritmus esetleges félbeszakításakor (pl. túl sok számítási idő vagy az iterációk nagy száma miatt) is rendelkezésünkre áll az (1.1)—(1.4) feladatnak a félbeszakításig talált legjobb célfüggvényértékű megoldása.

Az algoritmus módosított változatának felépítéséhez térjünk rá először az LD ($\mathbf{y}^i$), LP ($\mathbf{y}^*$) lineáris programozási feladatok vizsgálatára.

Az LD ($\mathbf{y}^i$) feladat helyett a vele ekvivalens, feltételeiben egyenlőséget tartalmazó lineáris programozási feladatot oidjuk meg. Ehhez igen gazdaságos a módosított szimplex módszer alkalmazása, mert:

1. Az i-edik feladat megoldásakor az utolsó megengedett bázis inverzét megőrizve, azt felhasználhatjuk az $i+1$-edik induló bázisinverzeként, hiszen az i-edik és az $i+1$-edik feladatok csupán célfüggvényükben különböznek. Ez jelentős gépidő-megtakarítást eredményez — a bázisinverz megőrzési módjától függően —

azzal a változattal szemben, amelyben a lineáris programozási feladatok megoldását minden alkalommal a kétfázisú szimplex módszer első fázisából indítjuk.

2. Ha az i-edik lineáris programozási feladatnak optimális megoldása van, akkor további számítások elvégzése nélkül rendelkezésünkre áll az LD $(\mathbf{y}^i)$ feladat duálisának, LP $(\mathbf{y}^i)$-nek egy optimális megoldása, nevezetesen az i-ediknek megoldott, az LD $(\mathbf{y}^i)$ feladattal ekvivalens lineáris programozási feladat optimális bázisához tartozó duál vektor (azaz a szimplex szorzókból álló vektor).

3. Az előző megjegyzés alapján az LP $(\mathbf{y}^*)$ feladat megoldása teljes egészében elhagyható, hiszen az utolsónak kiszámított LD $(\mathbf{y}^i)$=LD $(\mathbf{y}^*)$ feladatból $\mathbf{x}^*$ a leírt módon megkapható.

Az LD $(\mathbf{y}^i)$ megoldása után két esetben folytatódik az algoritmus újabb *Benders-alfeladat* (BSP $(i+1)$) megoldásával:

a) ha az LD $(\mathbf{y}^i)$ célfüggvénye nem korlátos,

b) ha az LD $(\mathbf{y}^i)$-nek optimális megoldása van, de nem teljesül az optimalitási kritérium.

Mindkét esetben azért kell az algoritmust folytatni, mert $\begin{pmatrix} x_0^i \\ \mathbf{y}^i \end{pmatrix}$ nem megengedett megoldása az (1.5)—(1.8) feladatnak. (Ennek bizonyítása is megtalálható a már említett művekben.) Azonban hangsúlyoznunk kell, hogy a nem megengedettség nem $\mathbf{y}^i$-re vonatkozik, hanem $\begin{pmatrix} x_0^i \\ \mathbf{y}^i \end{pmatrix}$-re. Nézzük meg pontosabban!

Az *a)* esetben — mint láttuk — új feltételként egy

$$\mathbf{v}^T F(\mathbf{y}) \leqq \mathbf{v}^T \mathbf{b} \tag{2.1}$$

alakú feltételt kell csatolnunk a BSP (i) feladathoz, kizárva ily módon a BSP $(i+1)$ megengedett megoldásai közül az $\begin{pmatrix} x_0^i \\ \mathbf{y}^i \end{pmatrix}$ vektort. Ez a feltétel azonban, bármekkora legyen is az x_0 változó értéke, kizárja az $\mathbf{y}^i$ vektort, mivel az LD $(\mathbf{y}^i)$ célfüggvénye nem korlátos, amiből az x_0-tól függetlenül következik a

$$\mathbf{v}^T(\mathbf{b} - F(\mathbf{y}^i)) < 0$$

reláció.

A *b)* esetben csatolt új feltétel

$$x_0 + \mathbf{u}^T F(\mathbf{y}) - f(\mathbf{y}) \leqq \mathbf{u}^T \mathbf{b} \tag{2.2}$$

alakú. Ez a feltétel csak az $\begin{pmatrix} x_0^i \\ \mathbf{y}^i \end{pmatrix}$ vektort zárja ki. Az $\mathbf{y}^i$ vektor része lehet bármely BSP (j) $(j>i)$ feladat megengedett, sőt optimális megoldásának, természetesen az x_0^i-nél kisebb x_0 értékkel (lásd a 4. szakasz P1 példáját!).

Nézzük most más szempontból e két esetet. Erre vonatkozólag a következő állítást mondjuk ki.

2.1. ÁLLÍTÁS: *a)* Ha az LD $(\mathbf{y}^i)$ feladat célfüggvénye nem korlátos, akkor az $\mathbf{y}^i$ vektorhoz nem található egyetlen $\mathbf{x}$ vektor sem úgy, hogy $\begin{pmatrix} \mathbf{x} \\ \mathbf{y}^i \end{pmatrix}$ az (1.1)—(1.4) feladat megengedett megoldása lenne.

b) Ha az LD ($\mathbf{y}^i$) feladatnak optimális megoldása van, akkor az adott $\mathbf{y}^i$ mellett az (1.1)—(1.4) feladat lehető legjobb célfüggvényértékét a 2. szakasz 2) megjegyzése alapján előállított $\mathbf{x}^i$ duál vektor szolgáltatja, és az

$$f(\mathbf{y}^i)+(\mathbf{b}-F(\mathbf{y}^i))^T\mathbf{u}=f(\mathbf{y}^i)+z^i$$

érték az (1.1)—(1.4) optimális célfüggvényértékének alsó becslése.

Bizonyítás: Mindkét esetben a dualitás tételre hivatkozunk, illetve annak PRÉKOPA [6]-ban szereplő bizonyítására, továbbá arra a tényre, hogy az LD ($\mathbf{y}^i$) duális feladata az LP ($\mathbf{y}^i$) feladat, az LP ($\mathbf{y}^i$) feladat pedig azonos az (1.1)—(1.4) feladattal az $\mathbf{y}=\mathbf{y}^i$ rögzítés mellett (csupán célfüggvényük különbözik a konstans $f(\mathbf{y}^i)$ értékben).

Ezért az *a)* esetben az LP ($\mathbf{y}^i$) feladatnak nincs megengedett megoldása, és így nem találhatunk az $\mathbf{y}^i$ vektort kiegészítő $\mathbf{x}$ vektort úgy, hogy $\begin{pmatrix}\mathbf{x}\\ \mathbf{y}^i\end{pmatrix}$ megengedett megoldása lenne az (1.1)—(1.4) feladatnak.

A *b)* esetben az LP ($\mathbf{y}^i$) feladatnak is optimális megoldása van (például $\mathbf{x}^i$), optimális célfüggvényértéke z^i. Ily módon $\begin{pmatrix}\mathbf{x}^i\\ \mathbf{y}^i\end{pmatrix}$ megengedett megoldása az (1.1)—(1.4) feladatnak és az $f(\mathbf{y}^i)+z^i$ érték egy alsó becslés annak optimumára.

A 2.1. állítás *b)* pontja alapján azokban az iterációkban, amelyekben az LD ($\mathbf{y}^i$) feladatnak optimális megoldása van, meg kell vizsgálnunk, hogy az $f(\mathbf{y}^i)+z^i$ érték nagyobb-e a megelőző legjobb alsó becslésnél. Ha a most kapott alsó becslés nagyobb a megelőzőnél, akkor meg kell őriznünk az alsó becslés értékét, az $\mathbf{y}^i$ és az $\mathbf{x}^i$ vektort, mint az (1.1)—(1.4) feladat eddig talált legjobb megoldását.

Ha már találtunk alsó becslést, akkor az algoritmust megállíthatjuk olyankor is, amikor az optimalitási kritérium még nem teljesül, de a kapott legjobb megoldás valamilyen más szempontnak (a továbbiakban leállítási kritériumnak) megfelel.

Példaként sorolunk néhány lehetséges leállítási kritériumot:

1. Az elért legjobb célfüggvényérték valamilyen számunkra elegendően nagy alsó korlátot meghaladott.

2. Az optimalitási kritérium nem teljesül ugyan, de a legjobb alsó és a felső becslés eltérése — a feladat gyakorlati tartalmát figyelembe véve — nem jelentős.

3. Az eltelt számítási időt és az elért legjobb célfüggvényértéket együttesen tekintve megelégszünk az eddigi legjobb megoldással.

Megjegyzés: A 2.1. állítás *b)* pontjának még további haszna is lehet. Tegyük fel, hogy az algoritmus során azt találjuk, hogy a BSP(j) feladatnak optimális célfüggvényértéke (felső becslés!) megegyezik a legjobb alsó becslés értékével. Ekkor a legjobb megoldásvektor optimális megoldása az (1.1)—(1.4) feladatnak. Ezekben az esetekben tehát nemcsak az LP($\mathbf{y}^*$), de az LD($\mathbf{y}^*$) feladat megoldása is mellőzhető.

Jelöljük az algoritmus során kapott legjobb célfüggvényértéket $\hat{x}_0$-val, a legjobb megoldást $\begin{pmatrix}\hat{\mathbf{x}}\\ \hat{\mathbf{y}}\end{pmatrix}$-val. A *módosított Benders-dekompozíció* lépései a következők:

1. lépés: Legyen $i=0$, $\mathbf{y}^0\in S$ tetszőleges vektor (lehetőség szerint olyan, hogy jó megengedett megoldást szolgáltasson), $x_0^0=+\infty$, $\hat{x}_0=-\infty$ és térjünk rá a 2. lépésre.

2. lépés: Oldjuk meg az LD($\mathbf{y}^i$) feladatot módosított szimplex módszerrel!

a) Ha az LD($\mathbf{y}^i$)-nek nincs megengedett megoldása, akkor az (1.1)—(1.4) feladatnak vagy nincs megengedett megoldása, vagy a célfüggvénye nem korlátos. *Az eljárás itt befejeződik.*

b) Ha a célfüggvény nem korlátos, akkor a BSP(i) feltételeihez csatoljuk a kapott $\mathbf{v}^*$ extremális irányból adódó:

$$\mathbf{v}^{*T}F(\mathbf{y}) \leqq \mathbf{v}^{*T}\mathbf{b}$$

feltételt, és a kapott $\mathbf{u}^*$ extremális pontból származó:

$$x_0+\mathbf{u}^{*T}F(\mathbf{y})-f(\mathbf{y}) \leqq \mathbf{u}^{*T}\mathbf{b}$$

feltételt is, és térjünk rá a 4. lépésre.

c) Optimális megoldás esetén, ha

$$\hat{x}_0 < z^i+f(\mathbf{y}^i),$$

akkor legyen

$$\hat{x}_0 := z^i+f(\mathbf{y}^i), \quad \hat{\mathbf{x}} := \mathbf{x}^i, \quad \hat{\mathbf{y}} := \mathbf{y}^i$$

és térjünk rá a 3. lépésre.

d) Optimális megoldás esetén, ha

$$\hat{x}_0 \geqq z^i+f(\mathbf{y}^i),$$

akkor térjünk rá a 3. lépésre.

3. lépés: Vizsgáljuk meg, hogy teljesül-e az optimalitási kritérium:

$$x_0^i \leqq \hat{x}_0 !$$

Ha teljesül, akkor $\begin{pmatrix}\hat{\mathbf{x}}\\ \hat{\mathbf{y}}\end{pmatrix}$ optimális megoldása az (1.1)—(1.4) feladatnak, $\hat{x}_0$ az optimális célfüggvényérték. *Itt az eljárás befejeződik.*

3/a lépés: Ha az $\begin{pmatrix}\hat{x}_0\\ \hat{\mathbf{x}}\\ \hat{\mathbf{y}}\end{pmatrix}$ legjobb megoldás megfelel a leállítási kritériumnak, akkor az *eljárás itt befejeződik.*

Egyébként a BSP(i) feladat feltételeihez csatoljuk az:

$$x_0+\mathbf{u}^{*T}F(\mathbf{y})-f(\mathbf{y}) \leqq \mathbf{u}^{*T}\mathbf{b}$$

feltételt és térjünk rá a 4. lépésre.

4. lépés: $i:=i+1$, oldjuk meg a BSP(i) feladatot.

a) Ha a BSP(i)-nek nincs megengedett megoldása, akkor az (1.1)—(1.4) feladatnak nincs megengedett megoldása. *Itt az eljárás befejeződik.*

b) Ha optimális megoldása van: $\begin{pmatrix}x_0^i\\ \mathbf{y}^i\end{pmatrix}$, és

$$x_0^i \leqq \hat{x}_0,$$

akkor $\begin{pmatrix}\hat{\mathbf{x}}\\ \hat{\mathbf{y}}\end{pmatrix}$ optimális megoldása az (1.1)—(1.4) feladatnak, $\hat{x}_0$ célfüggvényértékkel. *Itt az eljárás befejeződik.*

c) Ha optimális megoldása van: $\begin{pmatrix}x_0^i\\ \mathbf{y}^i\end{pmatrix}$ és az új felső becslés ismeretében $\begin{pmatrix}\hat{\mathbf{x}}\\ \hat{\mathbf{y}}\end{pmatrix}$ megfelel a leállítási kritériumnak, akkor *az eljárás itt befejeződik.* Egyébként térjünk rá a 2. lépésre.

3. Numerikus példák

A *módosított Benders-dekompozíció* működését két lineáris, vegyes, egészértékű feladaton mutatjuk be.

P1) A feladat a következő [2]:

$$\begin{aligned}
15x_1+4x_2-8x_3 &\leqq 120\\
x_1+4x_3 &\leqq 32\\
7x_1+2x_2+6x_3 &\leqq 70\\
x_1-8y_1 &\leqq 0\\
x_2-30y_2 &\leqq 0\\
x_3-8y_3 &\leqq 0\\
x_j &\geqq 0,\quad j=1,2,3\\
y_j &\in\{0,1\},\quad j=1,2,3\\
\max\,(6x_1+4x_2+7x_3-20y_1-80y_2-10y_3)&
\end{aligned}$$

$\mathbf{y}^0$-nak az $(1, 1, 0)^T$ vektort választjuk. Az algoritmus főbb lépései az optimum megtalálásáig a következők: LD($\mathbf{y}^0$)-nak optimális megoldása van, $\hat{x}_0=20$, a legjobb megoldás: $(0, 30, 0, 1, 1, 0)^T$.

Első iteráció:

BSP(1)-nek optimuma van, $\mathbf{y}^1=(0, 0, 1)^T$, $x_0^1=119{,}6$, LD($\mathbf{y}^1$)-nek optimális megoldása van, $\hat{x}_0=46$, a legjobb megoldás: $(0, 0, 8, 0, 0, 1)^T$.

Második iteráció:

BSP(2)-nek optimuma van, $\mathbf{y}^2=(1, 1, 1)^T$, $x_0^2=91{,}6$, LD($\mathbf{y}^2$)-nek optimális megoldása van, nem kapunk belőle jobb megoldást, mert $f(\mathbf{y}^2)+z^2=21{,}6667<46=\hat{x}_0$.

Harmadik iteráció:

BSP(3)-nak optimuma van, $\mathbf{y}^3=(1, 0, 1)^T$, $x_0^3=51{,}6667$, LD($\mathbf{y}^3$)-nak optimális megoldása van, nem kapunk belőle jobb megoldást, mert $f(\mathbf{y}^3)+z^3=43<46=\hat{x}_0$.

Negyedik iteráció:

BSP(4)-nek optimuma van, $\mathbf{y}^4=\hat{\mathbf{y}}=(0, 0, 1)^T$, $x_0^4=\hat{x}_0=46$, $\hat{\mathbf{x}}=(0, 0, 8)^T$.
Az eljárás itt befejeződik.

P2) A második feladatban csak a folytonos és a diszkrét mátrixot adjuk meg, elhagyva a feladat egyenlőtlenséges formában való felírását. A folytonos változók mátrixa:

$$\begin{matrix} 3 & -1 & -3 & 1 & 1 \\ -2 & 1 & -2 & 2 & -1 \\ 2 & -1 & -2 & 3 & 1 \\ 3 & 1 & 0 & -28 & -1 \\ 4 & -1 & 1 & 0 & 1 \\ -20 & -20 & -20 & -20 & -20 \\ -20 & 2 & 1 & 5 & 2 \\ 3 & -137 & 2 & 5 & -400 \\ 1 & 1 & 1 & 1 & 1 \\ 2 & -1 & -1 & 1 & 1, \end{matrix}$$

a diszkrét változók mátrixa:

$$\begin{matrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 3 & 3 & 3 & 3 & 2 & 2 & 2 & 2 & 0 & 0 \\ 0 & -5 & -8 & -3 & 0 & 7 & 5 & 3 & 1 & 4 \\ 1 & 2 & 3 & 4 & 5 & 0 & -1 & -2 & -3 & -4 \\ 8 & 0 & 7 & 0 & 6 & 5 & 4 & 3 & 2 & 1 \\ -20 & 11 & 0 & 9 & -9 & -2 & 2 & 9 & 1 & -1 \\ -5 & 2 & 3 & -5 & 2 & 3 & 0 & -5 & 2 & 3 \\ -2 & -2 & -2 & -2 & -2 & -2 & -2 & 0 & -2 & -4 \\ 10 & 11 & 12 & -14 & -15 & -16 & -13 & 14 & 15 & 16 \\ 4 & -5 & -6 & 7 & 1 & 2 & 3 & -4 & -5 & 6 \end{matrix}$$

$$\begin{matrix} 10 & -1 & -7 & -3 & 1 & 0 & 2 & 23 \\ -1 & 7 & 2 & -5 & 2 & -1 & 4 & -1 \\ 2 & 5 & 41 & 24 & 3 & 1 & 6 & -2 \\ -3 & 4 & 1 & 14 & 4 & 2 & 8 & 4 \\ 4 & -7 & -5 & -4 & 16 & -2 & -8 & 5 \\ 0 & 3 & -2 & -6 & 5 & 5 & -6 & 0 \\ 5 & 0 & 7 & -7 & 6 & 3 & 100 & -3 \\ -4 & 5 & -9 & -1 & 7 & -3 & -70 & -1 \\ 3 & 3 & 1 & -2 & 8 & 4 & 11 & -4 \\ 2 & 2 & 2 & -5 & 9 & -4 & 2 & -4, \end{matrix}$$

a **b** vektor: $(11, 22, 7, 8, 42, -1, 20, -30, 25, 5)^T$, a folytonos változók célfüggvényvektora: $(1, -1, 1, 1, -2)^T$, a diszkrét változók célfüggvényvektora: $(10, 11, 8, 10, 9, -2, 10, 13, 17, 14, 5, -2, 30, 71, 10, -5, 25, 15)^T$. A feladat minden feltétele kisebb-egyenlőséges, a folytonos változókra nemnegativitási feltételt teszünk, a diszkrét változók 0 vagy 1 értékűek, a célfüggvény maximalizálandó. $\mathbf{y}^0$-nak a $(0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1)^T$ vektort választjuk. Az algoritmus főbb lépései az optimum megtalálásáig a következők:

LD($\mathbf{y}^0$) célfüggvénye nem korlátos.

Első iteráció:

BSP(1)-nek optimuma van, $\mathbf{y}^1=(0, 0, 0, 1, 1, 1, 1, 0, 1, 0, 1, 0, 1, 1, 0, 0, 1, 1)^T$, $x_0^1=248{,}52$. LD($\mathbf{y}^1$) célfüggvénye nem korlátos.

Második iteráció:

BSP(2)-nek optimuma van $\mathbf{y}^2=(0, 0, 0, 1, 1, 1, 1, 0, 1, 0, 0, 0, 1, 1, 0, 0, 1, 1)^T$, $x_0^2=224{,}42$. LD($\mathbf{y}^2$) célfüggvénye nem korlátos.

Harmadik iteráció:

BSP(3)-nak optimuma van, $\mathbf{y}^3=(0, 1, 0, 1, 1, 0, 1, 1, 1, 0, 0, 0, 1, 1, 0, 0, 0, 1)^T$, $x_0^3=216{,}75$. LD($\mathbf{y}^3$)-nak optimális megoldása van, $\hat{x}_0=216{,}61$, $\hat{\mathbf{x}}=(0{,}41;\ 0{,}54;\ 29{,}9;\ 0{,}84;\ 0)^T$, $\hat{\mathbf{y}}=\mathbf{y}^3$.

Negyedik iteráció:

BSP(4)-nek optimális megoldása van, $\mathbf{y}^4=\mathbf{y}^3=\hat{\mathbf{y}}$, $x_0^4=\hat{x}_0=216{,}61$.

Az eljárás itt befejeződik.

Az első feladatban rögtön egy megengedett megoldást kaptunk. Ugyanezt a második feladatnál csak a harmadik iterációban értük el. Az első feladatnál, ha valamilyen gyakorlatias megfontolásból megelégszünk a 20-szal, vagy a 46-tal, mint célfüggvényértékkel, akkor az első, illetve a második iterációnál megállíthatjuk a számításokat. Mind az első, mind a második esetben a számítások nagy részét megtakaríthatjuk, és — szerencsénkre — a második esetben éppen az optimális megoldást kapjuk meg. A második mintapélda megoldása során a harmadik iterációban nem teljesül ugyan az optimalitási kritérium, de az optimális célfüggvényérték 216,75-nál nem lehet nagyobb. Az $\begin{pmatrix}\hat{\mathbf{x}}\\ \hat{\mathbf{y}}\end{pmatrix}$ vektort elfogadhatjuk optimálisnak anélkül, hogy optimális voltáról további számításokkal meggyőződnénk, hiszen a hozzá tartozó célfüggvényérték 216,61.

Ha optimális megoldásra törekednénk, akkor mindkét mintafeladat esetében, az algoritmus a BSP(4) feladat megoldásával zárulna, ily módon el lehet hagyni az eredeti javaslat szerint szükséges LD($\mathbf{y}^4$), LP($\mathbf{y}^4$) feladatok megoldását.

4. Számítástechnikai tapasztalatok vegyes, egészértékű feladatok esetében

Az algoritmus vizsgálatára FORTRAN nyelvű program íródott az MTA CDC 3300-as gépére, STRAZICKY BEÁTA lineáris programozási szubrutinjának [7] és VIZVÁRI BÉLA egészértékű feladatokat megoldó szubrutinjának felhasználásával.

A feladatok optimális megoldásának megtalálásához szükséges teljes számítási időt az eredeti javaslat szerinti algoritmust alkalmazva mértük. Elhagytuk azonban az LP(y^*) feladat megoldását a 2. szakasz 3. megjegyzése alapján; az LD(y^i) feladatok megoldásánál pedig a 2. szakasz 1. és 2. megjegyzésének ötletét használtuk fel.

Pozitív optimumú feladatok esetén az egyes részfeladatok megoldása után megvizsgáltuk az addig kapott legjobb megengedett megoldást: ha az optimum legjobb alsó becslése már pozitív volt, akkor kiszámítottuk a

$$\lambda = \frac{\hat{x}_0}{x_0^i} \cdot 100$$

hányadost. Minthogy x_0^i felső becslés az (1.1)—(1.4) feladat optimum értékére, ezért a λ kiszámításáig elért legjobb megoldáshoz tartozó célfüggvényérték legalább λ százaléka az optimum értékének. Az algoritmus során a megtalált legjobb megengedett megoldásokat öt kategóriába soroltuk a λ értéke szerint:

1. $90{,}0 \leqq \lambda < 95{,}0$
2. $95{,}0 \leqq \lambda < 99{,}0$
3. $99{,}0 \leqq \lambda < 99{,}5$
4. $99{,}5 \leqq \lambda < 99{,}9$
5. $99{,}9 \leqq \lambda.$

Kiszámítottuk, hogy mennyi idő telt el a számítások megkezdésétől az egyes kategóriákba eső, elsőnek megtalált megengedett megoldások előállításáig. Ezeket az időmennyiségeket t_i-vel ($i=1, \ldots, 5$), az optimum előállításához szükséges időmennyiséget t_{teljes}-sel jelölve, kiszámítottuk a

$$\mu_i = \frac{t_i}{t_{\text{teljes}}} \cdot 100$$

hányadosokat. Ily módon azt mondhatjuk, hogy az optimum értékének legalább λ_i százalékát elérő célfüggvényértékű megengedett megoldást a teljes számítási idő μ_i százaléka alatt nyertünk.

Számítástechnikai tapasztalataink közül a legérdekesebbeket említjük itt meg. A feladatokat P1, P2, P3, P4, P5-tel jelöljük. P1, illetve P2 a 3. szakaszban már szerepeltek. P_3 a P_2-nek további 14 egész értékű változóval való bővítése a jobb oldali

vektor változatlanul tartása mellett. A 14 új változónak megfelelő feltételi együttható mátrix:

0	3	−5	10	1	−1	2	1	1	1	1	1	1	1
2	2	−7	11	1	−2	−3	5	5	5	5	0	0	−3
−2	1	−2	12	1	−1	0	−3	−17	8	7	6	−2	6
3	0	−1	13	2	−1	6	−2	6	7	8	−17	−3	0
−3	−1	−8	14	1	−2	1	2	3	4	5	6	7	8
1	−2	−4	15	1	−1	1	−1	1	−1	1	−1	1	−1
−1	−3	−3	16	1	−2	1	1	1	1	−1	−1	−1	−1
1	4	−7	17	0	−1	−3	−3	−3	−3	−3	−3	−3	−3
11	6	−1	18	1	−1	5	0	5	−3	0	5	−2	10
−1	5	−9	19	1	0	2	−2	0	0	−7	−8	10	7

A célfüggvénybeli együtthatók rendre:

1	3	−32	99	0	−3	10	−3	15	16	17	18	20	9.

A P_4, P_5 feladatok véletlenszerűen generált, 20%-os kitöltöttségű mátrixúak. A feladatok méreteit az alábbi táblázatban foglaljuk össze:

	Feltételek száma: m	Folytonos változók száma: n_1	Egészértékű változók száma: n_2
P1	6	3	3
P2	10	5	18
P3	10	5	32
P4	45	45	35
P5	45	45	35

A teljes számítási időt, az iterációk számát és az algoritmus során előállított megengedett megoldásokat nagy mértékben befolyásolja az $\mathbf{y}^0$ választása. Ezt támasztják alá az alábbi eredmények (a táblázatokban található vonások azt jelzik, hogy az algoritmus során nem találtunk a megfelelő kategóriába eső megengedett megoldást):

P1	*a)*	*b)*
iterációk száma	5	3
t_{teljes}(sec)	5,5	3,2
μ_1 (%)	—	91,38
μ_2 (%)	76,15	—
μ_3 (%)	—	—
μ_4 (%)	—	—
μ_5 (%)	95,29	100,0

a) esetben $\mathbf{y}^\circ = (0, 0, 1)^T$ az egészértékű változók optimális értékei,

b) esetben $\mathbf{y}^\circ = (0, 1, 1)^T$.

P2	*a)*	*b)*	*c)*
iterációk száma	2	4	4
t_{teljes}(sec)	4,0	14,6	15,8
μ_1 (%)	—	—	—
μ_2 (%)	—	—	—
μ_3 (%)	—	—	—
μ_4 (%)	—	—	—
μ_5 (%)	91,42	69,87	73,4

a) esetben $\mathbf{y}^\circ = \mathbf{y}^*$
b) esetben $\mathbf{y}^\circ = (0, 1, 0, 1, \ldots)^T$
c) esetben $\mathbf{y}^\circ = \mathbf{0}$.

P3	*a)*	*b)*
iterációk száma	2	3
t_{teljes}(sec)	5,4	56,9
μ_1 (%)	—	—
μ_2 (%)	—	—
μ_3 (%)	—	—
μ_4 (%)	—	—
μ_5 (%)	94,3	100,0

a) esetben $\mathbf{y}^\circ = \mathbf{y}^*$
b) esetben $\mathbf{y}^\circ = (0, 1, 0, 1, \ldots)^T$

P4	*a)*	*b)*
iterációk száma	3	4
t_{teljes}(sec)	238,6	322,6
μ_1 (%)	—	92,74
μ_2 (%)	—	94,47
μ_3 (%)	93,31	98,78
μ_4 (%)	96,11	—
μ_5 (%)	98,62	100,0

a) esetben $\mathbf{y}^\circ = \mathbf{y}^*$
b) esetben $\mathbf{y}^\circ = \mathbf{0}$.

P5	*a*	*b)*
iterációk száma	5	10
t_{teljes}(sec)	228,4	473,3
μ_1 (%)	—	62,27
μ_2 (%)	72,14	67,29
μ_3 (%)	76,31	70,14
μ_4 (%)	—	73,59
μ_5 (%)	93,91	95,3

a) esetben $\mathbf{y}^\circ = \mathbf{y}^*$
b) esetben $\mathbf{y}^\circ = \mathbf{0}$.

Tapasztalatainkat három pontban foglaljuk össze:

1. A megvizsgált feladatok alapján ZOUTENDIJK [8] véleményével ellentétben azt sejtjük, hogy a számítások jó egészértékű megoldással való indítása a számítási idő csökkenéséhez vezet a feladatok többségében. A felsorolt tesztfeladatok közül kizárólag P1-ben követelt több időt az optimális megoldás megtalálása $\mathbf{y}^0 = \mathbf{y}^*$ esetén más induló vektor választásához képest. P2-ben az optimálistól eltérő, induló egész vektorral a számítások kb. 3,5-szer, P3-ban kb. 10-szer, P4-ben kb. 1,3-szer, P5-ben kb. 2-szer annyi ideig tartottak, mint az $\mathbf{y}^0 = \mathbf{y}^*$ választással.

2. Egyetlen esetben fordult csak elő, a P3 feladat *b)* esetében, hogy egészen az optimális megoldás megtalálásáig nem találtunk megengedett megoldást.

3. Valamennyi többi tesztfeladatnál valamelyik megengedett megoldás elfogadása, és ily módon az algoritmus megállítása esetén a számítási idő megtakarításának aránya nagyobb, mint amekkora engedményt tettünk a célfüggvényértékre. Különösen nagy az eltérés a P2 és a P5 feladatok esetében. Megjegyezzük még, hogy a mintafeladatok *a)* esetében mindig, de a többi esetekben is többször előfordult, hogy az algoritmus hamar megtalálta az optimális megoldást, csak további lépések voltak szükségesek az optimalitás felismerésére. Ilyen esetekben természetesen a jó megengedett megoldás egyben az optimális volt; a módosított algoritmus is az optimális megoldást szolgáltatta.

IRODALOM

[1] BENDERS, J. F., "Partitioning procedures for solving mixed variables programming problems", *Numerische Mathematik* **4** (1962) 238—252.
[2] FABIAN, CS., „Verknüpfung des Dekompositionsprinzips von Benders mit dem Prinzip des Lexikographischen Suchens", Dissertation, Bonn, 1975.
[3] GEOFFRION, A. M., "Generalized Benders decomposition", *Journal of Optimization Theory and Applications* **4** (1972) 237—260.
[4] KOVÁCS, L. B., *A diszkrét programozás kombinatorikus módszerei* (Bolyai János Matematikai Társulat, Budapest, 1969) 173—202.
[5] LASDON, L. S., *Optimization for Large Systems* (The Macmillan Company, New York, 1972).
[6] PRÉKOPA, A., *Lineáris programozás I.* (Bolyai János Matematikai Társulat, Budapest, 1968).
[7] STRAZICKY, B., „PRIMAL" rutin a CDC 3300-as gépekre (Magyar Tudományos Akadémia Számítástechnikai és Automatizálási Kutató Intézete, CDC 3300 Felhasználói ismertetők 2. 1973. május) 16—18.
[8] ZOUTENDIJK, G., "Mixed integer programming and the warehouse allocation problem", *Applications of Mathematical Programming Techniques* Ed. E. M. L. Beale (The English Universities Press Ltd., London, 1971/203—215.

(Beérkezett: 1979. június 12.)

HOFFER JÁNOS
MTA SZÁMÍTÁSTECHNIKAI ÉS AUTOMATIZÁLÁSI KUTATÓ INTÉZET
1502 BUDAPEST XI. KENDE U. 13—17.

BENDERS' PARTITIONING PROCEDURE COMPLETED BY THE EXAMINATION OF FEASIBLE SOLUTIONS

J. HOFFER

In this paper we present a modification of the *Benders' partitioning procedure*. For a maximization problem the method produces a series of feasible solutions with increasing objective function values. Two numerical examples and some numerical experiences are given for the illustration of the method.

Alkalmazott Matematikai Lapok 5 (1979) 191—202

MÓDSZEREK A RENDEZÉSI ALGORITMUSOK ELVI KORLÁTAINAK MEGHATÁROZÁSÁRA

VARECZA ÁRPÁD
Nyíregyháza/Budapest

Dolgozatunkban adott véges teljesen rendezett halmaz bizonyos tulajdonságú elemeinek kiválasztására szolgáló algoritmusokat ismertetünk. Célunk a rendezési algoritmusok elvi korlátainak meghatározására szolgáló módszerek bemutatása. A 3. fejezetben szereplő legnagyobb és legkisebb elem kiválasztásához szükséges lépésszám bizonyítására úgynevezett stratégia javításos módszert alkalmazunk. Igaz, így a bizonyítás kicsit hosszabb, mint [12]-ben, de véleményünk szerint általános módszerként jobban alkalmazható.

1. Bevezetés

A gyakorlatban előforduló gyakori probléma, hogy egy számítógép memóriájában tárolt számadatokat nagyság szerint sorba kell rendezni, vagy csak a legnagyobbat kell közülük megkeresni. Az is előfordulhat, hogy éppen a középen levő elemet kell megtalálnunk. Általában csupán egyetlen műveletet használhatunk erre: kiválasztunk közülük egy párt és azokat összehasonlítjuk. A matematikai cél nyilvánvaló. Olyan módszert kell kidolgozni, amely bizonyos értelemben a leggyorsabban vezet célhoz.

Egy másik, talán tréfásnak ható problémakör a sportversenyek problémája, amely hasonló matematikai modellhez vezet.

Adott néhány sportoló, akik páronként mérik össze erejüket, s meghatározandó például a legjobb három (feltéve persze, hogy a sportolók között van egy fix sorrend és nincs semmi véletlen, a jobb legyőzi a rosszabbat). A probléma komolyabbnak tűnik, ha meggondoljuk azt, hogy a pingpongversenyek kieséses formája nem biztosítja azt, hogy az ezüstérmes legyen a második legjobb. Vagy a sakkversenyek lebonyolítása táblázatok szerint megy végbe, mert ha az első néhány fordulóban csak úgy „össze-vissza" párosítjuk a versenyzőket, akkor a versenyt valószínűleg nem lehet úgy folytatni, hogy mindenki mindenkivel pontosan egyszer játsszék és minden fordulóban mindenki asztalhoz üljön (tegyük fel, hogy páros számú sakkozó van).

Fogalmazzuk meg ezek után a problémát a matematika nyelvén.

Legyen adott egy H n számosságú véges teljesen rendezett halmaz, amelynek rendezését nem ismerjük. Ki akarjuk választani a H halmazból a H elemeinek páronkénti összehasonlításával H bizonyos tulajdonságú elemét, vagy elemeit. Például a legnagyobbat, a legnagyobbat és az utána következőt stb. Kérdés az, hogy ezen elemek kiválasztásához legalább mennyi összehasonlítást kell végeznünk. Ezen problémakörben csak néhány pontos eredmény ismeretes. Legtöbb esetben csak alsó és felső korlátok bizonyítottak. A következőkben főleg olyan problémák fognak szerepelni, amelyeknél pontos értékek ismertek. Cikkünk áttekintő jellegű, de koránt-

sem törekszik teljességre, inkább csak az érdeklődés felkeltésére a szép és egyszersmind hasznos témakör iránt.

Mielőtt az egyes problémák vizsgálatával foglalkoznánk, definiáljuk matematikailag precízen a *kiválasztásra szolgáló stratégiát* úgy, hogy az minden továbbiakban szereplő problémánál alkalmazható legyen, s meghatározzuk a minimalizálandó célt is.

2. A stratégia

Tegyük fel, hogy ki akarjuk választani a H halmaz bizonyos tulajdonságú elemeit. Az első összehasonlítandó elempárt, mondjuk (a, b)-t, jelölje S_0 és ε_1 legyen 1 vagy 0 aszerint, hogy $a>b$ vagy $a<b$. Az ε_1 választól függően választunk egy $S_1(\varepsilon_1)$ párt, mondjuk $c(\varepsilon_1)$-t és $d(\varepsilon_1)$-et és ε_2-t 1-nek definiáljuk, ha $c(\varepsilon_1)>d(\varepsilon_1)$ különben pedig 0-nak. Ugyanígy folytatva, bizonyos

$$\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{k-1}$$

sorozatokra ($\varepsilon_i=0$ vagy 1, $i=1, \ldots, k-1$) megadjuk a

$$S_{k-1}(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{k-1}) \tag{2.1}$$

párt azzal a kikötéssel, hogy ha az

$$S_{k-1}(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{k-1})$$

definiálva van, akkor az

$$S_{k-2}(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{k-2})$$

szintén definiálva van. Az ε_k értéke 1 vagy 0 aszerint, hogy a (2.1) pár első vagy második tagja nagyobb. A kérdéseknek így — a közbeeső válaszok függvényeként — megadott sorozatát a H halmaz *bizonyos tulajdonságú elemeinek meghatározására szolgáló stratégiának* (a továbbiakban csak röviden stratégiának) *nevezzük,* ha minden

$$\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_l$$

sorozat esetén, amikor az

$$\begin{aligned} &S_{l-1}(\varepsilon_1, \ldots, \varepsilon_{l-1}) \quad \text{meg van határozva, de} \\ &S_l(\varepsilon_1, \ldots, \varepsilon_l) \quad \text{már nincs,} \end{aligned} \tag{2.2}$$

akkor az

(2.3)

$\varepsilon_1, \ldots, \varepsilon_l$ válaszok (az $S_0, S_1(\varepsilon_1), \ldots, S_{l-1}(\varepsilon_1, \ldots, \varepsilon_{l-1})$ kérdésekkel együtt)

egyértelműen meghatározzák a H keresett elemeit.

Jelölje $T_k(\varepsilon_1, \ldots, \varepsilon_{k+1})$ azt az egyenlőtlenséget, amit az $S_k(\varepsilon_1, \ldots, \varepsilon_k)$ párból készíthetünk az ε_{k+1} válasz alapján. A (2.3) feltételt úgy is fogalmazhatjuk, hogy a

$$T_0(\varepsilon_1), T_1(\varepsilon_1, \varepsilon_2), \ldots, T_{l-1}(\varepsilon_1, \ldots, \varepsilon_l) \tag{2.4}$$

egyenlőtlenségek egyértelműen meghatározzák a keresett elemeket.

Például a legnagyobb elem keresése esetén egy kivételével minden elemről belátható, hogy van nála nagyobb a (2.4) egyenlőtlenségek alapján.

Itt felmerül az a kérdés, hogy milyen H elemei közötti egyenlőtlenségek következnek a (2.4) egyenlőtlenségekből. Azaz adottak valamilyen $x_1, x_2, \ldots, x_n$ változók és ezek között adott egy

$$x_j < x_l \quad (1 \leqq j,\ l \leqq n,\ j \neq l) \tag{2.5}$$

alakú egyenlőtlenségekből álló egyenlőtlenségrendszer. Ennek kell tekintenünk a megoldásait egy adott n elemű H rendezett halmazból, vagy egy, legalább n elemű H'-ből (gyakori eset az, amikor H' a valós számok halmaza). Erre ad választ a következő lemma.

2.1. LEMMA: Ha a (2.5) egyenlőtlenségrendszerben nincs irányított kör és x_{i_j}, x_{i_k} között nincs irányított út, akkor (2.5)-nek van olyan megoldása is H'-ben, amelyre $x_{i_j} < x_{i_k}$ és olyan is, amelyre $x_{i_j} > x_{i_k}$.

A lemma bizonyításával itt nem foglalkozunk, mert bár egyszerű, viszonylag hosszadalmas és nem is igen tartozik témánkhoz. A témakör irodalmában említés nélkül is természetesnek veszik. Véleményünk szerint azonban a bizonyítások többségében ki van használva, így legalábbis említésre érdemes. A lemma szerint tehát a (2.5) egyenlőtlenségekből csak olyan egyenlőtlenségek következnek, amelyek a meglevő egyenlőtlenségekből azok egymás utáni alkalmazásával a szokásos értelemben következnek.

Az (2.2), (2.3) feltételeket kielégítő $\varepsilon_1, \ldots, \varepsilon_l$ sorozatra azt mondjuk, hogy erre a *stratégia befejeződik*. A *stratégia hosszán* a maximális hosszúságú olyan sorozat hosszát értjük, amelyre a stratégia befejeződik. A továbbiakban a stratégiát $\mathcal{S}$ és a stratégia hosszát $L(\mathcal{S})$ fogja jelölni. Az $\mathcal{S}$ stratégia egy $(\varepsilon_1, \ldots, \varepsilon_i)$ állapotán az $S_{i-1}(\varepsilon_1, \ldots, \varepsilon_{i-1})$ kérdésre adott válasz utáni helyzetet fogjuk érteni. A rövidebb írásmód kedvéért — ha az félreértésre nem ad okot — az ε-okat elhagyjuk. A vizsgálatok célja az lesz minden problémánál, hogy megkeressük azt a stratégiát(-kat), amelyre $L(\mathcal{S})$ minimális.

Természetesen minimalizálhatnánk például a maximális kérdésszám helyett az átlagos kérdésszámot is. Sokszor a gyakorlatban ez is fontos. Az irodalomban több ilyen cikk is szerepel, de bizonyításuk jellege más, ezért itt nem foglalkozunk velük.

3. A legnagyobb, a legnagyobb és legkisebb, a legnagyobb és az utána következő elemek kiválasztása

Először a H legnagyobb elemének meghatározásával foglalkozunk.

Legyen $\mathcal{S}$ egy erre szolgáló stratégia. Ha az $(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_l)$ állapotban $\mathcal{S}$ befejeződik, akkor definíció szerint

$$T_0(\varepsilon_1), T_1(\varepsilon_1, \varepsilon_2), \ldots, T_{l-1}(\varepsilon_1, \ldots, \varepsilon_l) \tag{3.1}$$

egyenlőtlenségek egyértelműen meghatározzák a legnagyobb elemet, azaz egy elem kivételével minden elemről belátható, hogy van nála nagyobb. A 2.1. lemma szerint ez csak úgy fordulhat elő, ha a bizonyos elem kivételével minden elem előfordul (3.1)-ben kisebbként, azaz a (3.1)-ben legalább $n-1$ egyenlőtlenség szerepel. Ennek alapján kimondható a következő ([9], [12])

3.1. TÉTEL: Ha $\mathscr{S}$ a legnagyobb elem meghatározására alkalmas tetszőleges stratégia, akkor

$$L(\mathscr{S}) \geqq n-1.$$

Figyeljük meg, hogy a tétel állításánál többet bizonyítottunk. Nemcsak azt, hogy a maximális lépésszám $n-1$, hanem azt, hogy sohase fejeződhet be a stratégiánk előbb. Ez persze a bonyolultabb problémák esetén már nem lesz igaz. Vegyük még észre, hogy a tétel pontos. A pingpongozók „kieséses versenye" adja azt a stratégiát, amelynek hossza $n-1$.

A bizonyításunk kicsit formálisnak tűnhet a sok jelöléssel, hiszen ha azt mondjuk egyszerűen, hogy „a győztes kivételével mindenkinek egyszer alul kell maradnia, tehát legalább $n-1$ mérkőzés kell", akkor majdnem pontosak vagyunk. A jelöléseket főleg azért használtuk itt, hogy az olvasó hozzászokjon. A „precízkedő" jelölésekre a későbbiekben szükség lesz. Hogy mennyire, azt az mutatja, hogy két legnagyobb elem problémájára két — enyhén szólva — nem teljes bizonyítás is megjelent ([15], [16]).

A következőkben a H halmaz maximális és minimális eleme egyszerre való meghatározásával foglalkozunk.

Legyen $\mathscr{S}$ a H halmaz maximális és minimális eleme egyszerre való meghatározásának egy tetszőleges stratégiája. Tegyük fel, hogy valamilyen $(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_i)$ állapotban van az $\mathscr{S}$ stratégia és ott az $S_i(\varepsilon_1, \ldots, \varepsilon_i)=(a, b)$ pár kerül összehasonlításra, ahol a már szerepelt nagyobbként, azaz a

$$T_0(\varepsilon_1), T_1(\varepsilon_1, \varepsilon_2), \ldots, T_{i-1}(\varepsilon_1, \ldots, \varepsilon_i) \tag{3.2}$$

egyenlőtlenségek egyike $a>c$ valamilyen c-vel és tegyük még fel, hogy b szerepelt már kisebbként, azaz (3.2)-ben szerepel $b<d$ valamilyen d-vel. (Feltehető, hogy $b\neq c$, $d\neq a$ azaz a négy elem közül legfeljebb a c és d lehet egyenlő.) Tekintsük az $\mathscr{S}$ stratégia következő módosítását.

$$S'_j(\delta_1, \ldots, \delta_j) = \begin{cases} S_j(\delta_1, \ldots, \delta_j), & \text{ha } j < i, \text{ vagy } (\delta_1, \ldots, \delta_i) \neq (\varepsilon_1, \ldots, \varepsilon_i) \\ S_j(\varepsilon_1, \ldots, \varepsilon_i, 1, \delta_{i+1}, \ldots, \delta_j), & \text{ha } (\delta_1, \ldots, \delta_i) = (\varepsilon_1, \ldots, \varepsilon_i) \end{cases} \tag{3.3}$$

ahol $\delta_1, \ldots, \delta_j$ egy olyan 0, 1 sorozat, hogy a jobb oldal értelmezve legyen. Igazoljuk a következőt:

3.1. LEMMA: A (3.3)-mal definiált kérdéssorozat stratégia.

Bizonyítás: Azt kell igazolni, hogy (2.2)-ből következik (2.4). Esetünkben ez azt jelenti, hogy a

$$T'_0(\delta_1), T'_1(\delta_1, \delta_2), \ldots, T'_{i-1}(\delta_1, \ldots, \delta_i) \tag{3.4}$$

egyenlőtlenségrendszer egyértelműen meghatározza a legnagyobb és legkisebb elemeket. A 2.1. lemma alapján ez akkor és csakis akkor áll fenn, ha az egyenlőtlenségekben minden elem szerepel egy-egy kivétellel nagyobbként is és kisebbként is. Nevezzük az ilyen egyenlőtlenségrendszert a továbbiakban „jó"-nak.

Ha $l<i$ vagy $(\delta_1, \ldots, \delta_i) \neq (\varepsilon_1, \ldots, \varepsilon_i)$, akkor nyilvánvaló, mert (3.4) írható vesszők nélkül is változtatás nélkül, és $\mathscr{S}$ stratégia. Tegyük fel tehát, hogy $(\delta_1, \ldots, \delta_i) = (\varepsilon_1, \ldots, \varepsilon_i)$. A (3.4) akkor (3.3) alapján a

$$\begin{aligned} &T_0(\varepsilon_1), \ldots, T_{i-1}(\varepsilon_1, \ldots, \varepsilon_i), \\ &T_{i+1}(\varepsilon_1, \ldots, \varepsilon_i, 1, \delta_{i+1}), \ldots, \\ &T_l(\varepsilon_1, \ldots, \varepsilon_i, 1, \delta_{i+1}, \ldots, \delta_l) \end{aligned} \tag{3.5}$$

alakot ölti. Lássuk be, hogy (3.5) jó egyenlőtlenségrendszer. Mivel $\mathscr{S}$ stratégia, a befejező

$$\varepsilon_1, \ldots, \varepsilon_i\ 1, \delta_{i+1}, \ldots, \delta_l$$

sorozatra

$$T_0(\varepsilon_1), \ldots, T_{i-1}(\varepsilon_1, \ldots, \varepsilon_i), \quad T_i(\varepsilon_1, \ldots, \varepsilon_i, 1),$$

$$T_{i+1}(\varepsilon_1, \ldots, \varepsilon_i, 1, \delta_{i+1}), \ldots, T_l(\varepsilon_1, \ldots, \varepsilon_i, 1, \delta_{i+1}, \ldots, \delta_l)$$

egyenlőtlenségrendszer jó.

Igazolnunk tehát csak azt kell, hogy a $T_i(\varepsilon_1, \ldots, \varepsilon_i, 1)$ azaz $a>b$ elhagyása ezen nem ront. Viszont a feltevés szerint a már korábban szerepelt nagyobbként, b pedig kisebbként, tehát $a>b$ elhagyása bajt nem okozhat. A többi elemre vonatkozó egyenlőtlenségen pedig az elhagyás nem változtat. A bizonyítást befejeztük. Vegyük észre, hogy a 3.1. lemma akkor is érvényes, ha a szerepel korábban kisebbként és b nagyobbként.

Kissé más a helyzet, ha csak a-ról tesszük fel, hogy már szerepelt nagyobbként, b-ről pedig azt tudjuk, hogy még eddig nem szerepelt összehasonlításban. Ekkor feltesszük valamilyen további $e(\neq b)$ elemről is, hogy még nem szerepelt és egy kissé más transzformációt alkalmazunk:

$$S_j^*(\delta_1, \ldots, \delta_j) = \begin{cases} S_j(\delta_1, \ldots, \delta_j), & \text{ha } j < i, \text{ vagy } (\delta_1, \ldots, \delta_i) \neq (\varepsilon_1, \ldots, \varepsilon_i) \\ (e, b), & \text{ha } (\delta_1, \ldots, \delta_i) = (\varepsilon_1, \ldots, \varepsilon_i) \text{ és } S_j(\varepsilon_1, \ldots, \varepsilon_i, 1, \\ \delta_{i+2}, \ldots, \delta_j), & \text{ha } (\delta_1, \ldots, \delta_i) = (\varepsilon_1, \ldots, \varepsilon_i) \text{ és } \delta_{i+1} = 1, \\ S_j(\varepsilon_1, \ldots, \varepsilon_i, 1, \delta_{i+2}, \ldots, \delta_j)\text{-ben } b \text{ és } e \text{ szerepe felcserélve,} \\ \text{ha } (\delta_1, \ldots, \delta_i) = (\varepsilon_1, \ldots, \varepsilon_i) \text{ és } \delta_{i+1} = 0 \end{cases} \tag{3.6}$$

ahol $\delta_1, \ldots, \delta_j$ egy olyan 0, 1 sorozat, melyre a jobb oldal értelmezve van. Igazoljuk a következőt:

3.2. LEMMA: A (3.6) képlettel definiált $\mathscr{S}^*$ kérdéssorozat stratégia.

A 3.2. lemma bizonyítása lényegében megegyezik a 3.1. lemmáéval. Figyeljük meg, hogy a 3.2. lemma akkor is igaz, ha a korábban kisebbként szerepelt, vagy ha a és b sorrendjét felcseréljük.

A H halmaz elemeit az $\mathscr{S}$ stratégia egy tetszőleges $(\varepsilon_1, \ldots, \varepsilon_i)$ állapotában négy részre oszthatjuk:

$B(\varepsilon_1, \ldots, \varepsilon_i)$ álljon azon elemekből, amelyek már kisebbként is, nagyobbként is előfordultak;
$A_1(\varepsilon_1, \ldots, \varepsilon_i)$ azokból, amelyek nagyobbként igen, kisebbként nem;
$A_2(\varepsilon_1, \ldots, \varepsilon_i)$ azokból, amelyek kisebbként igen, nagyobbként nem;
$A_3(\varepsilon_1, \ldots, \varepsilon_i)$ azokból, amelyek még egyáltalán nem szerepeltek.

Legtöbbször az $(\varepsilon_1, \ldots, \varepsilon_i)$ félreértés veszélye nélkül elhagyható. Egy $S_i(\varepsilon_1, \ldots, \varepsilon_i)$ párról azt mondjuk, hogy pl. (A_2, B) típusú, ha a pár egyik eleme (első vagy második) $A_2(\varepsilon_1, \ldots, \varepsilon_i)$-beli, a másik pedig $B(\varepsilon_1, \ldots, \varepsilon_i)$-beli. Hasonlóan értendő a többi típus is.

3.3. LEMMA: Ha $\mathcal{S}$ a maximális és minimális elemek meghatározására szolgáló stratégia, akkor van ugyanerre egy $\mathcal{S}_1$ stratégia, melyre $L(\mathcal{S}_1) \leqq L(\mathcal{S})$ és csak $(A_1, A_1), (A_2, A_2), (A_3, A_3)$ típusú összehasonlítást tartalmaz, illetve (B, A_3), $(A_1, A_3), (A_2, A_3)$ típusút csak akkor, ha $A_3(\varepsilon_1, \ldots, \varepsilon_i)$ egyelemű.

Bizonyítás: Könnyen látható, hogy a 3.1. és 3.2. lemmában szereplő transzformációk a stratégia hosszát nem növelik:

$$L(\mathcal{S}'), L(\mathcal{S}^*) \leqq L(\mathcal{S}).$$

Elindulva $\mathcal{S}$ egy ágán, az első olyan állapotnál, amelyben $(B, B), (B, A_1), (B, A_2)$ vagy (A_1, A_2) típusú összehasonlítás szerepel, alkalmazzuk a 3.1. lemma transzformációját. Könnyen látható, hogy véges sok ilyen lépéssel megszabadulhatunk az összes ilyen típusúaktól.

Ezután ugyanezzel a módszerrel, de a 3.2. lemma alkalmazásával megszüntetjük a $(B, A_3), (A_1, A_3), (A_2, A_3)$ ($|A_3| > 1$) típusúakat is. A lemma bizonyításának vázlatát ezzel befejeztük.

Ezek után könnyen igazolhatjuk a következőt:

3.2. TÉTEL: Ha $\mathcal{S}$ a véges n-elemű H halmaz legnagyobb és legkisebb eleme egyszerre való meghatározásának stratégiája, akkor

$$L(\mathcal{S}) \geqq n + \left[\frac{n}{2}\right] - 2.$$

($[x]$ jelöli az x-nél nem kisebb legkisebb egészet.)

Bizonyítás: Vegyük a 3.3. lemma által biztosított $\mathcal{S}_1$ stratégiát. A tételt elég erre bizonyítanunk. Tegyük fel, hogy $\varepsilon_1, \ldots, \varepsilon_i$-re $\mathcal{S}_1$ befejeződik, azaz

$$T_0(\varepsilon_1), \ldots, T_{i-1}(\varepsilon_1, \ldots, \varepsilon_i) \tag{3.7}$$

egyenlőtlenségrendszer jó.

Tegyük fel, hogy n páros és tekintsük (3.7)-ben az (A_3, A_3) típusú összehasonlításokat. H minden eleme pontosan egyszer fordul elő bennük, hiszen kétszer nyilvánvalóan nem fordulhat elő: másrészt $\mathcal{S}_1$ definíciója szerint egy A_3-ba tartozót nem lehet összehasonlítani másfajtával, míg legalább kettő van belőle, ezért mindegyik legalább egyszer előfordul. Így az (A_3, A_3) típusú összehasonlítások száma $\frac{n}{2}$. Jelölje $a_1, \ldots, a_{\frac{n}{2}}$ azokat, amelyek ezen összehasonlításokban nagyobbaknak; $b_1, \ldots, b_{\frac{n}{2}}$ azokat, amelyek kisebbeknek bizonyultak. $\mathcal{S}_1$ definíciója miatt az a-k és b-k nem kerülhetnek összehasonlításra. Viszont (3.7)-ben egy kivételével minden a elem szerepel kisebbként és egy kivételével minden b szerepel nagyobbként és ezen kétfajta egyenlőtlenségek nem eshetnek össze. Ezek száma tehát legalább $2\left(\frac{n}{2} - 1\right)$. A (3.7) egyenlőtlenségeinek száma tehát legalább $\frac{3n}{2} - 2$.

Ha n páratlan, az (A_3, A_3) típusú összehasonlítások száma $\frac{n-1}{2}$. Jelölje a bennük nagyobb, ill. kisebb elemeket $a_1, \ldots, a_{\frac{n-1}{2}}$, illetve $b_1, \ldots, b_{\frac{n-1}{2}}$. Az egyetlen elemet, ami nem fordul elő bennük jelölje c. Az $a_1, \ldots, a_{\frac{n-1}{2}}, c$ közül egy kivételével mindegyik szerepel kisebbként, és $b_1, \ldots, b_{\frac{n-1}{2}}, c$ közül pedig nagyobbként, és ezen egyenlőtlenségek között nincs közös. Tehát a számuk legalább $2\frac{n-1}{2}$, azaz az egyenlőtlenségek száma legalább $3\frac{n-3}{2}$.

A tételt bebizonyítottuk.

A 3.2. tétel [12]-ben szereplő bizonyítása rövidebb, az itt közölt bizonyítás viszont alkalmasabbnak tűnik általánosabb módszerként.

A 3.2. tétel becslése pontos, azaz van olyan stratégia, amelynek hossza ennyi: párosítsuk a „versenyzőket". A győztesek és a vesztesek is (beleértve az esetleg egy kimaradt páratlant is) játszanak „kieséses versenyt".

A következő probléma, amivel foglalkozunk igen régi. Még STEINHAUS vetette fel 1930-ban. Keresendő a legrövidebb algoritmus, amely képes meghatározni a rendezett halmaz első két elemét páronkénti összehasonlítások alapján.

Ha a pingpongversenyek kieséses stratégiáját alkalmazzuk, akkor jól látható, hogy legfeljebb $[\log_2 n]$ „forduló" van, azaz az első legfeljebb $[\log_2 n]$ játékost győz le közvetlenül. Másodikként csak ezek jöhetnek számításba, tehát ezek egymás között döntik el, ki közöttük a legjobb, azaz az abszolút második. Ehhez, mint tudjuk $[\log_2 n]-1$ játék elegendő. Így tehát olyan stratégiát már ismerünk, amely $n+[\log_2 n]-2$ lépésben elvégzi feladatát. A következő tétel azt mondja ki, hogy ennél jobb stratégia nincs is.

3.3. TÉTEL: Ha $\mathcal{S}$ egy olyan stratégia, amely egy n-elemű rendezett halmaz első két elemét határozza meg, akkor

$$L(\mathcal{S}) \geqq n+[\log_2 n]-2.$$

A 3.2. tételre alkalmazott módszerünk, amelyet *stratégia—javításos* módszernek nevezhetünk, itt is alkalmazható, csak kissé hosszadalmas lenne. Ezért továbbá, hogy egy újabb módszert mutassunk be, KNUTH ([10.211—212]) bizonyítását adjuk itt meg: Ennek a módszernek a lényege az, hogy ügyesen megadjuk az összehasonlításokra adandó válaszokat előre úgy, hogy a legrosszabb ágba kerüljön a stratégia, és csak ennek az ágnak a hosszát becsüljük meg alulról. A stratégia úgy fog működni, mintha az „ellenségünk" mondaná a válaszokat, ezért nevezik az irodalomban ezt a módszert *„ellenséges"* módszernek.

Itt jegyezzük meg, hogy a [15] és [16] bizonyítások hiányosak, az első helyes (de hosszadalmas) bizonyítás KISLINCITŐL [9] származik.

Térjünk rá ezek után a tétel bizonyítására.

Bizonyítás:

1. Először azt bizonyítjuk be, hogy egy tetszőleges stratégiában, amely az első elemet választja ki, van olyan ág, melyben a legnagyobb elem legalább $[\log_2 n]$-szer

kerül összehasonlításra. Induljunk ki az S_0-ból. Tegyük fel, hogy már megadtuk az $\varepsilon_1, \ldots, \varepsilon_i$ válaszsorozatot és az

$$S_i(\varepsilon_1, \ldots, \varepsilon_i) = (c, d)$$

összehasonlításra kerül sor. Legyen az erre adott ε_{i+1} válasz 1, ha a

$$T_0(\varepsilon_1), T_1(\varepsilon_1, \varepsilon_2), \ldots, T_{i-1}(\varepsilon_1, \ldots, \varepsilon_i) \tag{3.8}$$

egyenlőtlenségekben d szerepel kisebbként, de c nem, vagy ha egyikük sem szerepel kisebbként, de c többször szerepel nagyobbként. Legyen $\varepsilon_{i+1}=0$, ha a fentiek c és d fölcserélésével teljesülnek, és legyen minden egyéb esetben ε_{i+1} tetszőleges, de olyan, hogy ne kerüljön ellentmondásba (3.8)-cal. Könnyű meggondolni, hogy az utóbbi mindig megtehető, és hogy az előbbi, egyértelműen meghatározott ε_{i+1}-ek sem vezetnek ellentmondáshoz. Ha az $\mathcal{S}$ stratégia így meghatározott ágán végig megyünk, akkor a

$$T_0(\varepsilon_1), T_1(\varepsilon_1, \varepsilon_2), \ldots, T_{l-1}(\varepsilon_1, \ldots, \varepsilon_l) \tag{3.9}$$

egyenlőségrendszerhez jutunk. Erről fogjuk belátni, hogy a legnagyobb elem (az egyetlen tehát, amelyik kisebbként szerepel benne), legalább $[\log_2 n]$ egyenlőtlenségben szerepel.

Tekintsük azokat az egyenlőtlenségeket (3.8)-ban, amelyekben valamelyik elem (ebben a sorrendben) először fordul elő kisebbként. Jelölje ezek halmazát $\mathcal{E}_i$. A következő lemmát fogjuk p-re vonatkozó teljes indukcióval bizonyítani.

3.4. LEMMA: Ha egy a elem (3.8)-ban kisebbként nem szerepel, de nagyobbként p-szer, akkor $\mathcal{E}_i$ alapján legfeljebb 2^p darab b-re állíthatjuk, hogy $b<a$ vagy $b=a$.

Bizonyítás: $p=0$-ra az állítás triviális. Tegyük fel, hogy $p>1$ és $p-1$-re az állítás igaz.

Igazoljuk, hogy akkor p-re is. Legyen $T_{j-1}(\varepsilon_1, \ldots, \varepsilon_j)$ az utolsó (3.8)-beli egyenlőtlenség, amelyben a előfordul. Legyen pl. $a>c$.

Most be fogjuk látni, hogy $\mathcal{E}_i$ alapján nem vezethető le több $b<a$, mint $\mathcal{E}_j$ alapján. Legyen ugyanis $b<a$ levezethető $\mathcal{E}_i$-ben, azaz léteznek b-k, hogy

$$b = b_0 < b_1 < \ldots < b_k = a$$

és legyen l a legkisebb egész, amelyre

$$b_l < \ldots < b_k = a$$

még megtalálható $\mathcal{E}_j$-ben. Probléma csak akkor van, ha $l>0$. Ekkor $b_{l-1}<b_l$-nek $\mathcal{E}_i-\mathcal{E}_j$-ben kell lenni. Ha $l=k$, akkor ez ellentmond annak, hogy $T_{j-1}(\varepsilon_1, \ldots, \varepsilon_j)$ a utolsó előfordulása. Ha pedig $0<l<k$, akkor viszont b_l előbb szerepel kisebbként, mint $b_{l-1}<b_l$ (ami b_{l-1} első kisebbsége) s ez ellentmond az ε-ok „ellenséges" megválasztásának.

Ha $T_{j-1}(\varepsilon_1, \ldots, \varepsilon_j)$ nincs $\mathcal{E}_j$-ben, akkor $\mathcal{E}_i$ helyett mindjárt $\mathcal{E}_{j-1}$-et vehetünk, a csak $p-1$-szer szerepel nagyobbként és így ebben az indukciós feltevés miatt legfeljebb 2^{p-1} $b<a$ vagy $b=a$ vezethető le.

Tegyük fel tehát, hogy $T_{j-1}(\varepsilon_1, \ldots, \varepsilon_j)$ benne van $\mathcal{E}_i$-ben, azaz c még nem volt kisebb. Az ε-ok konstrukciója miatt tehát a c legfeljebb $p-1$-szer volt nagyobb ez előtt.

Hány $b<a$ vezethető le $\mathscr{E}_j$-ben $c<a$ használatával? Annyi, ahány $b<c$ vagy $b=c$ $\mathscr{E}_{j-1}$-ben.

Ezek száma az indukciós feltevés miatt legfeljebb 2^{p-1}. Másrészt $c<a$ használata nélkül $\mathscr{E}_{j-1}$-ben legfeljebb 2^{p-1} darab $b<a$ vagy $b=a$ vezethető le. Azaz az $\mathscr{E}_j$-ben levezethető $b<a$ vagy $b=a$ összefüggések száma legfeljebb 2^p. A lemma bizonyítását befejeztük.

(3.9)-ben egy kivételével — jelöljük ezt a-val — minden b elemhez található olyan egyenlőtlenségsorozat, amelyek egymás után alkalmazásával $b<a$ nyerhető. Igaz ez $\mathscr{E}_l$-ben is, mert b $\mathscr{E}_l$-ben is előfordul kisebbként: $b<b_1$, de b_1 is $b_1<b_2$ és így tovább el kell jutnunk egy olyan b_k elemhez, ami $\mathscr{E}_l$-ben nem szerepel kisebbként, de akkor (3.9)-ben sem, azaz $b_k=a$. Ha tehát a (3.9)-ben p-szer szerepel, akkor a lemma alapján $2^p \geqq n$, vagyis $p \geqq [\log_2 n]$. A bizonyítás első pontját befejeztük.

2. Legyen most $\mathscr{S}$ egy olyan stratégia, amely az első két elemet meghatározza. Akkor egyben az elsőt is, tehát alkalmazható rá az első rész eredménye, hogy van egy olyan ág ((3.9)), melyben az első elem legalább $[\log_2 n]$-nel van összehasonlítva. Ezek közül egy kivételével mindegyiknek legalább még egyszer kell kisebbnek bizonyulnia. (3.9)-ben, tehát egy kivételével minden elem legalább egyszer szerepel kisebbként és legalább $[\log_2 n]$ elem legalább kétszer szerepel. Vagyis (3.9)-ben az egyenlőtlenségek száma legalább $n-1+[\log_2 n]-1$. A tétel bizonyítását befejeztük.

A három legnagyobb elem megkereséséhez szükséges algoritmus hosszának minimumára van egy bonyolult formula ([8]), azonban általában a t első elem esetére csak becslések vannak. KISLICIN ([9]) bizonyította, hogy van olyan algoritmus, melyre

$$L(\mathscr{S}) \leqq n-t+\sum_{n+1-t<j\leqq n} [\log_2 j] \quad (n \geqq t)$$

HADIAN és SOBEL ([4]) pedig olyan algoritmust találtak, amely a t-edik elemet találja meg legfeljebb.

$$L(\mathscr{S}) \leqq n-t+(t-1)[\log_2 (n+2-t)] \tag{3.10}$$

lépésben. HYAFIL pedig az ([5])

$$n-t+(t-1)\left[\log_2 \frac{n}{t-1}\right]$$

alsó becslést adta.

4. Teljes rendezés és medián keresés

Tulajdonképpen a H halmaz bizonyos elemeinek meghatározására szolgáló stratégiák maximumának minimuma pontos értéke csak a 3. fejezetben tárgyalt esetekben ismert. A további keresési problémáknál csak alsó és felső korlátok vannak. Ebben a fejezetben megemlítünk néhányat ezek közül.

Először foglalkozunk a H halmaz elemeinek teljes rendezésével.

A $|H|=n$, ezért H elemeinek $n!$ sorrendje lehet. Ezen $n!$ sorrend közül kell kiválasztani egyet, amelyben — mondjuk — növekvő sorrendben szerepelnek H elemei. Ha ennek stratégiája $\mathscr{S}$, akkor $\mathscr{S}$ minden állapotában a számításba

jöhető sorrendek számának legfeljebb a fele kerül kizárásra. Ez azt jelenti, hogy ha az $\mathscr{S}$ stratégia l lépésben befejeződik, akkor

$$n! \leqq 2^l$$

azaz

$$\log_2 n! \leqq l.$$

Ezzel igazoltuk a következőt:

4.1. TÉTEL: Ha $\mathscr{S}$ a H halmaz elemeinek sorbarendezésének stratégiája, akkor

$$L(\mathscr{S}) \geqq [\log_2 n!] \quad (\sim n \log_2 n)$$

A fenti bizonyítást heurisztikusan úgy is mondhatjuk, hogy $n!$ lehetséges kimenetel lévén, az információtartalom $\log_2 n!$. Egy összehasonlítással egységnyi információt nyerhetünk, így a szükséges lépések száma a fenti. Ezért hívjuk ezt a módszert információelméleti módszernek.

A H halmaz elemeinek sorbarendezésére STEINHAUS a következő algoritmust adta ([2, 276—277.]).

Első lépésben összehasonlítjuk H két tetszőleges elemét, és ha már H tetszőleges k elemét sorba rendeztük, akkor veszünk egy tetszőleges $k+1$-edik elemet és összehasonlítjuk a már rendezett k elem közül a középsővel vagy valamelyik középsővel páros k esetén. Ha ennél nagyobb, akkor a középső elemtől nagyobb, ha kisebb, akkor a középső elemtől kisebb elemek középsőjével hasonlítjuk össze és így tovább. Így felezésekkel meghatározzuk a $k+1$-edik elem helyét a k elem között. Utána veszünk egy $k+2$-edik elemet és így tovább. A $k+1$-edik elem elhelyezésére így $[\log_2(k+1)]$ összehasonlítás szükséges. Az eljárást minden elemre elvégezve

$$L(\mathscr{S}) \leqq \sum_{k=2}^{n} [\log_2 k] \quad (\sim n \log_2 n)$$

adódik. FORD és JOHNSON ([3], [2, 275—277.]) a STEINHAUS-módszert továbbfejlesztve javítani tudták a felső korlátot. Megadtak olyan $\mathscr{S}$ stratégiát, amelyre

$$L(\mathscr{S}) \leqq \sum_{k=2}^{n} \left[\log_2 \left(\frac{3}{4} k\right)\right]$$

teljesül. Ez eddig a legjobb.

Mi a helyzet, ha az n elemből csupán a középső elemet, más szóval a mediánst akarjuk megkeresni? Első ránézésre nem tudunk sokkal jobbat csinálni, mint a teljes sorrendet meghatározni, amihez körülbelül $n \log_2 n$ lépés kell. (3.10)-ből is csupán $\frac{1}{2} n \log_2 n$-et kapunk. Sokáig tartott, míg ezt (nagy meglepetésre) sikerült konstansszor n lépsre leszorítani.

Az előző problémákban a stratégia megadása könnyű volt. Nehézség csak a jó alsó becslés megadásánál merült fel. Ilyen alsó becslések bizonyítására mutattunk be három különböző módszert. Itt először találkozunk azzal a helyzettel, hogy a konstrukció megtalálása is bonyolult. Az alábbi bizonyítás ([1]) indukciót használ és nemcsak a középső elem meghatározására, hanem a tetszőleges t-edik ($1 \leqq t \leqq n$) elem meghatározására is érvényes.

4.2. TÉTEL: Ha $1 \leqq t \leqq n$, akkor létezik olyan stratégia, amely az n-elemű rendezett halmaz t-edik elemét meghatározza és

$$L(\mathcal{S}) \leqq 22n.$$

Bizonyítás: Bizonyítsuk először az állítást kis n-ekre. $\sum_{k=2}^{n} [\log_2 k] \leqq n[\log_2 n]$ lépés elég a teljes rendezés meghatározásához. Ha tehát $[\log_2 n] \leqq 22$, vagyis $n \leqq 2^{22}$, akkor az állítás igaz. Ezek után n-re vonatkozó teljes indukciót alkalmazunk. A konstrukciót könnyebb lesz végezni $5(2q+1)$ alakú számokra, tegyük tehát fel, hogy az állítás bizonyítva van $5(2(q-1)+1)$-ig minden n-re (és minden t-re). Először bebizonyítjuk az állítást $n=5(2q+1)$-re, majd utána a közbenső számokra is.

Osszuk a H halmaz elemeit 5-ös csoportokba, és határozzuk meg ezen 5-ös csoportok mediánsait. Ezt 6 összehasonlítással megtehetjük, de kevesebbel nem ([9]). Így $6(2q+1)$ összehasonlítást végzünk.

Az 5-ös csoportok mediánsainak száma $2q+1$ és ezek x mediánsának meghatározására — indukciós feltevésünk szerint — $22(2q+1)$ összehasonlítás elegendő. Az ötösök mediánsai közül így q nagyobb x-nél, q pedig kisebb. Az x-nél nagyobbaknál 2—2 elem kisebb és az x-nél kisebbeknél 2—2 elem nagyobb, így a nagyobb mediánsoknál kisebb $2q$ elem és a kisebb mediánsoknál nagyobb $2q$ elem olyan, amelyekről nem tudjuk, hogy nagyobbak vagy kisebbek x-nél. Ezt $4q$ összehasonlítással eldönthetjük. Ezek után H minden eleméről tudjuk, hogy x-nél nagyobb vagy kisebb.

Tegyük fel, hogy γ elem nagyobb x-nél. Ha $\gamma+1=t$, akkor készen vagyunk, x a t-edik elem.

Ha $t<r+1$, akkor az r elemből kell kiválasztanunk a t-ediket; és ha $t>r+1$, akkor az $n-1-r$ x-nél kisebb elemből a $t-r-1$-ediket. Ez lesz H t-edik eleme. Viszont

$$r \leqq 2q+3q+2q+2 = 7q+2$$

Mivel az x-nél nagyobb mediánsoknál kisebb $2q$ és az x-nél kisebb mediánsoknál nagyobb $2q$ elemhez vagy az x-nél nagyobb mediánsoknál nagyobb vagy egyenlő $3q$ elem jön számításba. Ugyanígy látható $n-1-r \leqq 7q+2$ is. Az indukciós feltevés szerint a legfeljebb $7q+2$ elemből legfeljebb $22(7q+2)$ összehasonlítással a t-edik (illetve $t-r-1$-edik) kiválasztható, így

$$6(2q+1)+22(2q+1)+4q+22(7q+2) = 214q+72 <$$

$$< 22 \cdot 5(2q+1) = 220q+110$$

lépés valóban elegendő. $n=5(2q+1)$ esetére az állítást bizonyítottuk. Legyen most $5(2(q-1)+1)<n<5(2q+1)$. Egészítsük ki a halmazt néhány új elemmel, hogy $5(2q+1)$ eleme legyen. A t-edik elem $214q+72$ lépéssel meghatározható. Lássuk be, hogy ez $\leqq 22n$. Mivel n lehetséges legkisebb értéke $5(2(q-1)+1)+1$ a $214q+72 \leqq 22[5(2(q-1)+1)+1] = 220q-88$-nak kell teljesülni. Ez igaz, ha $q \geqq 27$, utóbbi viszont feltehető a bizonyítás elején mondottak alapján.

A bizonyítást befejeztük.

A fenti gondolattal jobb konstans is elérhető. Az idők folyamán több különböző bizonyítás is született, különböző konstansokkal. Ebben magyarok is értek el eredményeket: PÓSA LAJOS és NAGY ZSIGMOND. A rekordot a [14] dolgozat tartja. Azt

bizonyítja, hogy van egy $\mathscr{S}$ stratégia a mediáns meghatározására, amelynek hosszára $L(\mathscr{S}) \sim 3n$ teljesül.

A téma iránt érdeklődő olvasóknak a [10] és [11] könyvek megfelelő fejezetét ajánljuk, további tanulmányozás céljából.

Végül köszönetemet fejezem ki KATONA GYULÁnak a dolgozat megírásához nyújtott segítségéért.

IRODALOM

[1] BLUM, M., PRATT, V., TARJÁN, R. and RIVEST, R., "Time bounds for selection", *J. Comp. and Sys. Sci.* 7 (1973) 448—461.
[2] BUSACKER, R. G. and SAATY, T. L., *Véges gráfok és hálózatok* (Műszaki Könyvkiadó, Budapest, 1969).
[3] FORD, L. R. and JOHNSON, S. M., "A tournament problem", *Amer. Math. Monthly* **66** (1959) 387—389.
[4] HADIAN, A. and SOBEL, M., "Selecting the *t*-th largest of items using binary comparisons", Tech. Rept. No. **121**. Dep. of Stat., Univ. of Minnesota.
[5] HYAFIL, L., "Bounds for selection", *SIAM J. Comput.* **8** (1976) 109—114.
[6] KATONA, G., "Combinatorial search problems", *A surrey of Combinatorial Theory, Ed. by. JN Srivastava et al.* (North Holland, 1973) 285—308.
[7] KIRKPATRICK, D. G., "Topics in the complexity of combinatorial algorithms", Tech. Report No. **74** (1974). Dep. of Comp. Sci. Univ. of Toronto.
[8] KIRKPATRICK, D. G., Személyes közlés *P. Ružičkától.*
[9] КИСЛИЦЫН, С. С., «О выделение *k*-ого элемента упорядоченной совокупности путем попарных сравнений», *Сиб. Мат. Ж.* V. №. 3 (1964) 557—564.
[10] KNUTH, D. E., *The Art of Computer Programming, Vol.* 3, *Sorting and Searching* (Addison—Wesley, New York, 1973).
[11] LOVÁSZ, L. és GÁCS, P., *Algoritmusok* (Műszaki Könyvkiadó, Bp., 1978).
[12] POHL, I., "A sorting problem and its complexity", *Com. of the ACM* **15** (1972).
[13] PRATT, V. and YAO, F., "On lower bounds for computing the *i*-th largest element". *Proc. 14th Ann. IEEE Symp. on Switching and Automata Theory* (1973) 70—81.
[14] SCHÖNHAGE, A., PATERSON, M. and PIPPENGER, N., "Finding the median", *Theory of Comp. Report No.* **6** (1975).
[15] SLUPECKI, J., "On the system of tournaments", *Wroclaw* (1951) *Coll. Math.* 286—290.
[16] SCHREIER, J., *Math. Pols.* **7** (1932) 154—160.

(Beérkezett: 1979. január 16.)

VARECZA ÁRPÁD
BESSENYEI GYÖRGY TANÁRKÉPZŐ FŐISKOLA
4400 NYÍREGYHÁZA PF. 166

MTA MATEMATIKAI KUTATÓ INTÉZET
1053 BUDAPEST V., REÁLTANODA U. 13—15.

METHODS FOR DETERMINATION OF PRINCIPLE BOUNDS OF ORDERING ALGORITHMS

Á. VARECZA

We care for several methods for determination of principle bounds of ordering algorithms. We utilize a correcting method for finding maximum and minimum elements of totally ordered set in the 3rd chapter. This proof is a little longer the one in [12]. However one can applay this proof given in my paper as general method better.

Alkalmazott Matematikai Lapok **5** (1979) 203—209

A külföldi szakirodalomból

OPTIMALIZÁLÁS ÉS DIFFERENCIÁLIS JÁTÉKOK

PONTRJAGIN, L. SZ.

Matematikusok esetében az a kérdés, hogy mivel foglalkozzanak, talán élesebben jelentkezik, mint más tudományterületek művelőinél. A tisztán alkalmazott tudományként keletkezett matematika alapvető feladata a jelenlegi korszakban is a minket körülvevő anyagi világ tanulmányozása abból a célból, hogy az emberiség szükségleteinek megfelelően felhasználjuk. Ugyanakkor a matematika fejlődésének megvan a maga belső logikája, és ezt követve a matematikusok olyan fogalmakat, sőt egész fejezeteket hoznak létre, amelyek tisztán szellemi tevékenység termékei, nincs kapcsolatuk a minket körülvevő anyagi valósággal és jelenleg semmiféle alkalmazásuk sincs. Ezeknek a fejezeteknek gyakorta nagyfokú harmóniájuk és bizonyos fajta szépségük van. Azonban ez a fajta szépség nem szolgálhat létezésük igazolására. A matematika nem zene, amelynek szépsége sok ember által hozzáférhető. A matematikai szépséget csupán néhány szakember tudja értékelni. Ily szépségek létrehozásán munkálkodva a matematikusok gyakorlatilag saját maguk számára dolgoznak. Nem lehet azonban azt sem állítani, hogy a belső harmóniával rendelkező, de alkalmazások nélküli matematikai fejezeteknek nincs joguk a létezésre. Ezek a tudomány belső szövetét alkotják és kivágásuk a szervezet egészét károsíthatja. Ezenkívül tapasztalható, hogy a matematika néhány olyan fejezete, amelyet sok évszázadon át nem alkalmaztak, később megtalálja alkalmazását. Klasszikus példát szolgáltatnak erre a másodrendű görbék; ezeket az ókorban, a tudomány belső szükséglete hívta életre, és csak később találtak igen fontos alkalmazásukra. Más részről a matematika egyes, csupán belső problémákkal foglalkozó fejezetei fokozatosan elkorcsosulnak és majdnem bizonyosan semmire sem használhatóaknak mutatkoznak. Ebben a helyzetben a kutatási téma kiválasztása a matematikusok számára szerfölött nyugtalanítóvá válik. Úgy vélem, hogy ha nem is az összes, de mindenesetre sok matematikusnak munkájában az ősforrások, vagyis az alkalmazások felé kell fordulnia. Ez egyrészt azért szükséges, hogy saját létezésüket igazolják, másrészt azért, hogy új, friss lendületet vigyenek a tudományos kutatásba. Ezekből a megfontolásokból kiindulva és a *Sztyeklov Intézet* vezetésének bizonyos nyomására én és három munkatársam, MISCSENKO E. F., GAMKRELIDZE R. V. és BOLTYANSZKIJ V. G., elhatároztuk, hogy kutatási területül alkalmazott témákat keresünk: a rezgéselméletben, pontosabban az elektronikus berendezések

Előadás a *Szovjetunió Tudományos Akadémiája Elnökségének* 1977. dec. 22-i ülésén, *Uszpehi Matematicseszkih Nauk*, **33** (1978) 22—28.

matematikai vizsgálatában és a szabályozáselméletben, amelyet most általánosabban ésszerűbb irányításelméletnek nevezni. Előre kizártuk vizsgálatainkból a mérnökök által már megfogalmazott matematikai feladatokat és kutatásainkat a műszaki problémákkal való megismerkedésre alapoztuk, aminek során kapcsolatot létesítettünk sok műszaki szakemberrel. Ennek során mi nem egyszerűen a matematika alkalmazására törekedtünk, hanem matematikailag is érdekes új feladatok megfogalmazására. A sok, általunk megismert műszaki probléma között szerepelt a következő. Egy repülési szakember ezt mondta: „Ha egy repülőgép egy másikat üldöz, akkor a követő gép pilótája természetesen tudja, hogyan kell ezt csinálnia, azonban érdekes lenne olyan elméletet alkotni, amely esetleg lehetővé tenné az üldözés végrehajtását automata segítségével." Hallomásból mindnyájan tudjuk azt most, hogy léteznek önmagukat rávezérlő rakéták, és ezeknek van valamilyen üldözéselméletük. Azonban az önvezérlésű rakéta valószínűleg olyan fölényben van sebesség és manőverező képesség dolgában a repülőgéppel szemben, hogy a viselkedésének alapjául szolgáló elmélet nagyon durva lehet. Rögtön fel szeretném hívni a figyelmet ennek a feladatnak, amely kezdetben teljesen megközelíthetetlennek tűnt számunkra, a különösségére. Valóban, az üldöző repülőgépnek nyilvánvalóan nem kell arra a helyre repülnie, ahol az adott időpontban a menekülő gép van, mivel az utóbbi természetesen elmegy arról a helyről, ahol jelenleg tartózkodik. Ugyanakkor értelmetlen dolog lenne feltételezni, hogy a menekülő gép egyenes vonalban halad; nyilván irányt változtathat és nem tudni, hogy merre. Repülőgép repülőgéppel való üldözésének feladata tudomásom szerint máig sincs megoldva. Az üldözés leegyszerűsített modelljeit vizsgáljuk; ezek a vizsgálatok alkotják az ún. differenciális játékok elméletének tárgyát. A „játék" szó arra a körülményre utal, hogy mindkét repülőgép jövőbeni viselkedése ismeretlen, a pilóta akaratától függ. „Differenciálisnak" azért nevezzük ezt a játékot, mert a repülőgép mozgását differenciálegyenletekkel írjuk le.

Ahhoz, hogy a matematikát egy műszaki feladat megoldására alkalmazhassuk, mindenekelőtt le kell írnunk a feladatot matematikailag. Ebben az esetben a repülőgép mozgásának matematikai leírásával kezdjük a dolgot. Ennek során, amint ezt a matematikusok mindig tenni szokták, eltekintünk a felesleges konkrétumoktól és arra törekszünk, hogy a megoldásra váró műszaki feladatnak csak a jellemző fővonásait ragadjuk meg. A repülőgépet térben mozgó pontnak tekintjük. Ismeretes, hogy pont helyzetét a térben három koordináta meghatározza. Ezeket x_1, x_2, x_3-mal jelöljük. Mivel a pont (a repülőgép) mozog, ezért sebesség-vektora is van. E vektor koordinátáit x_4, x_5, x_6-tal jelöljük. Az $x_1, x_2, \ldots, x_6$ értékek a mozgó pont állapotát adott időpillanatban meghatározzák; ezeket a pont fáziskoordinátáinak nevezzük. Abból a célból, hogy a felesleges konkrétumoktól elvonatkoztassunk, olyan objektumot fogunk vizsgálni, amelynek állapotát adott időpillanatban nem hat, hanem tetszőleges számú fáziskoordináta határozza meg. Ezeket $x_1, x_2, \ldots, x_n$-nel jelöljük. E mennyiségek együttesét egy betűvel szokás jelölni, így $x=(x_1, x_2, \ldots, x_n)$. Itt x objektumunk fázisterének pontja, vagy objektumunk fázisvektora. Az objektum tetszőleges fáziskoordinátáját x_i-vel jelöljük, ahol i az $i=1, 2, \ldots, n$ értékeket veheti fel. Mivel az objektum állapota az időben változik, ezért az x_i mennyiség is változik az időben, változási sebességét rendszerint $\dot{x}_i$-tal jelöljük. Ez az x_i mennyiségnek a t idő szerinti deriváltja. Az objektum viselkedésének fizikai törvényszerűsége rendszerint abban áll, hogy az objektum $x_1, x_2, \ldots, x_n$ fáziskoordinátái egyértelműen meghatározzák az x_i fáziskoordináta $\dot{x}_i$ változási sebességét, amit

matematikailag az

$$\dot{x}_i = f_i(x_1, x_2, \ldots, x_n) = f_i(x), \quad (i = 1, 2, \ldots, n) \tag{1}$$

formulával írunk le. Ez azt jelenti, hogy $\dot{x}_i$ az $x_1, x_2, \ldots, x_n$ mennyiségek függvénye, vagyis kiszámítható, ha $x_1, x_2, \ldots, x_n$ ismertek. Itt n számú, időben változó, ismeretlen mennyiségünk van, $x_1, x_2, \ldots, x_n$, vagyis ezek az idő függvényei: $x_i = x_i(t)$, és van n számú differenciálegyenletünk úgy, hogy a feladat matematikailag megoldható, azaz megkapható az állapot időbeli változásának törvényszerűsége, x mint az idő függvénye: $x = x(t)$. A legkülönfélébb objektumok leírhatók (1) alakú egyenletek segítségével, nemcsak mechanikai objektumok, hanem másfélék, mint például kémiai folyamatok is. Ez utóbbi esetben objektumunk $x_1, x_2, \ldots, x_n$ fáziskoordinátái a reakcióban részt vevő különböző anyagok tömegei. Ilyen egyenletekkel biológiai folyamatok is leírhatók, mint például farkasok, nyulak és a fű együttélése egy szigeten. Gazdasági törvényszerűségeket is le lehet írni (1) típusú egyenletrendszerek segítségével.

A repülőgép mozgásának itt vázolt leírása nem tartalmazza a számunkra legfontosabb momentumot. A repülőgépben pilóta ül, aki akaratának megfelelően változtathatja a gép mozgásának törvényszerűségeit, működésbe hozva az irányítás kormányműveit. Így a pilóta megváltoztathatja a motor tolóerejét, a farokkormány, a fékszárnyak helyzetét. Az irányítás minden egyes elemének helyzetét valamilyen szám határozza meg. Ezeket a számokat $u_1, u_2, \ldots, u_r$-rel és ezek együttesét egy betűvel $u = (u_1, u_2, \ldots, u_r)$-rel jelöljük. Itt u az a vektor, amelynek koordinátái a kormányok helyzetét határozzák meg. Ily módon a repülőgép mozgását nem az (1) egyenlet, hanem az

$$\dot{x}_i = f_i(x, u) \quad (i = 1, 2, \ldots, n) \tag{2}$$

egyenlet írja le, ahol a jobb oldalon szerepel az irányítás u vektora. Az irányítás u vektora a gép pilótája akaratának megfelelően változik az időben és ezért az idő adott $u = u(t)$ függvénye. Így aztán, valójában a (2) egyenlet alakja a következő:

$$\dot{x}_i = f_i(x, u(t)) \quad (i = 1, 2, \ldots, n), \tag{3}$$

ahol $u(t)$ az objektumnak az időben konkrétan megvalósított irányítása. A (3) egyenletrendszert már meg lehet oldani. Egy igen fontos körülményt meg kell jegyezni. A kormányok helyzetét meghatározó $u_1, u_2, \ldots, u_r$ értékek nem lehetnek tetszőlegesek. Így ha u_1 a motor tolóereje, akkor világos, hogy csak bizonyos határok, 0 és valamilyen a szám között változhat, $0 \leqq u_1 < a$. Ugyanúgy a farokkormány csak meghatározott határok között fordulhat el úgy, hogy ha u_2 ennek az elfordulásnak a szöge, akkor valamilyen $-b \leqq u_2 \leqq b$ egyenlőtlenséget elégít ki. Abból a célból, hogy a felesleges konkrétumoktól elvonatkoztassunk, egyszerűen azt mondjuk, hogy u az r-dimenziós térnek nem tetszőleges, hanem e tér valamely adott Ω halmazához tartozó vektora. A (2) differenciálegyenlet-rendszer az adott Ω halmazzal együtt írja le matematikailag az irányítható objektum viselkedésének lehetőségeit. Az ilyen objektumot irányíthatónak fogjuk nevezni, mivel viselkedése attól függ, hogy az időnek milyen $u(t)$ függvénye az u irányítás.

Avégett, hogy elkezdhessük repülőgép repülőgéppel történő üldözési feladatának megoldását, le kellene írnunk a második repülőgépet is irányítható objektumként és azután pontosan megfogalmaznunk az üldözés feladatát. Azonban, amint ezt már

korábban említettem, a feladat játékként való megfogalmazása annyira sajátos, hogy jobbnak láttuk először egy olyan másik feladat megoldását megkísérelni, amelyben a játéknak nincs szerepe. Feltételeztük, hogy a második objektum mozdulatlan, vagy a repülőgép nyelvén beszélve, azzal kezdtünk foglalkozni, hogy a repülőgépet az egyik állapotból a másikba vezéreljük a legrövidebb idő alatt. Matematikailag ezt a feladatot a következőképpen fogalmazzuk meg. A kezdő pillanatban megadjuk az objektum valamilyen kiindulási fázisállapotát, amit x^0-lal jelölünk. Ezenkívül adott az objektum valamely másik x^1 fázisállapota. Ha az objektumot valamilyen módon irányítva át tudjuk vezetni az x^0 fázisállapotból az x^1 állapotba, akkor felmerül a kérdés, melyik irányítás viszi át az x^0 állapotból x^1-be a legrövidebb idő alatt. Ez a gyorsaságra való optimalizálás feladata. Az ennek a feladatnak megoldásaként nyert $u(t)$ irányítást, illetve az objektum mozgását optimálisnak, illetve optimális mozgásnak nevezzük a gyorsaság értelmében. Ha az objektum mozgása során nemcsak az idő változik, hanem valamilyen más, számunkra különösen fontos mennyiség is, például fogy az üzemanyag, akkor fel lehet vetni az üzemanyag-fogyasztás optimalizálásának kérdését az x^0 állapotból az x^1 állapotba való átmenet során. Ez a feladat különösen fontos például űrhajóknak egyik pályáról a másikra történő átállítását vizsgálva, amikor a minimális üzemanyag-felhasználásnak óriási szerepe van.

Az így megfogalmazott optimalizálási feladatot a variációszámítás segítségével lehetne megoldani, ha az u irányító vektorra nem lennének korlátozó feltevések, vagyis ha u tetszőleges vektor lehetne. Az a körülmény, hogy az u vektor az adott Ω halmaznak eleme, egyszerre kiemeli az előbb megfogalmazott optimalizálási feladatot a klasszikus variációszámítás segítségével megoldható feladatok köréből. Ha az u vektor tetszőleges, akkor a megfogalmazott probléma a klasszikus variációszámítás körébe tartozik. Meg kell azonban jegyeznünk, hogy az itt adott megfogalmazásban a variációszámítás sohasem oldotta meg azt. A klasszikus variációszámításban megfogalmazott feladatok sokkal általánosabb jellegűek, mint az itt leírt probléma, és hiányzik belőlük az a konkrétság, amely a mi esetünkben onnan származik, hogy műszaki objektumot vizsgálunk. Kitűnt, hogy a variációs feladatnak ez a konkrétabb jellege, amely azzal kapcsolatos, hogy irányítható objektumot vizsgálunk, a probléma megoldásának új lehetőségeihez vezetett, olyan ötletek keletkezésének lehetőségét teremtette meg, amelyekhez az általános variációs feladattal kapcsolatban nagyon nehéz lett volna eljutni.

Ismertetem most a gyorsaságra való optimalizálás feladatának általunk adott megoldását. Bevezetünk n számú segédmennyiséget, $\psi_1, \psi_2, \ldots, \psi_n$, amelyek együttesét egyetlen betűvel $\psi=(\psi_1, \psi_2, \ldots, \psi_n)$-nel jelöljük. Bevezetjük továbbá a

$$H(\psi, x, u) = \psi_1 f_1(x, u)+\psi_2 f_2(x, u)+\ldots+\psi_n f_n(x, u) \tag{4}$$

segédfüggvényt. Látható, hogy H három vektortól, ψ-től, x-től és u-tól függ. Az újonnan bevezetett (4) segédfüggvényt azért jelöltük H-val, mert a reá vonatkozó, számunkra szükséges egyenletek nagyon hasonlítanak a mechanikából mindenki által ismert *Hamilton-egyenletekre*. Ezek az egyenletek a következők:

$$\dot{x}_i = \frac{\partial H(\psi, x, u)}{\partial \psi_i}, \quad \dot{\psi}_i = -\frac{\partial H(\psi, x, u)}{\partial x_i}. \tag{5}$$

Az (5) differenciálegyenlet-rendszer $2n$ számú egyenletből áll. Ismeretlen függvények bennük: $x_1, x_2, \ldots, x_n$, $\psi_1, \psi_2, \ldots, \psi_n$, $u_1, u_2, \ldots, u_r$, vagyis az ismeretlen függvények száma $2n+r$. Ily módon a rendszer nem teljes. Megoldani nem lehet. Ezt az egyenletrendszert a következő feltétel egészíti ki. Az irányítás u vektorát úgy kell megválasztani, hogy ψ és x minden rögzített értéke mellett a $H(\psi, x, u)$ függvény a maximumát ennél az u értéknél vegye fel. Az ezzel a feltétellel kiegészített (5) egyenletrendszer már teljes, és éppen ezt a feltételrendszert kell kielégítenünk ahhoz, hogy megkapjuk a gyorsaságra való optimalizálás feladatának megoldását. Ezt az eredményt nevezzük maximumelvnek. Valamely más, az időtől különböző mennyiség, például az üzemanyag-felhasználás optimalizálásának feladata hasonló módon oldható meg. Itt ennek a megoldását nem ismertetem. Az objektum mozgása céljának az ő meghatározott x^1 fázisállapotát tekintjük, vagyis azt, hogy a pont meghatározott helyre kerüljön meghatározott sebességgel. A maximumelv azonban alkalmas más feladatok megoldására is, például a cél lehet meghatározott helyre juttatás tetszőleges sebességgel.

Ha az irányítás vektora tetszőleges értékeket vehet fel és nincs megkötve azzal, hogy az Ω halmaz eleme legyen, akkor abból a feltételből, hogy a $H(\psi, x, u)$ függvénynek az u változóban maximuma van, következik, hogy az $u_1, u_2, \ldots, u_r$ változók szerinti összes parciális deriváltja zérus, vagyis teljesül az r számú feltétel

$$\frac{\partial H(\psi, x, u)}{\partial u_j} = 0, \quad (j = 1, 2, \ldots, r). \tag{6}$$

Ez az eredmény a klasszikus variációszámítás általános eredményeiből következik, de ebben a formában sohasem fogalmazták meg, minthogy a klasszikus variációszámításban egyáltalán nem foglalkoztak irányítható objektumokkal. Meg kell még jegyeznünk azt, hogy tetszőlegesen választható u esetén is a (6) feltétel gyengébb, mint az, hogy H-nak u-ban maximuma legyen.

Most a gyorsaságra való optimalizálás egy nagyon egyszerű feladatának olyan megoldását mutatjuk meg, amelyet a maximumelv segítségével meg lehet kapni, de a klasszikus variációszámítás módszereivel nem.

Tekintsük a matematikai ingát, vagyis egy olyan pont mozgását az egyenesen, amelyet az egyenesnek egy rögzített O pontja vonz a tőle való távolsággal arányos erővel. Az egyenes, amelyen a pont mozog, legyen az abszcissza tengely és az O pont az origó. A mozgó pont koordinátáját x-szel jelöljük. Ekkor e pont mozgásegyenlete az

$$\ddot{x} + x = 0 \tag{7}$$

alakban írható, ahol $\ddot{x}$ az x koordináta idő szerinti második deriváltja, vagyis a mozgó pont gyorsulása. A (7) egyenletet át lehet írni két elsőrendű egyenletből álló rendszerré:

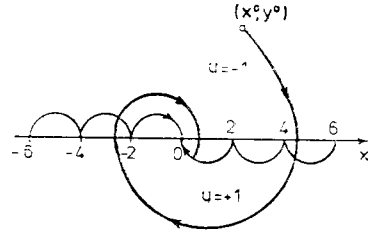
$$\dot{x} = y, \quad \dot{y} = -x. \tag{8}$$

Legyen $x=x(t)$, $y=y(t)$ a (8) rendszer tetszőleges megoldása. Geometriai ábrázolása végett tekintsük az (x, y) változók fázissíkján az $(x(t), y(t))$ pontot, amely a t idő változásával mozog. A pontnak a fázissíkon való mozgása eredményeképpen ily módon kapott pályát fázistrajektóriának nevezzük. A (8) rendszer esetében ez origó középpontú kör, amelyen a pont egy radián másodpercenkénti állandó szög-

sebességgel mozog az óramutató járásával megegyező irányban. Tételezzük fel most, hogy mozgó pontunkra, x-re u nagyságú külső erő hat, amely abszolút értékben nem lehet nagyobb egynél. Ekkor a pont mozgásegyenlete $\ddot{x}+x=u$ alakú, vagy egyenletrendszerré átírva:

$$\dot{x} = y, \quad \dot{y} = -x+u. \tag{9}$$

A (9) egyenletrendszer írja le az irányítható objektum mozgását, ahol u az irányító paraméter. Célunk az, hogy az u irányító paraméter felhasználásával, a kezdő időpontban a tetszőlegesen adott (x^0, y^0) helyzetben levő pontot minimális idő alatt nyugalmi helyzetbe, vagyis a fázissík koordináta-rendszerének origójába vezéreljük. A maximumelvből azonnal következik, hogy optimális irányítás esetén u csak a ± 1 értékeket veheti fel. Ha $u=1$, akkor a (9) rendszer fázistrajektóriája olyan kör, amelynek középpontja az $(1, 0)$ pont, ha pedig $u=-1$, akkor a fázistrajektória olyan kör, amelynek középpontja a $(-1, 0)$ pont. Tudva azt, hogy u optimális értéke plusz, vagy mínusz egy, csak azt kell meghatároznunk, hogyan veszi fel a mozgás során u váltakozva e két értéket. A maximumelvből könnyen levezethető, hogy u értéke csak a fázispontnak a fázissíkon felvett helyzetétől függ, pontosabban az egész fázissík két részre osztható, az egyiken u-nak 1-gyel, a másikon



1. ábra

-1-gyel kell egyenlőnek lennie. A fázissík felosztása két részre azzal a görbével történik, amelyek az ábrán látható. Ez a görbe egységsugarú félkörökből áll, amelyek az abszcisszatengelyen levő átmérőikre támaszkodnak, és az abszcisszatengely pozitív felén lefelé, negatív felén pedig felfelé állnak. Az origóval szomszédos két félkör maga is optimális trajektória úgy, hogy ha a kezdeti helyzet egyikükön van, akkor az origó felé irányuló mozgás ezen a félkörön történik. Továbbá bebizonyítható, hogy ha a kezdeti helyzet a síkot felosztó görbe alatt van, akkor u-nak az 1 értéket, ha pedig a görbe felett van, akkor u-nak a -1 értéket kell adni. Könnyű megrajzolni tetszőleges (x^0, y^0) kezdeti helyzetből kiindulva a pont optimális mozgásának pályáját (lásd az ábrát). A sík valamely (x^0, y^0) pontjából kiindulva a mozgást a (9) egyenlet határozza meg, ahol $u=\pm 1$, és u-nak éppen érvényben levő értéke akkor vált át az ellenkezőjére, amikor a megfelelő pálya eléri a síkot felosztó átkapcsolási görbét. Végül is a pont a síkot felosztó görbe egyik, az origóval szomszédos félkörére jut, majd ezután e félkörön mozog az origóig.

A maximumelv az optimalizációs feladatok megoldásának mindent átfogó, univerzális módszere. A tudomány legkülönbözőbb ágaiban sokszor került alkalmazásra és jelentős hatást gyakorolt a variációszámítás fejlődésére. Játék-feladatokban ilyen általános jellegű eredményeket nem sikerült elérnünk. Ilyen feladatokkal

most sok matematikus foglalkozik, közöttük meg kell említenünk a *Sztyeklov Intézet* munkatársainak egy csoportját és KRASZOVSZKIJ N. N. iskoláját *Szverdlovszk*ban. Ők jelentős eredményeket értek el. Itt arra szorítkozom, hogy egy konkrét példát idézzek fel az üldözési feladatra.

A tetszőleges n-dimenziós R térben, ahol $n \geqq 2$, tekintsünk két pontot, x-et és y-t, amelyeket egyidejűleg vektorokként is kezelünk. Az x pontot üldöző pontnak, az y pontot pedig menekülőnek nevezzük. Az üldözés folyamatát befejezettnek tekintjük, ha x egybeesik y-nal. E pontok mozgását írják le a következő egyenletek:

$$\ddot{x}+\alpha\dot{x} = u, \quad \ddot{y}+\beta\dot{y} = v. \tag{10}$$

Itt u és v az R tér vektorai. Feladatunkban ezek az irányítás vektorai. Irányuk tetszőlegesen választható, de hosszuk korlátos, teljesülniük kell az $|u|<\varrho$, $|v|<\sigma$ feltételeknek. Az $\alpha, \beta, \varrho, \sigma$ számok pozitívak. Ily módon a (10) egyenlet leírja a lineáris α csillapítású pont mozgását az u külső erő hatása alatt, amely utóbbinak iránya tetszőlegesen választható, nagysága azonban nem lehet nagyobb a ϱ számnál. Hasonló megállapítás érvényes az y pontra is. Az üldözés folyamatát két szemszögből lehet vizsgálni. Az egyik nézőpont az, amikor az üldözővel azonosítjuk magunkat. Ekkor feladatunk az üldözés végrehajtása az u irányítás alkalmas megválasztásával. Eközben az üldözés folyamata során végig megfigyeljük a menekülő objektum viselkedését. A második nézőpont az, amikor a menekülő objektummal azonosítjuk magunkat és feladatunk abban áll, hogy az üldözés elől elmeneküljünk a v irányítás alkalmas megválasztásával. Eközben végig megfigyeljük a minket üldöző objektumot. Az itt rendelkezésünkre álló alapvető eredmény a következő.

1. Az üldözés feladatát meg lehet oldani, vagyis az üldözést sikeresen be lehet fejezni mindig, ha fennáll a következő két egyenlőtlenség:

$$\frac{\varrho}{\alpha} > \frac{\sigma}{\beta}, \quad \varrho > \sigma. \tag{11}$$

2. A menekülés feladata mindig megoldható, ha teljesül a $\sigma > \varrho$ egyenlőtlenség. Bebizonyítható, hogy ha teljesülnek a (11) feltételek, akkor az üldözés feladatának megoldása során, mindig az üldöző rendelkezésére áll egy legjobb viselkedési mód, vagyis egyetlen, optimális $u(t)$ irányítás, amelytől való eltérés elkerülhetetlenül megnöveli az üldözés időtartamát. Az üldöző optimális $u(t)$ irányítása a menekülő objektum viselkedésétől függően, a t idő előrehaladtával lépésenként határozható meg.

FORDÍTOTTA:

FARKAS MIKLÓS
BME GÉPÉSZMÉRNÖKI KAR, MATEMATIKA TANSZÉK
1521 BUDAPEST, STOCZEK U., H ÉP. IV. EM.

A kiadásért felel az Akadémiai Kiadó igazgatója
Műszaki szerkesztő: Marton Andor
A kézirat a nyomdába érkezett: 1979. VII. 26. — Terjedelem: 18,55 (A/5 ív)
79-3807 — Szegedi Nyomda — F. v.: Dobó József igazgató

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 a felelős szerkesztő címére kell beküldeni:

Prékopa András, főszerkesztő, MTA SZTAKI
1502 Budapest XI., Kende u. 13—17.

A kéziratok szerkezeti felépítésének a következő követelményeket kell kielégíteni. A fejlécnek tartalmaznia kell a dolgozat címét, a szerző teljes nevét, valamint annak a városnak a nevét, ahol a szerző dolgozik. A fejléc után egy, képletet nem tartalmazó, legfeljebb 200 szóból álló kivonatot kell minden esetben megadni. A dolgozatot címmel ellátott szakaszokra kell bontani, és az egyes szakaszokat arab sorszámmal kell ellátni. Az esetleges bevezetésnek mindig az első szakaszt kell alkotnia. Az irodalomjegyzék mindig az utolsó szakasz kell hogy legyen, és azt nem kell sorszámmal ellátni. Az irodalomjegyzék után, a kézirat befejezéseképpen fel kell tüntetni a szerző teljes nevét és a munkahelye (illetve lakása) pontos postai címét. A dolgozatban előforduló képleteket szakaszonként újrakezdődően, a képlet előtt két zárójel közé írt kettős számozással kell azonosítani. Természetesen nem szükséges minden képletet számozással ellátni. Az esetleges definíciókat és tételeket (segédtételeket és lemmákat) ugyancsak szakaszonként újrakezdődő, kettős számozással kell ellátni. Kérjük a szerzőket, hogy ezeket, valamint a tételek bizonyítását a szövegben kellő módon emeljék ki. Minden dolgozathoz csatolni kell egy angol, német, francia vagy orosz nyelvű, külön oldalra gépelt összefoglalót. Amennyiben lehetséges, kérjük a nyomtatás számára különösen nehézkes matematikai jelölések használatának az elkerülését.

A dolgozat ábráit és az esetleges lábjegyzeteket a dolgozat végén, különálló lapokon kérjük beküldeni. Mind az ábrákat, mind a lábjegyzeteket a dolgozat szakaszokra bontásától független, folytatólagos arab sorszámozással kell ellátni. Az ábrák elhelyezését a dolgozat megfelelő helyén, széljegyzetként feltüntetett, ábraazonosító sorszámokkal kell megadni. A lábjegyzetekre a dolgozaton belül az azonosító sorszám felső indexkénti használatával lehet hivatkozni.

Az irodalmi hivatkozások formája a következő. Minden hivatkozást fel kell sorolni a dolgozat végén található irodalomjegyzékben, a szerzők, illetve társszerzők esetén az első szerző neve szerinti alfabetikus sorrendben úgy, hogy külön, de folytatólagos sorszámozású listát alkossanak a latin és a cirill betűs nevű szerzők műveire vonatkozó hivatkozások, és mindkét részben a megfelelő alfabetikus sorrend legyen kialakítva. 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 Ungelichungen«, *Journal für die reine und angewandte Mathematik* **124** (1902) 1—27.

[2] Kéri, G., „DUALSIMP", rutin a CDC 3300-as gépekre (Magyar Tudományos Akadémia Számítástechnikai és Automatizálási Kutató Intézete, CDC 3300 felhasználói ismertetők 2. 1973. május) 19—20.

[3] Prékopa, A., „Sztohasztikus 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 100 darab különlenyomatot kapnak, ezek költsége — nyomott oldalanként 25 forint — a szerzői díjat terheli.

Ára: 42,— Ft

INDEX: 26498
ISSN 0133—3399

TARTALOMJEGYZÉK

Hatvani László: Nem-autonóm differenciálegyenlet-rendszerek megoldásainak stabilitása és parciális stabilitása 1
Tóth János és *Hárs Vera:* A rekeszrendszerek inverz feladatáról 49
Béda Gyula: Egy képlékenységtani vizsgálat matematikai módszere 63
Nemetz Tibor és *Szilléry András:* Nyelvstatisztikai táblázatok 69
Asztalos Domonkos: Véges forrású tömegkiszolgálási modellek alkalmazása számítógépes rendszerekre 89
Németh György: Tömegkiszolgálás szimulációja sorbakapcsolt kiszolgálóhelyek esetén 103
Manigáti Csaba: Számítógéphálózatok bizonyos üzenetirányítási eljárásainak egyfajta matematikai modellje 123
Erol Gelenbe: Az optimális ellenőrzési intervallumhosszról 141
Kas Péter és *Mayer János:* A nemlineáris folyamprobléma egy megoldási módjáról 157
Soós Zsolt: A bimátrix játék egyensúlyi pontjainak megkereséséről 165
Hoffer János: Megengedett megoldások vizsgálatával bővített Benders-dekompozíciós algoritmus 177
Varecza Árpád: Módszerek a rendezési algoritmusok elvi korlátainak meghatározására 191

A külföldi szakirodalomból

Pontrjagin, L. Sz.: Optimalizálás és differenciális játékok 203

INDEX

Hatvani, L., "Stability and partial stability of nonautonomous systems of differential equations" 1
Tóth, J. and *Hárs, V.*, "On the inverse problem of compartment systems" 49
Béda, Gy., "A mathematical method for investigations in plasticity theory" 63
Nemetz, T. and *Szilléry, A.*, "Hungarian language-statistics" 69
Asztalos, D., "Application of finite source queueing models for computer systems" 89
Németh, G., "Simulation of waiting lines in series" 103
Manigáti, Cs., "A mathematical model for design of computer networks" 123
Gelenbe, E., "On the optimum checkpoint interval" 141
Kas, P. and *Mayer, J.*, "An algorithm for the solution of the nonlinear network flow problem" 157
Soós, Zs., "On the finding of bimatrix game's equilibrium points" 165
Hoffer, J., "Benders' partitioning procedure completed by the examination of feasible solutions" 177
Varecza, Á., "Methods for determination of principle bounds of ordering algorithms" 191

From the foreign literature

Pontrjagin, L. Sz., "Optimization and differential games" 203

Alkalmazott matematikai lapok

1979/3-4

A MAGYAR TUDOMÁNYOS AKADÉMIA
MATEMATIKAI ÉS FIZIKAI TUDOMÁNYOK
OSZTÁLYÁNAK KÖZLEMÉNYEI

5.

KÖTET

AKADÉMIAI KIADÓ, BUDAPEST

A MAGYAR TUDOMÁNYOS AKADÉMIA

MATEMATIKAI ÉS FIZIKAI TUDOMÁNYOK OSZTÁLYÁNAK

ALKALMAZOTT MATEMATIKAI LAPJA

V. kötet 3—4. szám

Szerkesztőség: 1502 Budapest XI., Kende u. 13—17.

Kiadóhivatal: 1055 Budapest V., Alkotmány u. 21.

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

Kéziratok a következő címre küldendők:

Prékopa András, főszerkesztő
1502 Budapest XI., Kende u. 13—17.

Ugyanerre a címre küldendő minden szerkesztőségi levelezés.

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 84 forint. Belföldi megrendelések az Akadémiai Kiadó, 1055 Budapest V., Alkotmány u. 21. címen (pénzforgalmi jelzőszám 215—11 488), külföldi megrendelések a Kultúra Külkereskedelmi Vállalat, H-1389 Budapest, Pf. 149. címen (pénzforgalmi jelzőszám 218—10 990) lehetségesek.

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

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

Alkalmazott Matematikai Lapok **5** *(1979)* 211—231

A POISSON-EGYENLET NUMERIKUS MEGOLDÁSAIRÓL

GERGELY JÓZSEF
Budapest

A dolgozatban áttekintést szeretnénk nyújtani a *Poisson-egyenlet* numerikus megoldási módszereiről. Megfogalmazzuk a *Poisson-egyenletre* vonatkozó *Dirichlet-feladat* megoldásának véges differencia módszerét, vizsgáljuk a differenciaegyenletek megoldásának iterációs és véges módszereit, majd a módszerek stabilitásának problémájával foglalkozunk. Végül összehasonlítást teszünk a módszerek közt.

1. A feladat megfogalmazása

Tekintsük az n dimenziós tér véges H tartományát, $\mathbf{x}\in H\subset R^n$, $\mathbf{x}=(x_1, x_2, \ldots, x_n)$; a H határa legyen G. Tekintsük a

$$\Delta u = \sum_{i=1}^{n} \frac{\partial^2 u}{\partial x_i^2} = -f(\mathbf{x}), \quad \mathbf{x}\in H. \tag{1.1}$$

Posson-egyenlet megoldását az

$$u = g_1(\mathbf{x}), \quad \mathbf{x}\in G, \quad \textit{(Dirichlet feladat)}, \tag{1.2}$$

$$\frac{\partial u}{\partial n} = g_2(\mathbf{x}), \quad \mathbf{x}\in G, \quad \textit{(Neumann feladat)}, \tag{1.3}$$

$$\frac{\partial u}{\partial n} + s(u - g_3(\mathbf{x})) = 0, \quad \mathbf{x}\in G, \quad \textit{(vegyes feladat)}, \tag{1.4}$$

peremfeltételek mellett, ahol $g_1(\mathbf{x}), g_2(\mathbf{x}), g_3(\mathbf{x})$ és $s=s(\mathbf{x})$ adott függvények, $\frac{\partial u}{\partial n}$ normális irányú deriváltat jelöl. A dolgozatban az (1.1) *Poisson-egyenlet* (1.2) peremfeltétel melletti megoldásának (vagyis a *Dirichlet-feladat*) numerikus módszerével foglalkozunk az $n=2$ esetben. Feltételezzük, hogy a feladatnak van megoldása és csak a megoldás numerikus módszereivel foglalkozunk.

A numerikus módszereket két nagy csoportra oszthatjuk: véges differencia és véges elemek módszerére. Mi csak a véges differencia módszereket vizsgáljuk. A véges elemek módszeréhez csak annyi megjegyzést fűzünk, hogy azok tanulmányozása, felhasználása legalább olyan jelentőséggel bír, mint a véges differencia módszereké, (lásd pl. [15]).

A véges differencia módszerek tanulmányozásához alapvető A. A. SZAMARSZKIJ akadémikus és iskolájának munkássága (lásd [1], [2], [3], [4] és [5]). A direkt módszerek tanulmányozásához jó alapot szolgáltat a [6] és a [7] dolgozatok tanulmányozása. A direkt megoldások új lehetősége vetődik fel R. E. BANK és D. J. ROSE [10] dolgozatában, aminek részletes elemzését a [8] és [9] dolgozatok adják.

Az említett munkákra támaszkodva a 2. pontban megfogalmazzuk a *Dirichlet-feladatra* vonatkozó véges differencia módszert kétdimenziós esetben. A 3. pontban a kapott differenciaegyenlet iterációs módszerekkel való megoldásával foglalkozunk. A 4. pontban a véges, az 5. pontban speciális módszereket vizsgálunk, majd a módszerek stabilitását vizsgáljuk. Végül összehasonlításokat teszünk a módszerek között. A módszerek megfogalmazását, tanulmányozását megkönnyíti az az egyszerűsítés, hogy a *Dirichlet-feladatot* csak téglalap tartományon vizsgáljuk.

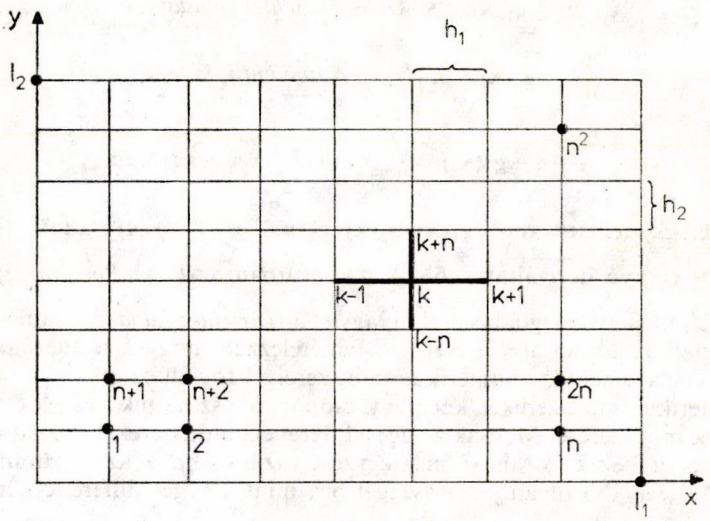
2. A véges differencia módszer

Tekintsük az (1.1) és (1.2) képletekkel megfogalmazott *Dirichlet-feladatot* kétdimenziós esetben a T téglalap tartományon.

$$\Delta u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = -f(x, y), \quad (x, y) \in T \subset R^2, \tag{2.1}$$

$$u_{(x,y)\in G} = g(x, y) \quad (G \text{ a } T \text{ határa}). \tag{2.2}$$

Vegyünk fel a T-ben egy derékszögű (a tengelyekkel párhuzamos) egyenlő lépésközű rácspont hálózatot, úgy hogy a téglalap mindkét oldalát felosztjuk $n+1$ egyenlő részre és az osztópontokon át a tengelyekkel párhuzamos egyenesek metszéspontjai alkossák a rácspontokat. Számozzuk be a belső rácspontokat az 1. ábrán látható módon. (A téglalap oldalai l_1 és l_2. $h_1 = l_1/(n+1)$, $h_2 = l/(n+1)$ a rácspont távolság x, ill. y irányban.)



1. ábra

Legyen u_k a függvény értéke a k-adik rácspontban. A (2.1) egyenletben szereplő differenciálhányadosok véges differencia közelítései egy belső k sorszámú rácspontban

$$\frac{\partial^2 u}{\partial x^2} \sim \frac{1}{h_1^2}(u_{k+1}-2u_k+u_{k-1}),$$

$$\frac{\partial^2 u}{\partial y^2} \sim \frac{1}{h_2^2}(u_{k+n}-2u_k+u_{k-n}).$$

A (2.1) egyenletünk véges differencia megközelítése a téglalap k-adik belső pontjában:

$$\Delta u \sim \frac{1}{h_1^2}(u_{k+1}-2u_k+u_{k-1})+\frac{1}{h_2^2}(u_{k+n}-2u_k+u_{k-n}) = -f(x_k, y_k) = -f_k. \tag{2.3}$$

Könnyen belátható, hogy a (2.3) egyenletben a közelítés pontossága $O(h_1^2)+O(h_2^2)$, (lásd [1]). Vagyis az

$$(L\mathbf{u})_k = \frac{1}{h_1^2}(u_{k+1}-2u_k+u_{k-1})+\frac{1}{h_2^2}(u_{k+n}-2u_k+u_{k-n}) \tag{2.4}$$

jelölést használva a téglalap belső pontjában

$$\Delta \mathbf{u} = L\mathbf{u}+O(h_1^2)+O(h_2^2).$$

Az L differenciaoperátor tulajdonságainak megfogalmazásához használjuk a következő belsőszorzat és norma jelöléseket:

$$(\mathbf{u}, \mathbf{v}) = \sum_{i=1}^{n} u_i v_i h, \quad \|\mathbf{u}\| = \sqrt{(\mathbf{u}, \mathbf{u})}. \tag{2.5}$$

Az L differenciaoperátor tulajdonságai a következőképpen fogalmazhatók meg:
1) Az L önadjungált operátor, vagyis

$$(L\mathbf{u}, \mathbf{v}) = (\mathbf{u}, L\mathbf{v});$$

2) A $-L$ operátor pozitív definit, vagyis

$$(-L\mathbf{u}, \mathbf{u}) \geqq \delta \|\mathbf{u}\|^2,$$

ahol δ a $-L$ operátor legkisebb sajátértéke (pozitív):

$$\delta = \frac{4}{h_1^2}\sin^2\frac{\pi h_1}{2l_1}+\frac{4}{h_2^2}\sin^2\frac{\pi h_2}{2l_2} \geqq \frac{4}{l_1^2}+\frac{4}{l_2^2};$$

3) Felírható a

$$\delta\|\mathbf{u}\|^2 \leqq (-L\mathbf{u}, \mathbf{u}) \leqq \gamma\|\mathbf{u}\|^2 \tag{2.6}$$

egyenlőtlenség, ahol γ a $-L$ operátor legnagyobb sajátértéke:

$$\gamma = \frac{4}{h_1^2}\cos^2\frac{\pi h_1}{2l_1}+\frac{4}{h_2^2}\cos^2\frac{\pi h_2}{l_2} < \frac{4}{h_1^2}+\frac{4}{h_2^2}.$$

A (2.6) egyenlőtlenséget szimbolikusan, rövidebben

$$\delta E \leqq -L \leqq \gamma E$$

alakban szokták felírni;
4) Explicit alakban megadható az $L\mathbf{v}+\lambda\mathbf{v}=0$, (ha $\mathbf{x}$ belső pont és $\mathbf{v}=\mathbf{0}$ a határon) sajátértékfeladat megoldása.
A sajátértékek:

$$\lambda_{ij} = 4\left(\frac{1}{h_1^2}\sin^2\frac{i\pi h_1}{2l_1}+\frac{1}{h_2^2}\sin^2\frac{j\pi h_2}{l_2}\right) \quad (1 \leqq i \leqq n,\ 1 \leqq j \leqq n), \tag{2.7}$$

míg a sajátvektorok:

$$\mathbf{v}_{ij} = \sqrt{\frac{4}{l_1 l_2}}\sin\frac{ih_1\pi}{l_1}\sin\frac{jh_2\pi}{l_2} \quad (1 \leqq i \leqq n,\ 1 \leqq j \leqq n). \tag{2.8}$$

A sajátvektorok ortonormált rendszert alkotnak a (2.5) skalárszorzatra nézve

$$(\mathbf{v}_{i_1 j_1}, \mathbf{v}_{i_2 j_2}) = \delta_{i_1 j_1}\delta_{i_2 j_2}.$$

A további egyszerűsítés kedvéért legyen $l_1=l_2$ és $h=h_1=h_2$, akkor a (2.3) differenciaegyenlet a (2.4) jelöléssel a k-adik belső rácspontban

$$h^2(L\mathbf{u})_k+f_k = 0. \tag{2.9}$$

Írjuk fel az összes (2.9) alakú egyenletet az 1. ábra számozásának megfelelő sorrendben. (Az olyan rácsponthoz tartozó egyenletbe, amelynek valamelyik szomszédos rácspontja határpont, értelemszerűen a határon adott függvényérték helyettesítendő.)

Az ilyen módon felírt lineáris egyenletek a következő lineáris egyenletrendszert adják

$$\mathbf{A}\mathbf{v} = \mathbf{b}, \tag{2.10}$$

ahol

$$\mathbf{A} = \begin{bmatrix} \mathbf{B} & -\mathbf{I} & & & & \\ -\mathbf{I} & \mathbf{B} & -\mathbf{I} & & & \\ & \ddots & \ddots & \ddots & & \\ & & -\mathbf{I} & \mathbf{B} & -\mathbf{I} & \\ & & & -\mathbf{I} & \mathbf{B} & \end{bmatrix}, \tag{2.11}$$

$$\mathbf{B} = \begin{bmatrix} 4 & -1 & & & & \\ -1 & 4 & -1 & & & \\ & \ddots & \ddots & \ddots & & \\ & & -1 & 4 & -1 & \\ & & & -1 & 4 & \end{bmatrix}, \tag{2.12}$$

$\mathbf{b}=\{b_i\}$, $i=1, \ldots, n^2$ és

$$b_i=\begin{cases} h^2f_i+h^2(g_{i_1}+g_{i_2}), & \text{ha } i=1,\ n,\ n^2-n,\ n^2,\ g_{i_1} \text{ és } g_{i_2} \text{ pedig a sarokpont melletti, a határon levő rácspontbeli adott függvényértékek;} \\ h^2f_i+h^2g_i, & \text{ha } i \text{ a határvonal melletti rácsok sorszáma (a sarokpontokat kivéve), } g_i \text{ pedig a szomszédos határon levő rácson a függvényérték:} \\ h^2f_i, & \text{a többi belső pontban.} \end{cases}$$

Minthogy az **A** mátrixot az L differenciaoperátor mátrixából kaphatjuk úgy, hogy azt $-h^2$-tel szorozzuk, ezért az **A** mátrixnak hasonló tulajdonságai vannak, mint a $-L$ operátor mátrixának: önadjungált pozitív definit és az **A** mátrixra megfogalmazott sajátértékfeladat is explicite megoldható.

A továbbiakban a (2.10) egyenletrendszer numerikus megoldásaival foglalkozunk, először az iterációs, majd a direkt módszerekkel.

3. Az iterációs eljárások

Az m ismeretlenes

$$\mathbf{Ax}=\mathbf{b} \tag{3.1}$$

lineáris egyenletrendszer iterációs módszerrel történő megoldásának alapgondolata a következőképpen fogalmazható meg: Valamilyen $\mathbf{x}^{(0)}$ közelítésből kiindulva az

$$\mathbf{x}^{(k+1)}=\mathbf{Bx}^{(k)}+\mathbf{d}$$

iterációs képlettel számítjuk az $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots$ közelítéseket. Ha az eljárás konvergens, akkor az $\mathbf{x}^{(n)}$ vektorsorozat konvergál az $\mathbf{x}$ megoldáshoz. Az $\mathbf{x}^{(0)}$ kiindulás tetszőlegesen választható, de előbb kapjuk meg a megoldást, ha annak valamilyen jó közelítésből indulunk ki. Általában az $\mathbf{x}^{(0)}=0$, vagy az az $\mathbf{x}^{(0)}=\mathbf{b}$ kiindulást szokták választani.

Az iterációs eljárás konvergenciájának elégséges feltétele, hogy az iterációs képletben szereplő **B** mátrix sajátértékei abszolút értékben 1-nél kisebbek legyenek. A konvergencia gyorsaságát is a **B** mátrix spektrál sugara határozza meg. Az iterációs eljárások az iterációs képletben szereplő **B** mátrix és a **d** vektor megválasztásában különböznek.

A lineáris egyenletrendszer iterációs eljárással való megoldásának vizsgálatát a klasszikus iterációs eljárások ismertetésével kezdjük (lásd [11]). Bontsuk fel az $\mathbf{A}=\{c_{ij}\}$ mátrixot $\mathbf{A}=\mathbf{D}-\mathbf{E}-\mathbf{F}$ alakba, ahol **D** a diagonális elemeket, $-\mathbf{E}$ a diagonális alatti, $-\mathbf{F}$ a diagonális feletti mátrixelemeket tartalmazza. Tegyük fel, hogy a diagonális elemek 0-tól különbözőek. A *Jacobi-iteráció* a (3.1) egyenletrendszer megoldására a következő képletekkel fogalmazható meg:

$$\mathbf{x}^{(k+1)}=\mathbf{D}^{-1}(\mathbf{E}+\mathbf{F})\mathbf{x}^{(k)}+\mathbf{D}^{-1}\mathbf{b},$$

vagy más formában:

$$x_i^{(k+1)}=-\sum_{\substack{j=1 \\ j\neq i}}^{m}\frac{a_{ij}}{a_{ii}}x_j^{(k)}+\frac{b_i}{a_{ii}}, \quad 1\leqq i\leqq m.$$

A *Gauss—Seidel-iteráció* képletei:

$$\mathbf{x}^{(k+1)} = (\mathbf{D}-\mathbf{E})^{-1}\mathbf{F}\mathbf{x}^{(k)} + (\mathbf{D}-\mathbf{E})^{-1}\mathbf{b},$$

vagy más formában:

$$x_i^{(k+1)} = -\sum_{j=1}^{i-1} \frac{a_{ij}}{a_{ii}} x_j^{(k+1)} - \sum_{j=i+1}^{m} \frac{a_{ij}}{a_{ii}} x_j^{(k)} + \frac{b_i}{a_{ii}}, \quad 1 \leqq i \leqq m.$$

A *szukcesszív túlrelaxálási módszer* (mely az angol irodalomban *Successive Over Relaxation,* SOR néven ismert) iterációs képletei:

$$\mathbf{x}^{(k+1)} = (\mathbf{I}-\omega\mathbf{L})^{-1}\{(1-\omega)\mathbf{I}+\omega\mathbf{U}\}\mathbf{x}^{(k)} + \omega(\mathbf{I}-\omega\mathbf{L})^{-1}\mathbf{D}^{-1}\mathbf{b},$$

vagy más formában:

$$x_i^{(k+1)} = x_i^{(k)} + \omega\left\{-\sum_{j=1}^{i-1} \frac{a_{ij}}{a_{ii}} x_j^{(k+1)} - \sum_{j=i+1}^{m} \frac{a_{ij}}{a_{ii}} x^{(k)} + \frac{b_i}{a_{ii}} - x_i^{(k)}\right\},$$

ahol

$$\mathbf{L} = \mathbf{D}^{-1}\mathbf{E}, \quad \mathbf{U} = \mathbf{D}^{-1}\mathbf{F}.$$

Az ω paraméter értékének megválasztása erősen befolyásolja a konvergencia gyorsaságát. A legjobb ω érték megválasztása külön feladat, szokásos az $\omega = 2/(1+\sqrt{1-\lambda^2})$ választás, ahol λ a $(\mathbf{D}-\mathbf{E})^{-1}\mathbf{F}$ mátrix legnagyobb abszolút értékű sajátértékének abszolút értéke. (Amit általában nem ismerünk, így annak valamilyen jó becslése is alkalmazható.)

Az egyes módszerek iterációs képleteiből könnyen kiolvasható az általánosan megfogalmazott képletben szereplő $\mathbf{B}$ mátrix. Így az egyes eljárások konvergenciájának elégséges feltételei is könnyen megfogalmazhatók.

Minthogy a *Dirichlet-feladat* esetén az $\mathbf{A}$ mátrix pozitív definit, a fenti *Jacobi-, Gauss—Seidel-* és a *szukcesszív túlrelaxál ási* eljárások konvergensek. A konvergencia gyorsasága azonban lehet lassú.

Az iterációs módszerek egy általános elméletét A. A. SZAMARSZKIJ dolgozta ki (lásd [1], [2], [5]). Az alábbiakban vázolni szeretnénk az általános elmélet egy rövid gondolatmenetét.

A (3.1) egyenletrendszer megoldására a

$$\mathbf{B}_k \frac{\mathbf{x}^{(k+1)}-\mathbf{x}^{(k)}}{\tau_{k+1}} + \mathbf{A}\mathbf{x}^{(k)} = \mathbf{b}, \quad k = 0, 1, \ldots. \tag{3.2}$$

iteráció képletet használjuk. Szeretnénk a $\mathbf{B}_k$ mátrixokat és a τ_k paramétereket úgy megválasztani, hogy az eljárásunk optimális legyen abban az értelemben, hogy a legkevesebb művelettel kapjuk meg a kellően pontos (adott ε pontosságú) megoldást.

A $\mathbf{B}_k = \mathbf{I}$ (egységmátrix) választás mellett jutunk az explicit iterációs eljáráshoz:

$$\frac{\mathbf{x}^{(k+1)}-\mathbf{x}^{(k)}}{\tau_{k+1}} + \mathbf{A}\mathbf{x}^{(k)} = \mathbf{b}, \quad \mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \tau_{k+1}(\mathbf{A}\mathbf{x}^{(k)}-\mathbf{b}). \tag{3.3}$$

Ebben az esetben már csak a τ_k paraméterek optimális megválasztása a feladat.

A megoldás pontosságát a

$$\|\gamma_k\| = \|\mathbf{A}\mathbf{x}^{(k)} - \mathbf{b}\| \leqq q_n \|\mathbf{A}\mathbf{x}^{(0)} - \mathbf{b}\| \tag{3.4}$$

eltérés normában mérjük. Pontosabban azt szeretnénk elérni, hogy adott ε mellett minél kevesebb iterációs lépésben teljesüljön a $q_n \leqq \varepsilon$ egyenlőtlenség.

Az elmélet lényeges pontja, hogy adott ε-hoz előre meg tudjuk becsülni az $n_0(\varepsilon)$ küszöbszámot, úgy hogy $n \geqq n_0(\varepsilon)$ esetén $q_n \leqq \varepsilon$ legyen. Ehhez a következő mennyiségeket kell kiszámolnunk, illetve megbecsülnünk: Legyen az $\mathbf{A}$ legkisebb és legnagyobb sajátértéke

$$\gamma_1 = \min \lambda(\mathbf{A}), \quad \gamma_2 = \max \lambda(\mathbf{A}),$$

$$\xi = \frac{\gamma_1}{\gamma_2}, \quad \varrho_0 = \frac{1-\xi}{1+\xi}, \quad \varrho_1 = \frac{1-\sqrt{\xi}}{1+\sqrt{\xi}}, \quad \tau_0 = \frac{2}{\gamma_1+\gamma_2},$$

akkor a

$$q_n = \frac{2\varrho_1^n}{1+\varrho_1^{2n}} \leqq \varepsilon$$

pontosság eléréséhez

$$n \geqq \frac{\ln\left(\frac{2}{\varepsilon}\right)}{\ln\left(\frac{1}{\varrho_1}\right)}$$

iterációs lépésre van szükségünk. Ennek felhasználásával

$$n_0(\varepsilon) = \frac{\ln\left(\frac{2}{\varepsilon}\right)}{2\sqrt{\xi}} = \frac{1}{2}\sqrt{\frac{\gamma_2}{\gamma_1}} \ln\left(\frac{2}{\varepsilon}\right). \tag{3.5}$$

Rögzítsük a (3.5) képlet segítségével kiszámolt $n = n_0(\varepsilon)$ számot és tekintsük az n-edfokú

$$T_n(t) = \cos(n \arccos t)$$

Csebisev-polinomot. Ennek gyökei explicit módon kiszámíthatók a

$$t_k = \cos\frac{2k-1}{2n}\pi, \quad k = 1, \ldots, n \tag{3.6}$$

képlettel. Az elmélet szerint a (3.3) iterációs képletben használható optimális paraméterek ezek után

$$\tau_k = \frac{\tau_0}{1+\varrho_0 t_k}, \quad k = 1, \ldots, n. \tag{3.7}$$

Összefoglalva: a (3.1) egyenletrendszernek a (3.3) explicit iterációs eljárással való optimális megoldása a következő lépéseket igényli:

1) Adott ε-hoz számítsuk ki az $n_0(\varepsilon)$ számot a (3.5) képlet segítségével;
2) A (3.6) és (3.7) képletekből határozzuk meg a τ_k paramétereket;
3) Alkalmazzuk a (3.3) iterációs képletet $k = 0, 1, \ldots, n_0(\varepsilon)$-ra.

Az $\mathbf{x}^{n_0(\varepsilon)}$ a (3.1) egyenletrendszer olyan megoldása lesz, amelyre

$$\|\mathbf{A}\mathbf{x}^{(n_0(\varepsilon))}-\mathbf{b}\| \leqq \varepsilon\|\mathbf{A}\mathbf{x}^{(0)}-\mathbf{b}\|.$$

Az általános iterációs eljárásból

$$\tau = \tau_k = \frac{2}{\gamma_1+\gamma_2}$$

választás mellett megkapjuk a *Jacobi-* vagy *egyszerű iterációs eljárást:*

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)}-\tau(\mathbf{A}\mathbf{x}^{(k)}-\mathbf{b}). \tag{3.8}$$

A (3.8) képlettel számolva

$$q_1 = \frac{2\varrho_1}{1+\varrho_1^2} = \varrho_0 \leqq \varepsilon$$

pontosság eléréséhez $n_0(\varepsilon)=\dfrac{\ln\left(\frac{1}{\varepsilon}\right)}{2\xi}$ iterációs lépésre van szükség.

A fent ismertetett általános iterációs eljárás használatára még nem alkalmas, ha $n_0(\varepsilon)$ nagy szám. Ugyanis a fenti módon való számolás instabil. Ez azt jelenti, hogy a (3.3) képlet sokszori alkalmazása a kerekítési hibák halmozódásához vezet. Ezen a problémán úgy lehet segíteni, hogy a *Csebisev-polinom* gyökeit nem a (3.6) képlettel előírt sorrendben vagyis $k=1, \ldots, n$-re számoljuk, hanem azokat egy más meghatározott sorrendben használjuk fel a (3.7) képletben.

A sorrend meghatározására pontos eljárás van adva az [1] könyvben. Itt ennek csak azt a speciális esetét ismertetjük, amikor $n=2^p$ alakú.

Legyen $Q_n(k)$, $k=1, \ldots, n$ az első n darab páratlan szám egy sorbarendezett alakja. Az [1]-ben található eljárás $n=2^p$ esetre:

$$Q = \{\theta_1(1)\} = \{1\}$$

$$Q_{2m}(2i-1) = \theta_m(i), \quad Q_{2m}(2i) = 4m(2i-1), \quad i = 1, \ldots, m.$$

Ez az elrendezési szabály például $n=16=2^4$-re a következőket eredményezi:

$$Q_1 = \{1\}$$

$$Q_2 = \{1, 3\}$$

$$Q_4 = \{1, 7, 3, 5\}$$

$$Q_8 = \{1, 15, 7, 9, 3, 13, 15, 11\}$$

$$Q_{16} = \{1, 31, 15, 17, 7, 25, 9, 23, 3, 29, 13, 19, 5, 27, 11, 21\}.$$

Ha a (3.2) képletben $\tau_k=1$ paramétert és $\mathbf{B}_k=\mathbf{D}-\mathbf{E}$ mátrixot választjuk, akkor a *Gauss—Seidel implicit iterációs képlethez* jutunk:

$$(\mathbf{D}-\mathbf{E})(\mathbf{x}^{(k+1)}-\mathbf{x}^{(k)})+\mathbf{A}\mathbf{x}^{(k)} = \mathbf{b}.$$

Ebben az esetben $n_0(\varepsilon)=O\left(\frac{1}{h^2}\ln\frac{1}{\varepsilon}\right)$.

A *szukcesszív túlrelaxálási* iterációs eljárást megkapjuk a

$$\tau_k = \omega \quad \text{és} \quad \mathbf{B}_k = \mathbf{D} - \omega\mathbf{E},$$

vagy a

$$\tau_k = 1 \quad \text{és} \quad \mathbf{B}_k = \frac{1}{\omega}\mathbf{D} - \mathbf{E}$$

választással. Az ω paraméter jó megválasztásával elérhetjük, hogy $n_0(\varepsilon) = O\left(\frac{1}{h}\ln\frac{1}{\varepsilon}\right)$ lesz.

A (3.2) képlettel megfogalmazott implicit iterációs eljárás részletes vizsgálatával nem foglalkozunk. Alkalmazásához idézünk csupán egy tételt az [1] könyvből: Tegyük fel, hogy teljesülnek a

$$\mathbf{B} = \mathbf{B}^* > 0, \quad \mathbf{A} = \mathbf{A}^* > 0, \quad \gamma_1\mathbf{B} \leqq \mathbf{A} \leqq \gamma_2\mathbf{B}, \quad \gamma_1 > 0$$

egyenlőtlenségek, akkor létezik optimális τ_k paraméter sorozat a

$$\mathbf{B}\frac{\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}}{\tau_{k+1}} + \mathbf{A}\mathbf{x}^{(k)} = \mathbf{b}, \quad k = 1, \ldots, n-1 \tag{3.9}$$

feladat megoldására és

$$\|\mathbf{A}\mathbf{x}^n - \mathbf{b}\|_{\mathbf{B}^{-1}} \leqq q_n \|\mathbf{A}\mathbf{x}^{(0)} - \mathbf{b}\|_{\mathbf{B}^{-1}}$$

ahol

$$q_n = \frac{2\varrho_1^n}{1 + 2\varrho_1^{2n}}, \quad q_1 = \frac{1 - \sqrt{\xi}}{1 + \sqrt{\xi}}, \quad \xi = \frac{\gamma_1}{\gamma_2}.$$

A paraméterek a

$$\tau_k = \frac{\tau_0}{1 + \varrho_0 t_k}, \quad t_k = \cos\frac{Q_n(k)}{2n}\pi$$

képletekkel nyerhetők.

Az *implicit iterációs módszer* egy konkrét esete a *„váltakozó-háromszög módszer"* ([1]). Ez röviden a következőképpen fogalmazható meg: Bontsuk fel az önadjungált $\mathbf{A}$ mátrixot

$$\mathbf{A} = \mathbf{R} + \mathbf{Q}$$

összegre, ahol az $\mathbf{R} = \{r_{ij}\}$ és $\mathbf{Q} = \{q_{ij}\}$ mátrixok elemeire

$$r_{ij} = \begin{cases} a_{ij}, & \text{ha } j < i, \\ 0, & \text{ha } j > i, \end{cases} \qquad q_{ij} = \begin{cases} 0, & \text{ha } j < i, \\ a_{ij}, & \text{ha } j > i, \end{cases}$$

és $r_{ii} = q_{ii} = 0{,}5a_{ii}$. Legyen

$$\mathbf{B} = (\mathbf{I} + \omega\mathbf{R})(\mathbf{I} + \omega\mathbf{Q}),$$

ahol $\omega > 0$ paraméter. Belátható, hogy az így definiált $\mathbf{B}$ mátrix esetén teljesülnek az implicit iteráció eljárásra fennt megfogalmazott tétel feltételei, így a tétel alkalmazható ebben az esetben. A (3.9) iterációs képlet alakja pedig:

$$(\mathbf{I} + \omega\mathbf{R})(\mathbf{I} + \omega\mathbf{Q})\frac{\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}}{\tau_{k+1}} + \mathbf{A}\mathbf{x}^{(k)} = \mathbf{b}. \tag{3.10}$$

A (3.10) egyenletet célszerű felbontani

$$(\mathbf{I}+\omega\mathbf{R})\mathbf{y} = \mathbf{t}_k \tag{3.11}$$

$$(\mathbf{I}+\omega\mathbf{Q})\mathbf{x}^{(k+1)} = \mathbf{y} \tag{3.12}$$

egyenletekre, ahol $\mathbf{t}_k = \mathbf{B}\mathbf{x}^{(k)} - \tau_{k+1}(\mathbf{A}\mathbf{x}^{(k)} - \mathbf{b})$.

A (3.11) és a (3.12) egyenletek mátrixa háromszögmátrix, ezért tetszőleges $\mathbf{x}^{(0)}$-ból kiindulva minden $\mathbf{x}^{(k)}$ esetén kevés számolással nyerhető a (3.11) és a (3.12) egymásutáni megoldása révén az $\mathbf{x}^{(k+1)}$. Ezen előnyök miatt a változóháromszög módszer gyors konvergenciát biztosít a *Dirichlet-egyenlet* numerikus megoldására.

Az [5] dolgozatban A. A. SZAMARSZKIJ összefoglalja a véges differencia módszerek elméletének legújabb eredményeit, amit ő és iskolája dolgozott ki az utóbbi években. Csak két általános érvényű tételt idézünk a dolgozatból:

3.1. Tétel: Tegyük fel, hogy az $\mathbf{A}$ és $\mathbf{B}$ differenciaoperátor nem függ a $t_n = n\tau$ paramétertől, $\mathbf{A}$ önadjungált és pozitív definit, valamint létezik a $\mathbf{B}^{-1}$. Akkor a (3.9) differenciaegyenlet stabilitásának szükséges és elégséges feltétele, hogy teljesüljön a

$$\mathbf{B}_0 = R_e\mathbf{B} \geqq 0{,}5\tau\mathbf{A},$$

vagy a

$$\mathbf{B} \geqq 0{,}5\tau\mathbf{A}$$

egyenlőtlenség.

3.2. Tétel: Legyenek $\mathbf{A}$ és $\mathbf{B}$ önadjungált operátorok és legalább az egyik potitív definit. Ha a (3.9) egyenlet stabil, akkor teljesülni kell a

$$\mathbf{B} \geqq 0{,}5\tau\mathbf{A}$$

egyenlőtlenségnek.

A differenciaegyenletek iteratív megoldásának gyakran használt módszere a *„változó irányok módszere"*. Mi a módszer rövid ismertetését a [11] könyv alapján adjuk meg. Legyen a megoldandó egyenletrendszerünk a (3.1) egyenletrendszer.

Bontsuk fel az $\mathbf{A}$ mátrixot

$$\mathbf{A} = \mathbf{H}+\mathbf{V}+\mathbf{S}$$

alakba úgy, hogy $\mathbf{H}$, $\mathbf{V}$ és $\mathbf{S}$ olyan differenciaoperátorok, melyek a vizsgált tartomány egy (x_0, y_0) belső pontjában a következőképpen értelmezhetők (h rácstávolság):

$$\mathbf{H}u = -u(x_0-h, y_0)+2u(x_0, y_0)-u(x_0+h, y_0),$$

$$\mathbf{V}u = -u(x, y_0-h)+2u(x_0, y_0)-u(x_0, y_0+h),$$

$$\mathbf{S}u = h^2u(x_0, y_0).$$

Alkalmas átrendezéssel a (3.1) egyenletrendszer felírható

$$\left(\mathbf{H}+\frac{1}{2}\mathbf{S}+r\mathbf{I}\right)\mathbf{x} = \left(r\mathbf{I}-\mathbf{V}-\frac{1}{2}\mathbf{S}\right)\mathbf{x}+\mathbf{b}$$

$$\left(\mathbf{V}+\frac{1}{2}\mathbf{S}+r\mathbf{I}\right)\mathbf{x} = \left(r\mathbf{I}-\mathbf{H}-\frac{1}{2}\mathbf{S}\right)\mathbf{x}+\mathbf{b}$$

alakban, valamilyen $r>0$ paraméter bevezetésével. Ezen felbontás segítségével felírhatjuk a következő iterációs képleteket:

$$\left(\mathbf{H}+\frac{1}{2}\mathbf{S}+r_{m+1}\mathbf{I}\right)\mathbf{x}^{\left(m+\frac{1}{2}\right)}=\left(r_{m+1}\mathbf{I}-\mathbf{V}-\frac{1}{2}\mathbf{S}\right)\mathbf{x}^{(m)}+\mathbf{b}, \tag{3.13}$$

$$\left(\mathbf{V}+\frac{1}{2}\mathbf{S}+r_{m+1}\right)\mathbf{x}^{(m+1)}=\left(r_{m+1}\mathbf{I}-\mathbf{H}-\frac{1}{2}\mathbf{S}\right)\mathbf{x}^{\left(m+\frac{1}{2}\right)}+\mathbf{b}.$$

A (3.13) képletekkel meghatározott iterációs eljárás a *változó irányok módszere.* Az elnevezés a (3.13) képlet segítségével szemléletesen úgy fogalmazható meg, hogy a (3.13) első képlete vízszintes a másik függőleges irányt jelöl ki a számolás menetére a kétdimenziós tartományban. Tetszőleges $\mathbf{x}^{(0)}$ kiindulásból az r_m paraméter megválasztása után váltakozva kell használni a (3.13) képleteit.

A változó irányok módszerét széles körben használják és sok általánosítása ismeretes. Az iteráció gyors konvergenciát biztosít a differenciaegyenletek megoldására. Kimutatható, hogy a konvergencia sebessége: $n^2 \ln n \ln \frac{1}{\varepsilon}$, ahol ε a pontosság, n az osztáspontok száma a téglalap tartomány egyik irányában.

A [14] dolgozatban N. Sz. Bahvalov olyan iterációs eljárást vizsgál, amely tetszőleges, változó együtthatós elliptikus egyenlet megoldására alkalmas és gyors konvergenciát biztosít. A konvergencia rendje $n^2 \ln \frac{1}{\varepsilon}$.

4. Véges módszerek

A 2. pontban megfogalmazott (2.10) egyenletben $(\mathbf{Az}=\mathbf{b})$ szereplő (2.11) alakú $\mathbf{A}$ mátrix jelölésére használjuk az

$$\mathbf{A}=[-\mathbf{I}, \mathbf{B}, -\mathbf{I}] \tag{4.1}$$

szimbólumot, ahol

$$\mathbf{B}=[-1, 4, -1], \tag{4.2}$$

az $\mathbf{I}$ n méretű egységmátrix, $\mathbf{B}$ mérete is n.

A $\mathbf{z}$ és a $\mathbf{b}$ is n^2 méretű vektorok, amelyeket n darab n méretű vektorra bontva használunk:

$$\mathbf{z}=\{\mathbf{z}^i\}, \quad \mathbf{b}=\{\mathbf{b}^i\}$$

és

$$\mathbf{z}^i=\{z_j\}, \quad \mathbf{b}^i=\{b_j\}, \quad j=(i-1)n+l, \quad l=1, \ldots, n.$$

Ezen jelölésekkel a (2.10) egyenletrendszer:

$$\begin{aligned} \mathbf{Bz}^1-\mathbf{z}^2&=\mathbf{b}^1, \\ -\mathbf{z}^{i-1}+\mathbf{Bz}^i-\mathbf{z}^{i+1}&=\mathbf{b}^i, \quad i=2, \ldots, n-1, \\ -\mathbf{z}^{n-1}+\mathbf{Bz}^n&=\mathbf{b}^n \end{aligned} \tag{4.3}$$

alakban írható fel.

A (4.1) alakú mátrixok mellett ebben a pontban annál általánosabb esetet is vizsgálunk:

$$\mathbf{A} = [\mathbf{A}_j, \mathbf{B}_j, \mathbf{C}_j]. \tag{4.4}$$

ahol $\mathbf{A}$ $n\times n$-es blokkból álló hiper mátrix, a blokkok mérete $m\times m$.

A (4.4) mátrixszal felírt egyenletrendszer pedig

$$\begin{aligned} &\mathbf{B}_1\mathbf{z}^1+\mathbf{C}_1\mathbf{z}^2 = \mathbf{b}^1, \\ &\mathbf{A}_i\mathbf{z}^{i-1}+\mathbf{B}_i\mathbf{z}^i+\mathbf{C}_i\mathbf{z}^{i+1} = \mathbf{b}^i, \quad 2 \leqq i \leqq n-1 \\ &\mathbf{A}_n\mathbf{z}^{n-1}+\mathbf{B}_n\mathbf{z}^n = \mathbf{b}^n. \end{aligned} \tag{4.5}$$

Lineáris egyenletrendszerek megoldásának legismertebb véges módszere a *Gauss-elimináció*, ami a (4.5) egyenletrendszer megoldására megfogalmazható a következő képletekkel:

$$\mathbf{v}^1 = \mathbf{B}_1^{-1}\mathbf{b}^1 \tag{4.6}$$

$$\left.\begin{aligned} \mathbf{v}^j &= (\mathbf{A}_j\mathbf{G}_{j-1}+\mathbf{B}_j)^{-1}(\mathbf{b}^j-\mathbf{A}_j\mathbf{v}^{j-1}) \\ \mathbf{G}_j &= -(\mathbf{A}_j\mathbf{G}_{j-1}+\mathbf{B}_j)^{-1}\mathbf{G}_j \end{aligned}\right\} \quad 2 \leqq j \leqq n, \tag{4.7}$$

$$\begin{aligned} \mathbf{z}^n &= \mathbf{v}^n, \\ \mathbf{z}^j &= \mathbf{v}^j+\mathbf{G}_j\mathbf{z}^{j+1}, \quad 1 \leqq j \leqq n-1. \end{aligned} \tag{4.8}$$

A *Gauss-elimináció* (4.6), (4.7) és (4.8)-ban megfogalmazott alakját faktorizációs módszer néven, az orosz irodalomban pedig *„pragonka"* elnevezéssel szokták használni. A (4.6) és (4.7) képletekkel számolt $\mathbf{G}_j$, $j=1, \ldots, n$ mátrixokat tárolnunk kell a (4.8) képletek használatához, ezért a módszer tárolási igénye $m=n$ esetben $O(n^3)$ nagyságrendű (n darab, n méretű mátrix tárolása), míg a számolás műveleti igénye $O(n^4)$ nagyságrendű (n-szer kell n méretű mátrixot invertálni).

A számolási munka csökkenthető a (4.3) egyenlet megoldására a $\mathbf{B}$ mátrix (4.2) tulajdonságát felhasználva. A megoldás ekkor megfogalmazható a következő módon. (Ez több irodalmi forrásban megtalálható, mi a [6] és [7] dolgozatot használjuk.) Tekintsük a következő mátrixokat:

$$\begin{aligned} \mathbf{P}_0(\mathbf{B}) &= \mathbf{I} \\ \mathbf{P}_j(\mathbf{B}) &= \prod_{i=1}^{j}[\mathbf{B}-r_i(j)\mathbf{I}], \quad r_i(j) = 2\cos\frac{i\pi}{j+1}. \end{aligned} \tag{4.9}$$

Ezek segítségével a (4.3) megoldása felírható a következő módon

$$\begin{aligned} \mathbf{z}^n &= \mathbf{P}_n^{-1}(\mathbf{B})\sum_{i=1}^{n}(-1)^{i+n}\mathbf{P}_{i-1}(\mathbf{B})\mathbf{b}_i \\ \mathbf{z}^j &= \mathbf{P}_j^{-1}(\mathbf{B})\left[\sum_{i=1}^{j}(-1)^{i+j}\mathbf{P}_{i-1}(\mathbf{B})\mathbf{b}_i-\mathbf{P}_{j-1}(\mathbf{B})\mathbf{z}^{(j+1)}\right], \quad 1 \leqq j \leqq n-1. \end{aligned} \tag{4.10}$$

A (4.9) és (4.10) képletek felhasználásával a (4.3) egyenletrendszer megoldása $6n^3$ nagyságrendű műveletben végezhető el.

A 60-as években törekvések folytak olyan módszerek kidolgozására, amelyek alkalmazásával a műveleti igény csökkenthető. Például ezzel kapcsolatos eredmény található a [12]-es dolgozatban. A dolgozat a (4.3) megoldására a

$$\begin{aligned} \mathbf{v}^1 &= \mathbf{b}^1, \\ \mathbf{v}^j &= \mathbf{P}_{j-1}(\mathbf{B})\mathbf{b}^j - \mathbf{v}^{j-1}, \quad 2 \leqq j \leqq n, \\ \mathbf{z}^n &= \mathbf{P}_n^{-1}(\mathbf{B})\mathbf{b}^n, \\ \mathbf{z}^{n-1} &= \mathbf{b}^n \mathbf{B}\mathbf{z}^n, \\ \mathbf{z}^j &= \mathbf{b}^{j+1} - \mathbf{B}\mathbf{z}^{j+1} - \mathbf{z}^{j+2}, \quad 1 \leqq j \leqq n-2 \end{aligned} \tag{4.11}$$

megoldást javasolja, ami $\frac{3}{2}n^3$ nagyságrendű műveletet igényel.

A fenti eredmények ismerete nélkül a szerző hasonló eredményekhez jutott a 70-es évek elején. A megoldás gondolatmenete a következőképpen fogalmazható meg a (4.3) egyenletrendszer megoldására:

Tegyük fel, hogy ismerjük a $\mathbf{z}^1$-et. Ekkor rendre kiszámolható

$$\begin{aligned} \mathbf{z}^2 &= \mathbf{B}\mathbf{z}^1 - \mathbf{b}_1 \\ \mathbf{z}^3 &= (\mathbf{B}^2 - \mathbf{I})\mathbf{z}^1 - \mathbf{B}\mathbf{b}^1 - \mathbf{b}^2 \\ \mathbf{z}^4 &= (\mathbf{B}^3 - 2\mathbf{B})\mathbf{z}^1 - (\mathbf{B}^2 - \mathbf{I})\mathbf{b}^1 - \mathbf{B}\mathbf{b}^2 - \mathbf{b}^3 \end{aligned}$$

és általában

$$\mathbf{z}^s = \mathbf{R}_s\mathbf{z}^1 - \mathbf{R}_{s-1}\mathbf{b}^1 - \mathbf{R}_{s-2}\mathbf{b}^2 - \ldots - R_1\mathbf{b}^{s-1} \tag{4.12}$$

ahol

$$\mathbf{R}_1 = \mathbf{I}, \quad \mathbf{R}_2 = \mathbf{B}, \quad \mathbf{R}_s = \mathbf{B}\mathbf{R}_{s-1} - \mathbf{R}_{s-2}, \quad s = 3, \ldots, k-1.$$

Fejezzük ki (4.12) segítségével a $\mathbf{z}^{n-1}$-et és $\mathbf{z}^n$-et és helyettesítsük azokat a (4.3) rendszer utolsó egyenletébe, azt kapjuk, hogy

$$(\mathbf{B}\mathbf{R}_n - \mathbf{R}_{n-1})\mathbf{z}^1 = \mathbf{b}^n + \mathbf{B}(\mathbf{R}_{n-1}\mathbf{b}^1 + \ldots + \mathbf{R}_1\mathbf{b}^{n-1}) - (\mathbf{R}_{n-1}\mathbf{b}^1 + \ldots + \mathbf{R}_1\mathbf{b}^{n-1}),$$

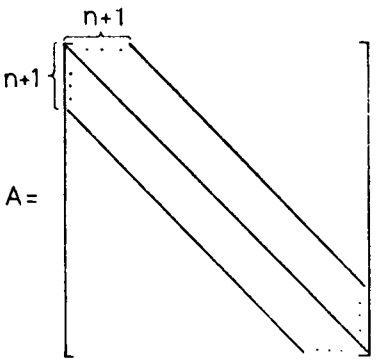
amiből $\mathbf{z}^1$ kifejezhető

$$\mathbf{z}^1 = \mathbf{R}_{n+1}^{-1}(\mathbf{B} - \mathbf{I})[\mathbf{R}_{n-1}\mathbf{b}^1 + \ldots + \mathbf{R}_1\mathbf{b}^{n-1}] + \mathbf{b}^n.$$

Ezt (4.12)-be helyettesítve megkapjuk a teljes megoldást. A fenti megoldás tárolási igénye n^3, műveleti igénye $5n^3$. A [13] dolgozatban ismertetett módszer a most leírt gondolatmeneten alapszik, de számolás-szervezése eltér attól. A módszer használatához mindössze $7n^2$ tárolási igény szükséges, a műveleti igénye $\left(5+\frac{1}{3}\right)n^3$. A [13] dolgozatban leírt módszer nem használja ki a $\mathbf{B}$ mátrix (4.2)-es specialitását, ezért bizonyos megszorítások mellett használható a (4.5) típusú egyenletrendszerek megoldására is.

A (4.10), (4.11) és a szerző által kidolgozott módszerek alapvető hiányossága, hogy azok nem stabil módszerek. Ennek a problémának az elemzésére még visszatérünk.

Az eddig tárgyalt véges módszerek (kivéve a [13] dolgozat módszerét) sajátsága, hogy olyan egyenletrendszert old meg, amelynek a mátrixa egy hipermátrix, aminek az elemei mátrixok. Azonban az eddigiekben tárgyalt lineáris egyenletrendszerek tekinthetők szalagmátrixú lineáris egyenletrendszereknek: az **A** mátrixnak n^2 sora és oszlopa van, a szalagszélesség $2n+1$ (lásd 2. ábra).



2. ábra

Az egyenletrendszer megoldására alkalmazhatjuk a szalagmátrixú lineáris egyenletrendszerek megoldására kidolgozott módszereket. A *Gauss-elimináció* szalagmátrixra kidolgozott változata használja a teljes szalag sávot, hiszen az elimináció közben telítődik az ottani, kezdetben ritka szalagsáv. Ezért a *Gauss-elimináció* használata $n^2(2n+1) \sim 2n^3$ helyet igényel, műveleti igénye pedig $n^4/3$ nagyságrendű.

Az egyenletrendszerek számításánál problémaként jelentkezik, hogy az egyenletrendszer mátrixa nem fér el a gép gyorsmemóriájában. Ilyenkor igénybe kell venni a megoldásához a háttértárolókat. A gyorsmemória és a háttértároló közti átmenetek erősen lassítják a számítási időt, ezért a módszerek hatékonyságának lényeges része, hogy mennyi a tárolók közötti adatmozgatás.

Ha az **A** mátrixot hipermátrixnak tekintjük, az adatmozgatás a hipermátrix elemeinek (amik ugyancsak mátrixok) a mozgatását jelentik. A szerző kidolgozott olyan módszereket és programokat, amelyek segítségével nagyméretű szalagmátrixú lineáris egyenletrendszerek oldhatók meg aránylag kevés háttértároló mozgatásával.

A [10]-es dolgozatban R. E. BANK és D. J. ROSE a kétdimenziós peremértékfeladat megoldásának új lehetőségeit tárgyalják. A megoldás alapötlete azonos a 70-es évek elején a szerző által vizsgált, már említett meggondolásokkal. Módszerüket *„marching algoritmus"*-nak nevezik. Ennek lényege röviden összefoglalva a következő:

Legyen az **A** mátrix (4.1) alakú. Az $\mathbf{Ax}=\mathbf{b}$ egyenletrendszer megoldását kezdjük azzal, hogy átrendezzük az egyenletrendszert úgy, hogy az egyenletrendszer

első n egyenletét az egyenletrendszer végére visszük. Így az átrendezett egyenletrendszer mátrixa

$$\bar{\mathbf{A}} = \begin{bmatrix} -\mathbf{I} & \mathbf{B} & -\mathbf{I} & & & \\ & -\mathbf{I} & \mathbf{B} & -\mathbf{I} & & \\ & \ddots & \ddots & \ddots & & \\ & & & -\mathbf{I} & \mathbf{B} & -\mathbf{I} \\ & & & & -\mathbf{I} & \mathbf{B} \\ \mathbf{B} & -\mathbf{I} & \dots & & & \mathbf{0} \end{bmatrix} = \begin{vmatrix} \mathbf{U}_1 & \mathbf{R} \\ \mathbf{C} & \mathbf{0} \end{vmatrix}$$

alakú lesz. Faktorizáljuk ezt a mátrixot:

$$\begin{bmatrix} \mathbf{U}_1 & \mathbf{R} \\ \mathbf{C} & \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{C}\mathbf{U}_1^{-1} & \mathbf{L}_2 \end{bmatrix} \begin{bmatrix} \mathbf{U}_1 & \mathbf{R} \\ \mathbf{0} & \mathbf{U}_2 \end{bmatrix}. \tag{4.13}$$

A részmátrixok között fennáll az

$$\mathbf{F} = -\mathbf{C}\mathbf{U}_1^{-1}\mathbf{R} = \mathbf{L}_2\mathbf{U}_2 \tag{4.14}$$

összefüggés. A (4.13) jobb oldala első tényezőjének utolsó sora így írható:

$$\mathbf{P} = [\mathbf{C}\mathbf{U}_1^{-1}\dots\mathbf{F}] = [\mathbf{S}_1\mathbf{S}_2\dots\mathbf{S}_{n-1}-\mathbf{S}_n], \tag{4.15}$$

ahol $S_0=\mathbf{I}$, $\mathbf{S}_1=\mathbf{B}$, $\mathbf{S}_{i+1}=\mathbf{B}\mathbf{S}_i-\mathbf{S}_{i-1}$.

Az $\mathbf{S}_i$ mátrixok úgynevezett *Csebisev-sorozatot* alkotnak és teljesül rájuk, hogy

$$\mathbf{S}_n = \prod_{i=1}^{n} (\mathbf{B}-r_n(i)\mathbf{I}), \quad r_i(j) = 2\cos\frac{j\pi}{i+1}, \quad 1 \leqq j \leqq i.$$

A (4.13) bontásnak megfelelően bontsuk fel az $\mathbf{x}$ és a $\mathbf{b}$ vektorokat $\mathbf{x}=(\mathbf{x}_1, \mathbf{x}_2)$, $\mathbf{b}=(\mathbf{b}_1, \mathbf{b}_2)$ részekre, akkor felírható, hogy

$$\bar{\mathbf{A}}\mathbf{x} = \begin{vmatrix} \mathbf{U}_1 & \mathbf{R} \\ \mathbf{C} & \mathbf{0} \end{vmatrix} \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} = \begin{vmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{C}\mathbf{U}_1^{-1} & \mathbf{L}_2 \end{vmatrix} \begin{vmatrix} \mathbf{U}_1 & \mathbf{R} \\ \mathbf{0} & \mathbf{U}_2 \end{vmatrix} \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{pmatrix}. \tag{4.16}$$

A (4.16) egyenletrendszer megoldását felbontjuk két egymásutáni egyenletrendszer megoldására:

$$\begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{C}\mathbf{U}_1^{-1} & \mathbf{L}_2 \end{pmatrix} \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{pmatrix} \tag{4.17}$$

$$\begin{pmatrix} \mathbf{U}_1 & \mathbf{R} \\ \mathbf{0} & \mathbf{U}_2 \end{pmatrix} \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix}. \tag{4.18}$$

A (4.17) egyenletrendszerből $\mathbf{y}_1=\mathbf{b}_1$ és

$$\mathbf{C}\mathbf{U}_1^{-1}\mathbf{y}_1+\mathbf{L}_2\mathbf{y}_2 = \mathbf{b}_2. \tag{4.19}$$

(4.19) első tagja kiszámítható az $\mathbf{U}_1\mathbf{z}=\mathbf{y}_1$ egyenletrendszer megoldása $\mathbf{z}=\mathbf{U}_1^{-1}\mathbf{y}_1$ segítségével és így (4.14) és (4.15) segítségével felírható, hogy

$$\mathbf{S}_n\mathbf{x}_2 = \prod_{i=1}^{n} (\mathbf{B}-r_n(i)\mathbf{I})\mathbf{x}_2 = -\mathbf{C}\mathbf{z}+\mathbf{b}_2, \tag{4.20}$$

amiből $\mathbf{x}_2$ nyerhető. Ezután $\mathbf{x}_1$-re megoldandó az

$$\mathbf{U}_1\mathbf{x}_1 = \mathbf{y}_1 - \mathbf{R}\mathbf{x}_2 = \mathbf{b}_1 - \mathbf{R}\mathbf{x}_2$$

egyenletrendszer, és ezáltal megkaptuk a teljes megoldást.

Az eljárásban lényeges, hogy az $\mathbf{U}_1$ mátrix háromszög mátrix, így a hozzá tartozó egyenletrendszerek megoldása nagyon gyors. Másrészt a (4.20) egyenlet mátrixa háromátlós mátrixok szorzata, így az is gyorsan megoldható. A módszer számolási műveleti igénye $11n^2-16n+4$ multiplikatív művelet, tárolási igénye pedig $6n^2-4n$.

A *marching algoritmus* nagyon gyors, kevés memóriát igénylő módszer. Hibája viszont, hogy erősen instabil.

A [8] és a [9] dolgozatok részletesen foglalkoznak a *marching algoritmussal* és az instabilitási probléma elkerülésére vizsgálják az *„általánosított marching algoritmust"*. Ezáltal javul a módszer stabilitása.

5. További módszerek

A 2. pont jelöléseit használva, legyen

$$L_i u = \frac{1}{h_i}(u_{k+1} - 2u_k + u_{k-1}), \quad i = 1, 2$$

és tekintsük a következő sajátérték-feladatot: $L_2 v + \lambda v = 0$, ha y belső pont és $v(0) = v(l_2) = 0$. Ennek a megoldása a következő módon írható fel:

$$u_k(jh_2) = \sqrt{\frac{2}{l_2}} \sin \frac{jh_2\pi}{l_2}, \quad \lambda_k = \frac{4}{h_2^2} \sin^2 \frac{j\pi h_2}{2l_2}, \quad j = 1, \ldots, n.$$

Tekintsük most a 2. pontban megfogalmazott

$$Lu = -f \tag{5.1}$$

egyenletet homogén peremfeltételek mellett. Keressük ennek megoldását

$$u_{ij} = \sum_{k=1}^{n} c_k(ih_1) v_k(jh_2)$$

alakban. Akkor a c_k úgynevezett *Fourier-együtthatók* az

$$L_1 c_k - \lambda_k c_k = -f_k, \quad k = 1, \ldots, n$$

differenciaegyenlet homogén peremfeltételek melletti megoldásából nyerhetők. Az (5.1) egyenlet ilyen módon való numerikus megoldását *Fourier-módszernek* nevezik.

Vázlatosan ismertetjük még az úgynevezett *„ciklikus redukciós módszert"* (lásd [7]).

Legyen a (4.5) egyenletben $\mathbf{A}_i = \mathbf{C}_i = \mathbf{T}$ és $\mathbf{B}_i = \mathbf{B}$, akkor a (4.5) egyenletrendszer három egymásutáni egyenlete:

$$\begin{aligned} \mathbf{T}\mathbf{z}^{j-2} + \mathbf{B}\mathbf{z}^{j-1} + \mathbf{T}\mathbf{z}^{j} &= \mathbf{b}^{j+1}, \\ \mathbf{T}\mathbf{z}^{j-1} + \mathbf{B}\mathbf{z}^{j} + \mathbf{T}\mathbf{z}^{j+1} &= \mathbf{b}^{j}, \\ \mathbf{T}\mathbf{z}^{j} + \mathbf{B}\mathbf{z}^{j+1} + \mathbf{T}\mathbf{z}^{j+2} &= \mathbf{b}^{j+1}. \end{aligned} \tag{5.2}$$

Szorozzuk meg ezek közül az elsőt és az utolsót **T**-vel, a középsőt **B**-vel és adjuk össze őket, azt kapjuk, hogy

$$\mathbf{T}\mathbf{z}^{j-2}+(2\mathbf{T}^2-\mathbf{A}^2)\mathbf{z}^j+\mathbf{T}^2\mathbf{z}^{j+2}=\mathbf{T}\mathbf{b}^{j-1}-\mathbf{A}\mathbf{b}^j+\mathbf{b}^{j+1}. \tag{5.3}$$

Az (5.3) típusú egyenletek hasonló szerkezetűek, mint az (5.2) kiindulási egyenletek voltak. Az (5.3) egyenletekből minden második ismeretlen hiányzik. Ezáltal az egyenletek és az ismeretlenek számát felére redukáltuk. Ezt a redukciós lépést többször is megismételhetjük. A több lépésben redukált kevés egyenletet és ismeretlent tartalmazó egyenletrendszert megoldva és a megoldása segítségével a kiküszöbölt ismeretlenek rendre kiszámíthatók.

A *Fourier- és ciklikus redukciós módszer* segítségével az n^2 egyenletből álló és n^2 ismeretlent tartalmazó egyenletrendszer is $O(n^2 \log_2 n)$ lépésszámban megoldható. De mindkét módszer használatát korlátozza az a körülmény, hogy csak vizsgált egyenletrendszer megoldására használható az adott tartományon és az általánosítás nehézkes, vagy nem lehetséges.

6. Stabilitás

A dolgozatban többször utaltunk a vizsgált módszer stabilitásának kérdésére. Ennek a fontos követelménynek alapos vizsgálata megtalálható a [3] könyvben. Röviden úgy lehetne megfogalmazni, hogy a módszer stabilitása a módszer használhatóságát jelenti.

A dolgozatban tárgyalt feladatok kapcsán kétféle stabilitási kérdést lehet felvetni: 1. a differenciálegyenlet közelítő megoldására megfogalmazott differenciaegyenlet megoldásának stabilitási kérdése; 2. a differenciaegyenlet megoldására alkalmazott numerikus módszer stabilitási kérdése. Először a felvetett első kérdésre adjuk meg a választ.

Mint ahogy azt a 2. pontban tárgyaltuk, a fellépő

$$\mathbf{A}\mathbf{y}=\mathbf{b} \tag{6.1}$$

differenciaegyenletünk mátrixa pozitív definit, így a (6.1) egyenletrendszer mindig egyértelműen megoldható. A pozitív definitség miatt tetszőleges y vektorra fennáll, hogy

$$(\mathbf{A}\mathbf{y}, \mathbf{y}) \geqq \delta \|\mathbf{y}\|^2, \tag{6.2}$$

ahol $\delta>0$. Szorozzuk meg a (6.1) egyenletünk mindkét oldalát skalárisan az **y** vektorral. A *Cauchy—Bunyakovszkij-egyenlőtlenség* felhasználásával kapjuk, hogy

$$(\mathbf{A}\mathbf{x}, \mathbf{y}) = (\mathbf{b}, \mathbf{y}) \leqq \|\mathbf{b}\| \, \|\mathbf{y}\|. \tag{6.3}$$

A (6.2) és a (6.3) egyenlőtlenségből adódik, hogy

$$\delta \|\mathbf{y}\|^2 \leqq \|\mathbf{b}\| \, \|\mathbf{y}\|,$$

vagyis

$$\|\mathbf{y}\| \leqq \frac{1}{\delta} \|\mathbf{b}\|. \tag{6.4}$$

Legyen most a vizsgált differenciálegyenletünk pontos megoldása az $u(\mathbf{x})$ függvény, a differenciálegyenletet approximáló (6.1) alakú differenciaegyenlet megoldása

pedig $y_i = y(x_i)$. A $z_i = y_i - u(x_i)$ különbségre felírható az

$$\mathbf{Az} = \mathbf{g}$$

egyenletrendszer, ahol $\mathbf{g} = \mathbf{b} - \mathbf{Au}$. A (6.4) egyenlőtlenség alapján fennáll, hogy

$$\|\mathbf{z}\| \leqq \frac{1}{\delta} \|\mathbf{g}\|. \tag{6.5}$$

A (6.5) egyenlőtlenségből következik, hogy a (6.1) egyenlet **y** megoldása folytonosan függ a **b** vektor megváltozásától, vagyis annak kis megváltozása, a megoldás kis megváltozását eredményezi. Minthogy a **b** vektor tartalmazza a vizsgált differenciálegyenletre előírt peremértékeket és a differenciálegyenletben szereplő inhomogenitást jelentő függvény értékeit is, ezért a most megfogalmazott stabilitás azt jelenti, hogy a peremértékek és az inhomogenitást jelentő függvény megváltozására nézve is folytonos lesz a megoldás változása. A fentebb megállapított egyértelmű megoldhatósággal együtt a most megfogalmazott stabilitás azt jelenti, hogy az elliptikus típusú parciális differenciálegyenletek megoldásának feladata differenciaegyenletek segítségével korrekt kitűzésű feladat.

Vizsgáljuk meg a felvetett 2. kérdést. Minthogy az **A** mátrix pozitív definit a (6.1) differenciaegyenletrendszer megoldása *Gauss elimináció* segítségével (*faktorizációs, pragonka módszerrel*) főelem választás nélkül is stabil (lásd [16]).

Az 5. pontban vizsgált többi véges módszer stabilitásával viszont problémák vannak. Vizsgáljuk meg ezt a kérdést az 5. pontban tárgyalt *marching algoritmus* esetében (lásd [8], [9] és [10]). Az algoritmus használata során többször is (például az $\mathbf{U}_1\mathbf{x} = \mathbf{y}$ egyenlet megoldásához) szükség van

$$\begin{aligned} \mathbf{x}_{n-1} &= -\mathbf{b}_{n-1}, \\ \mathbf{x}_{n-2} &= \mathbf{B}\mathbf{x}_{n-1} - \mathbf{b}_{n-2}, \\ \mathbf{x}_{n-j} &= \mathbf{B}\mathbf{x}_{n-j+1} - \mathbf{x}_{n-j+2} - \mathbf{b}_{n-j}, \quad 3 \leqq j \leqq n-1 \end{aligned} \tag{6.6}$$

típusú rekurziós helyettesítésre. Az említett dolgozatokból látható, hogy az itt szereplő **B** mátrix normája

$$\|\mathbf{B}\| = 4 + 2\cos\frac{\pi}{n+1} \sim 6.$$

Ezért a (6.6) rekurzióban a hibák lépésről lépésre meghatszorozódnak. Ez viszont elég sok lépés után oda vezet, hogy a kerekítési hibákból adódó, az első lépésekben még kicsi hibák olyan nagyra nőnek a rekurzió folyamán, hogy használhatatlan eredményeket kapunk. A dolgozatokból láthatóan, valamint a szerző próbaszámolásai alapján ez az instabilitási probléma (számítógéptől függően más n-nél) egy 10 decimális szóhosszal működő gépnél $n = 15, 20$ körüli értéknél jelentkezik. Ez azt jelenti, hogy $n < 15$ esetén a módszer nagyon jól, gyorsan működik, pontos eredményeket ad. $15 \leqq n \leqq 18$ esetekben az eredmények nagyon pontatlanok lesznek. $n = 20$ esetén viszont teljesen használhatatlan eredményeket kapunk. Sőt $n > 20$ esetben a gépi lebegőpontos túlcsordulás veszélye is fennáll. A súlyos instabilitási probléma miatt a gyorsnak tűnő *marching algoritmus* nagy n-ek esetén használhatatlan. Mint általában a numerikus módszerek területén az elliptikus típusú differenciálegyenletek numerikus módszereinél is a módszerek használhatóságának alapvető kritériuma a numerikus stabilitás.

7. A módszerek összehasonlítása

A tárgyalt numerikus módszerek összehasonlításával egyidejűleg több szempontot kell összehasonlítani. Ezek közül a legfontosabbak:

1) az adott feladat megoldásához szükséges műveletek száma;
2) a módszer stabilitása;
3) a számítógépen való használhatóság szempontjai;
4) az általánosíthatóság.

A felsorolt szempontok néha ellentmondó követelményeket állítanak elénk. Például tekintsük a 4. pontban tárgyalt *Gauss-eliminációt*. Ennek műveleti igénye, mint láttuk $O(n^4)$ nagyságrendű és a módszer stabil. Ezzel szemben a marching algoritmus műveleti igénye $O(n^2)$ nagyságrendű, de instabil. Külön problémát jelent a véges és az iterációs módszerek összehasonlítása egymással.

Az iterációs módszerek minden egyes iterációja általában $O(n^2)$ nagyságrendű műveletet igényel. Kérdés csupán az, hogy hány iterációt kell végezni. A 3. pontban tárgyalt SOR iterációnál adott ε pontosság eléréséhez az iterációk száma $O\left(n \lg \frac{1}{\varepsilon}\right)$ nagyságrendű. Ez azt jelenti, hogy a módszer műveleti igénye $O\left(n^3 \ln \frac{1}{\varepsilon}\right)$ lesz. Az iterációs módszerek rendkívüli előnye a véges módszerekkel szemben, hogy könnyen programozhatók, továbbá az egyenletrendszer mátrixa számolás közben sose változik, így nagy mátrixok is könnyen kezelhetők a számítás alatt.

A [6] dolgozatban F. W. DORR végez számszerű összehasonlítást a módszerek között. A [17] dolgozatban A. A. SZAMARSZKIJ azt vizsgálja, hogy milyen esetekben milyen módszert (melyik iterációs vagy véges módszert) célszerű használni. A dolgozat összehasonlítást tesz a legújabb eredmények és a korábban ismert módszerek között is. Az alábbiakban két táblázatot közlünk a módszerek összehasonlítására, amik a [6] dolgozat táblázataiból és a [17] dolgozatból, valamint a jelen dolgozat eredményeiként összegezhetők.

1. TÁBLÁZAT

	Szukcesszív túlrelaxálás	Változó irányok módszere	Váltakozó háromszög módszer
iterációk száma	$\frac{n}{\pi} \ln n$	$0{,}94 \log_{10} n \log_e n$	$\log_e \frac{2}{\varepsilon} \Big/ 3{,}54 \sqrt{\frac{1}{n}}$
műveletek száma interációnként	$7n^2$	$20n^2$	$18n^2$
teljes műveleti igény	$\frac{3}{2} n^3 \log_2 n$	$4n^2 (\log_2 n)^2$	$\frac{18n^2 \log_e \frac{2}{\varepsilon}}{3{,}54 \sqrt{\frac{1}{n}}}$

(n a felosztás száma, ε a pontosság)

2. TÁBLÁZAT

Módszer	Műveleti igény
Gauss elimináció	$\frac{n^4}{3}$
Gauss elimináció specializált változat	$6n^3$
marching algoritmus	$11n^2-16n+4$ (instabil)
Fourier módszer	$5n^2 \log_2 n$
ciklikus redukció	$\frac{9}{2} n^2 \log_2 n$
SOR	$\frac{3}{2} n^3 \log_2 n$
változó irányok módszere	$4n^2 (\log_2 n)^2$

IRODALOM

[1] Самарский, А. А., *Теория разностных схем* (Наука, Москва, 1977).
[2] Самарский, А. А., *Введение в теорию разностных хсем* (Наука, Москва, 1974).
[3] Самарский, А. А., Гулин, А. В., *Устойчивость разностных хсем* (Наука, Москва, 1973).
[4] Самарский, А. А. и Андреев, В. Б., *Разностные методы для эллиптических уравнений* (Наука, Москва, 1976).
[5] Самарский, А. А., «Новые результаты теории разностных схем», Препинт ИМП *П* 87, 1978, Москва.
[6] Dorr, F. W., "The direct solution of the directe Poisson equation on a rectangle", *SIAM Rev.* **12** (1970) 248—263.
[7] Buzbee, B. L., Galub, G. H. and Nielson, C. W., "On direct methods for solving Poisson equation", *SIAM J. Numer. Anal.* 7 (1970) 627—656.
[8] Bank, R. E. and Rose, D. J., "Marching algoritmus for elliptic boundary value problems. I: the constant coefficient case", *SIAM J. Numer. Anal.* **14** (1977) 792—829.
[9] Bank, R. E., "Marching algoritms for elliptic boundary value problems. II: the variable. coefficient case", *SIAM J. Numer. Anal.* **14** (1977) 950—69.
[10] Bank, R. E. and Rose, D. J., "An $o(n^2)$ method for solving constant coefficient boundary value problems in two dimensions", *SIAM J, Numer. Anal.* **12** (1975) 529—540.
[11] Varga, R. S., *Matrix Iterative Analysis* (Prentice-Hall, Englewood Cliffs, New Jersey, 1962)
[12] Schechter, S., "Quasi-tridiagonal matrices and type-insensitive difference equations", *Quart. Appl. Math.* **18** (1960) 285—295.
[13] Gergely, J., „Szalagmátrixú lineáris egyenletrendszerek megoldása", *MTA SZTK Közlemények* **6** (1971) 36—39.
[14] Н. С. Бахвалов, «О сходимости одного релаксадионного метода при естественных ограничениях на эллиптический операторе», *Ж. В. М. и М. Ф.* **6** 1966 861—883.
[15] Ciarlet, P. G., *The finite element method for elliptic problems* (North-Holl., Amsterdam, New York, Oxford, 1978).
[16] Wilkinson, J. H., *The algebraic eigenvalue problem* (Clarendon Press, Oxford, 1965).
[17] Szamarszkij, A. A., "New iterative methods for difference equations", Coll. Math. Soc. János Bolyai, 22. Numerical Methods, Keszthely (Hungary) North-Holl., 1977.

(Beérkezett: 1979. március 5.)

DR. GERGELY JÓZSEF

MTA SZÁMÍTÁSTECHNIKAI ÉS AUTOMATIZÁLÁSI KUTATÓ INTÉZET
1250 BUDAPEST, ÚRI U. 49.

ON NUMERICAL SOLUTION OF THE POISSON EQUATION

J. GERGELY

The paper gives a survey about the numerical solution methods of the *Poisson equation.* It is formulated the finite difference method of the *Dirichlet problem* for linear second order equations, discussed the iterative and direct solutions of difference equations and the stability of the methods is investigated. Finally a comparison among the methods is done.

Alkalmazott Matematikai Lapok 5 (1979) 233—240

A LINEÁRIS EGYENLETRENDSZEREK ÁLTALÁNOS MEGOLDÁSÁNAK EGY DIREKT MÓDSZEROSZTÁLYA

ABAFFY JÓZSEF
Budapest

A cikkben a lineáris egyenletrendszerek direkt megoldási módszereinek egy osztályát adjuk meg, amely alulhatározott egyenletrendszerek esetén *Moore—Penrose*-féle megoldást szolgáltat.

1. Bevezetés

Az utóbbi időben főleg nagyméretű lineáris egyenletrendszerek megoldására a klasszikus direkt módszereken kívül egyéb direkt eljárásokat is használunk ([1]). Ezek a módszerek azonban, bár eredményeik általában elfogadhatóak, nem tudják figyelembe venni az együtthatómátrixok esetleges speciális tulajdonságait. A továbbiakban ebből a problémából kiindulva meghatároztuk a direkt megoldási módszerek egy teljes osztályát, amely három lényegében tetszőleges, lépésenként meghatározható paramétertől függ. A paraméterek lépésenkénti megválasztására három javaslatot adunk, amelyeket egy konkrét feladat megoldása során általában kombináltan célszerű alkalmazni. A paramétereket végtelen sok módon be lehet állítani, azok megválasztása mindig a konkrét feladattól függ.

Megmutatjuk továbbá, hogy alulhatározott lineáris egyenletrendszerek esetén a módszerosztály által *Moore—Penrose*-értelemben vett megoldást kapunk.

2. A direkt eljárások egy módszerosztálya

Legyen

$$\mathbf{A}^T\mathbf{x} = \mathbf{b}, \quad \mathbf{x}, \mathbf{b} \in R^n \tag{2.1}$$

és $\mathbf{A}$ $n \times m$-es ($m \leqq n$) mátrix. Legyen az $\mathbf{A}$ mátrix rangja,

$$r(\mathbf{A}) = k, \quad k \leqq \min(m, n)$$

és az egyszerűség érdekében tegyük fel, hogy az $\mathbf{A}$ mátrix k lineárisan független sora az első k sorban helyezkedik el. A függetlenség eldöntésével kapcsolatban utalunk HUANG [1] cikkére. A módszerosztályt a következőképpen definiáljuk $i=0,1, \ldots, k-1$-re.

Legyen $\mathbf{H}_0 = \mathbf{I}$ $n \times m$-es egységmátrix, $\mathbf{x}_0 = \mathbf{0}$ és

$$\mathbf{p}_{i+1} = \mathbf{H}_i^T \mathbf{a}_{i+1}, \tag{2.2}$$

$$\mathbf{x}_{i+1} = \mathbf{x}_i - \frac{r_{i+1}(\mathbf{x}_i)}{\mathbf{p}_{i+1}^T \mathbf{a}_{i+1}} \mathbf{p}_{i+1}, \tag{2.3}$$

(2.4) $$\mathbf{y}_{i+1} = k_i \mathbf{a}_{i+2} - l_i \mathbf{a}_{i+1}, \quad k_i^2 + l_i^2 > 0,$$

$$l_i \neq \frac{k_i \mathbf{a}_{i+1}^T \mathbf{H}_i \mathbf{a}_{i+1}}{\mathbf{p}_{i+2}^T \mathbf{a}_{i+1}}.$$

(2.5) $$\mathbf{H}_{i+1} = \mathbf{H}_i + \gamma_i \mathbf{H}_i \mathbf{a}_{i+1} \mathbf{p}_{i+1}^T - \frac{\gamma_i \mathbf{p}_{i+1}^T \mathbf{a}_{i+1} + 1}{\mathbf{y}_{i+1}^T \mathbf{H}_i \mathbf{a}_{i+1}} \mathbf{H}_i \mathbf{a}_{i+1} \mathbf{y}_{i+1}^T \mathbf{H}_i,$$

ahol $r_i(\mathbf{x}) = \mathbf{a}_i^T \mathbf{x} - b_i$ a (2.1) lineáris egyenletrendszer hibavektorának i-edik komponense, $\mathbf{a}_i$ $\mathbf{A}^T$ egy sora és k_i, l_i, γ_i tetszőleges, a megadott feltételeket kielégítő konstansok.

A fentiekben definiált módszerosztály véges sok lépésben adja a lineáris egyenletrendszer egzakt vagy a *Moore—Penrose*-értelemben definiált megoldását.

Ennek bizonyításához szükségünk van a következő lemmákra.

2.1. Lemma. A (2.5) kifejezésben definiált mátrixsorozatra fennáll

(2.6) $$\mathbf{H}_i \mathbf{H}_j = \mathbf{H}_i, \quad 0 \leqq j \leqq i.$$

(2.7) $$\mathbf{H}_i^T \mathbf{H}_j = \mathbf{H}_i^T, \quad 0 \leqq j \leqq i.$$

Bizonyítás. A bizonyítást az i indexre vonatkozó teljes indukcióval végezzük el. Először a (2.6) állítást látjuk be. Minthogy $\mathbf{H}_0$ egységmátrix, $i=0$ esetén az állítás nyilvánvalóan igaz. Tegyük fel, hogy $\mathbf{H}_i \mathbf{H}_j = \mathbf{H}_i$, $0 \leqq j \leqq i$, és lássuk be érvényességét $i = i+1$-re. Ezt két lépésben végezzük el. Legyen először $i+1 > j$. Ekkor az indukciós feltevést és a (2.5), (2.2) kifejezéseket felhasználva kapjuk, hogy

$$\mathbf{H}_{i+1} \mathbf{H}_j = \mathbf{H}_i \mathbf{H}_j + \gamma_i \mathbf{H}_i \mathbf{a}_{i+1} \mathbf{a}_{i+1}^T \mathbf{H}_i \mathbf{H}_j - \frac{\gamma_i \mathbf{p}_{i+1}^T \mathbf{a}_{i+1} + 1}{\mathbf{y}_{i+1}^T \mathbf{H}_i \mathbf{a}_{i+1}} \mathbf{H}_i \mathbf{a}_{i+1} \mathbf{y}_{i+1}^T \mathbf{H}_i \mathbf{H}_j = \mathbf{H}_{i+1}.$$

Az $\mathbf{y}_{i+1}^T \mathbf{H}_i \mathbf{a}_{i+1}$ érték, a módszerosztály definíciójánál megadott k_i, l_i értékekre vonatkozó feltételek miatt nem lehet nulla. Legyen most $j = i+1$. Felhasználva ismét az indukciós feltevést, a (2.5) és a (2.2) kifejezéseket, valamint bevezetve a $\delta = \frac{\gamma_i \mathbf{p}_{i+1}^T \mathbf{a}_{i+1} + 1}{\mathbf{y}_{i+1}^T \mathbf{H}_i \mathbf{a}_{i+1}}$ jelölést, a következőt kapjuk

$$\mathbf{H}_{i+1} \mathbf{H}_{i+1} = (\mathbf{H}_i + \gamma_i \mathbf{H}_i \mathbf{a}_{i+1} \mathbf{a}_{i+1}^T \mathbf{H}_i - \delta \mathbf{H}_i \mathbf{a}_{i+1} \mathbf{y}_{i+1}^T \mathbf{H}_i).$$

$$(\mathbf{H}_i + \gamma_i \mathbf{H}_i \mathbf{a}_{i+1} \mathbf{a}_{i+1}^T \mathbf{H}_i - \delta \mathbf{H}_i \mathbf{a}_{i+1} \mathbf{y}_{i+2}^T \mathbf{H}_i) =$$

$$= \mathbf{H}_i + \gamma_i \mathbf{H}_i \mathbf{a}_{i+1} \mathbf{a}_{i+1}^T \mathbf{H}_i - \delta \mathbf{H}_i \mathbf{a}_{i+1} \mathbf{y}_{i+1}^T \mathbf{H}_i + \gamma_i \mathbf{H}_i \mathbf{a}_{i+1} \mathbf{a}_{i+1}^T \mathbf{H}_i +$$

$$+ \gamma_i^2 \mathbf{H}_i \mathbf{a}_{i+1} \mathbf{a}_{i+1}^T \mathbf{H}_i \mathbf{H}_i \mathbf{a}_{i+1} \mathbf{a}_{i+1}^T \mathbf{H}_i - \gamma_i \delta \mathbf{H}_i \mathbf{a}_{i+1} \mathbf{y}_{i+1}^T \mathbf{H}_i \mathbf{H}_i \mathbf{a}_{i+1} \mathbf{a}_{i+1}^T \mathbf{H}_i -$$

$$- \delta \mathbf{H}_i \mathbf{a}_{i+1} \mathbf{y}_{i+1}^T \mathbf{H}_i - \delta \gamma_i \mathbf{H}_i \mathbf{a}_{i+1} \mathbf{a}_{i+1}^T \mathbf{H}_i \mathbf{H}_i \mathbf{a}_{i+1} \mathbf{y}_{i+1}^T \mathbf{H}_i +$$

$$+ \delta^2 \mathbf{H}_i \mathbf{a}_{i+1} \mathbf{y}_{i+1}^T \mathbf{H}_i \mathbf{H}_i \mathbf{a}_{i+1} \mathbf{y}_{i+1}^T \mathbf{H}_i =$$

$$= \mathbf{H}_{i+1} + \gamma_i \mathbf{H}_i \mathbf{a}_{i+1} \mathbf{a}_{i+1}^T \mathbf{H}_i + \gamma_i^2 \mathbf{a}_{i+1}^T \mathbf{H}_i \mathbf{a}_{i+1} \mathbf{H}_i \mathbf{a}_{i+1} \mathbf{a}_{i+1}^T \mathbf{H}_i -$$

$$- \delta \gamma_i \mathbf{y}_{i+1}^T \mathbf{H}_i \mathbf{a}_{i+1} \mathbf{H}_i \mathbf{a}_{i+1} \mathbf{a}_{i+1}^T \mathbf{H}_i -$$

$$- \delta \mathbf{H}_i \mathbf{a}_{i+1} \mathbf{y}_{i+1}^T \mathbf{H}_i - \delta \gamma_i \mathbf{a}_{i+1}^T \mathbf{H}_i \mathbf{a}_{i+1} \mathbf{H}_i \mathbf{a}_{i+1} \mathbf{y}_{i+1}^T \mathbf{H}_i +$$

$$+\delta^2 \mathbf{y}_{i+1}^T \mathbf{H}_i \mathbf{a}_{i+1} \mathbf{H}_i \mathbf{a}_{i+1} \mathbf{y}_{i+1}^T \mathbf{H}_i =$$

$$= \mathbf{H}_{i+1} + \gamma_i (1 + \gamma_i \mathbf{a}_{i+1}^T \mathbf{H}_i \mathbf{a}_{i+1} - \delta \mathbf{y}_{i+1}^T \mathbf{H}_i \mathbf{a}_{i+1}) \mathbf{H}_i \mathbf{a}_{i+1} \mathbf{a}_{i+1}^T \mathbf{H}_i -$$

$$- \delta (1 + \gamma_i \mathbf{a}_{i+1}^T \mathbf{H}_i \mathbf{a}_{i+1} - \delta \mathbf{y}_{i+1}^T \mathbf{H}_i \mathbf{a}_{i+1}) \mathbf{H}_i \mathbf{a}_{i+1} \mathbf{y}_{i+1}^T \mathbf{H}_i .$$

Visszaírva δ helyébe a megfelelő kifejezést, és felhasználva a (2.2) képletet a zárójelben levő kifejezésekre a következőt kapjuk:

$$1 + \gamma_i \mathbf{a}_{i+1}^T \mathbf{H}_i \mathbf{a}_{i+1} - \delta \mathbf{y}_{i+1}^T \mathbf{H}_i \mathbf{a}_{i+1} = 1 + \gamma_i \mathbf{a}_{i+1}^T \mathbf{H}_i \mathbf{a}_{i+1} -$$

$$- \frac{\gamma_i \mathbf{a}_{i+1}^T \mathbf{H}_i \mathbf{a}_{i+1} + 1}{\mathbf{y}_{i+1}^T \mathbf{H}_i \mathbf{a}_{i+1}} \mathbf{y}_{i+1}^T \mathbf{H}_i \mathbf{a}_{i+1} = 0,$$

tehát

$$\mathbf{H}_{i+1} \mathbf{H}_{i+1} = \mathbf{H}_{i+1} .$$

Így a (2.6) állítást beláttuk. Minthogy a (2.7) állítást teljesen hasonló módon lehet belátni, így annak bizonyítását elhagyjuk.

2.2. Lemma. A (2.5) kifejezésben a $\mathbf{H}_{i+1} - \mathbf{H}_i$ mátrixok ($i = 0, 1, \ldots, k-2$) 1 rangú diádok, amelyeknek nem zérus sajátértéke -1 és sajátvektora $\mathbf{H}_i \mathbf{a}_{i+1}$.

Bizonyítás. A (2.5) kifejezés a következő alakba írható

$$\mathbf{D}_i = \mathbf{H}_{i+1} - \mathbf{H}_i = \mathbf{H}_i \mathbf{a}_{i+1} \left(\gamma_i \mathbf{p}_{i+1} - \frac{\gamma_i \mathbf{p}_{i+1}^T \mathbf{a}_{i+1} + 1}{\mathbf{y}_{i+1}^T \mathbf{H}_i \mathbf{a}_{i+1}} \mathbf{H}_i^T \mathbf{y}_{i+1} \right)^T ,$$

azaz $\mathbf{D}_i$ 1 rangú mátrix. Minthogy $\mathbf{D}_i$-nek egyetlen nem zérus 1 sajátértéke a diád nyoma (hiszen a többi sajátértéke nulla), így felhasználva a 2.1. lemmát és a (2.2) kifejezést λ-ra a következőt kapjuk:

$$\lambda_i = \gamma \mathbf{p}_{i+1}^T \mathbf{H}_i \mathbf{a}_{i+1} - \frac{\gamma_i \mathbf{p}_{i+1}^T \mathbf{a}_{i+1} + 1}{\mathbf{y}_{i+1}^T \mathbf{H}_i \mathbf{a}_{i+1}} \mathbf{y}_{i+1}^T \mathbf{H}_i \mathbf{H}_i \mathbf{a}_{i+1} =$$

$$= \gamma_i \mathbf{a}_{i+1}^T \mathbf{H}_i \mathbf{H}_i \mathbf{a}_{i+1} - \frac{\gamma_i \mathbf{a}_{i+1}^T \mathbf{H}_i \mathbf{a}_{i+1} + 1}{\mathbf{y}_{i+1}^T \mathbf{H}_i \mathbf{a}_{i+1}} \mathbf{y}_{i+1}^T \mathbf{H}_i \mathbf{H}_i \mathbf{a}_{i+1} = -1, \quad (i = 0, 1, \ldots, k-2),$$

amiből az is következik, hogy $\mathbf{H}_i \mathbf{a}_{i+1}$ sajátvektor és ezzel a lemma állítását beláttuk. A fenti lemmák segítségével bizonyítható, hogy módszerosztályunk véges sok lépésben megoldáshoz vezet.

2.3. Tétel. A (2.2)—(2.5) kifejezések által definiált módszerosztályra teljesül, hogy

(2.8) $$r_j(\mathbf{x}_i) = 0, \quad 1 \leqq j \leqq i,$$

és

(2.9) $$r_j(\mathbf{x}) = 0, \quad 1 \leqq j \leqq i$$

esetén, ahol

(2.10) $$\mathbf{x} = \mathbf{x}_i + \mathbf{H}_i^T \mathbf{q}, \quad \mathbf{q} \in R^n$$

tetszőleges.

Bizonyítás. A (2.5) kifejezés jól definiált, azaz a nevező nem válhat zérussá a k_i, l_i értékekre vonatkozó feltételek miatt.

A (2.3) kifejezés jól definiáltságát lépésenként látjuk be (2.13), (2.20).

Minthogy felhasználva (2.5)-öt

$$\mathbf{H}_1\mathbf{a}_1 = \mathbf{H}_0\mathbf{a}_1 + \gamma_0\mathbf{H}_0\mathbf{a}_1\mathbf{p}_1^T\mathbf{a}_1 - \frac{\gamma_0\mathbf{p}_1^T\mathbf{a}_1 + 1}{\mathbf{y}_1^T\mathbf{H}_0\mathbf{a}_1}\mathbf{H}_0\mathbf{a}_1\mathbf{y}_1^T\mathbf{H}_0\mathbf{a}_1 = \mathbf{0}, \tag{2.11}$$

(2.2) és (2.11) miatt

$$\mathbf{p}_2^T\mathbf{a}_1 = \mathbf{a}_2^T\mathbf{H}_1\mathbf{a}_1 = 0, \tag{2.12}$$

és

$$\mathbf{H}_2\mathbf{a}_1 = \mathbf{H}_1\mathbf{a}_1 + \gamma_1\mathbf{H}_1\mathbf{a}_2\mathbf{p}_2^T\mathbf{a}_1 - \frac{\gamma_1\mathbf{p}_2^T\mathbf{a}_2 + 1}{\mathbf{y}_2^T\mathbf{H}_1\mathbf{a}_2}\mathbf{H}_1\mathbf{y}_2^T\mathbf{H}_1\mathbf{a}_1 = \mathbf{0}.$$

A $\mathbf{H}_2\mathbf{a}_2$-re (2.11)-hez hasonlóan azt kapjuk, hogy $\mathbf{H}_2\mathbf{a}_2 = \mathbf{0}$.

Az $\mathbf{x}_0 = \mathbf{0}$ és $r_i(\mathbf{x})$ definíciója miatt

$$r_1(\mathbf{x}_1) = r_1\left(\mathbf{x}_0 - \frac{r_1(\mathbf{x}_0)}{\mathbf{p}_1^T\mathbf{a}_1}\mathbf{p}_1\right) = 0, \tag{2.13}$$

(hiszen $\mathbf{p}_1^T\mathbf{a}_1 > 0$, mert $\mathbf{H}_0$ egységmátrix), valamint (2.12) miatt

$$r_1(\mathbf{x}_2) = r_1\left(\mathbf{x}_1 - \frac{r_2(\mathbf{x}_1)}{\mathbf{p}_2^T\mathbf{a}_2}\mathbf{p}_2\right) = r_1(\mathbf{x}_1) - \frac{r_2(\mathbf{x}_1)\mathbf{a}_1^T}{-\mathbf{p}_2^T\mathbf{a}_2}\mathbf{p}_2 = 0,$$

($\mathbf{p}_2^T\mathbf{a}_2 > 0$ voltát később látjuk be, l. (2.20))

$$r_2(\mathbf{x}_2) = r_2(\mathbf{x}_1) - \frac{r_2(\mathbf{x}_1)\mathbf{a}_2^T}{\mathbf{p}_2^T\mathbf{a}_2}\mathbf{p}_2 = 0,$$

és a (2.13), (2.11) kifejezések miatt

$$r_1(\mathbf{x}) = r_1(\mathbf{x}_1 + \mathbf{H}_1^T\mathbf{q}) = r_1(\mathbf{x}_1) + \mathbf{a}_1^T\mathbf{H}_1\mathbf{q} = 0.$$

Indukciós lépésként tehát feltehetjük a következőket

$$\mathbf{H}_i\mathbf{a}_j = \mathbf{0}, \quad 1 \leqq j \leqq i, \tag{2.14}$$

$$\mathbf{p}_i^T\mathbf{a}_j = 0, \quad 1 \leqq j < i, \tag{2.15}$$

$$r_j(\mathbf{x}_i) = 0, \quad 1 \leqq j \leqq i, \tag{2.16}$$

$$r_j(\mathbf{x}) = 0, \quad 1 \leqq j \leqq i, \tag{2.17}$$

$\mathbf{x}$ (2.8) szerinti, és

$$\mathbf{p}_i^T\mathbf{a}_i > 0. \tag{2.18}$$

Először azt mutatjuk meg, hogy a (2.14) indukciós feltevésből az következik, hogy

$$\mathbf{H}_{i+1}\mathbf{a}_j = \mathbf{0}, \quad 1 \leqq j \leqq i+1.$$

Felhasználva ugyanis $\mathbf{p}_i$ (2.2)-beli definícióját és a (2.5) kifejezést, azt kapjuk, hogy $\mathbf{H}_{i+1}\mathbf{a}_j=\mathbf{0}$, $1\leqq j<i+1$ az indukciós feltevés miatt, másrészt a (2.11)-hez teljesen hasonlóan adódik, hogy

$$\mathbf{H}_{i+1}\mathbf{a}_{i+1}=\mathbf{0},$$

tehát

$$\mathbf{H}_{i+1}\mathbf{a}_j=\mathbf{0},\quad 1\leqq j\leqq i+1. \tag{2.19}$$

(Az $\mathbf{y}_{i+1}^T\mathbf{H}_i\mathbf{a}_{i+1}$ nevező értéke nem válhat zérussá a k_i, l_i értékekre vonatkozó feltételek miatt.)

Meggondolásainkat most a (2.18) feltételre alkalmazzuk. Először is belátjuk, hogy

$$r(\mathbf{H}_i)=n-i.$$

Minthogy $\mathbf{H}_0$ n-edrendű egységmátrix és minden lépésben a 2.2. lemma miatt egy 1 rangú -1-es sajátértékű mátrixszal módosítunk, és $\mathbf{H}_i$-nek a $\mathbf{H}_i\mathbf{a}_{i+1}$ szintén sajátvektora, $\mathbf{H}_i$ rangja legalább $n-i$. Az előbb igazolt (2.14) összefüggés miatt viszont a rang legfeljebb $n-i$, tehát $r(\mathbf{H}_i)=n-i$. Továbbá a (2.14) és a 2.2. lemma miatt $\mathbf{H}_i$ $n-i$ db nem nulla sajátértéke csak 1 lehet. Így, mivel $\mathbf{a}_{i+1}$ lineárisan független az $\mathbf{a}_j$, $j=1, 2, \ldots, i$ vektoroktól, és ezek a $\mathbf{H}_i$ zérus sajátértékeihez tartozó sajátvektorok, következik az, hogy

$$\mathbf{p}_{i+1}^T\mathbf{a}_{i+1}=\mathbf{a}_{i+1}^T\mathbf{H}_i\mathbf{a}_{i+1}>0,\quad i+1\leqq k\text{-ra}, \tag{2.20}$$

tehát a (2.3) kifejezés nevezője sem válhat zérussá.

A (2.15) indukciós feltevés helyessége $i=i+1$-re (2.19)-ből azonnal adódik, tehát

$$\mathbf{p}_{i+1}^T\mathbf{a}_j=0,\quad 1\leqq j<i+1. \tag{2.21}$$

A (2.21) kifejezésből, a (2.3) definícióból és a (2.16) indukciós feltevésből következik, hogy

$$r_j(\mathbf{x}_{i+1})=r_j\left(\mathbf{x}_i-\frac{r_{i+1}(\mathbf{x}_i)}{\mathbf{p}_{i+1}^T\mathbf{a}_{i+1}}\mathbf{p}_{i+1}\right)=r_j(\mathbf{x}_i)-\frac{r_{i+1}(\mathbf{x}_i)}{\mathbf{p}_{i+1}^T\mathbf{a}_{i+1}}\mathbf{a}_j^T\mathbf{p}_{i+1}=0,\quad j\leqq i-1$$

és

$$r_{i+1}(\mathbf{x}_{i+1})=r_{i+1}\left(\mathbf{x}_i-\frac{r_{i+1}(\mathbf{x}_i)}{\mathbf{p}_{i+1}^T\mathbf{a}_{i+1}}\mathbf{p}_{i+1}\right)=r_{i+1}(\mathbf{x}_i)-\frac{r_{i+1}(\mathbf{x}_i)}{\mathbf{p}_{i+1}^T\mathbf{a}_{i+1}}\mathbf{a}_{i+1}^T\mathbf{p}_{i+1}=0$$

$$r_j(\mathbf{x}_{i+1})=0,\quad 1\leqq j\leqq i+1\quad\text{esetén}. \tag{2.22}$$

Végül, felhasználva a (2.22), (2.19) kifejezéseket, a (2.10) definíciót, (2.17)-re $i=i+1$ esetén azt kapjuk, hogy

$$r_j(\mathbf{x})=r_j(\mathbf{x}_{i+1}+\mathbf{H}_{i+1}\mathbf{q})=r_j(\mathbf{x}_{i+1})+\mathbf{a}_j^T\mathbf{H}_{i+1}\mathbf{q}=0,$$

amivel a tétel állításait beláttuk.

3. Speciális esetek

Ebben a fejezetben a 2.3. tétel két speciális esetével foglalkozunk:

a) $r(\mathbf{A})=n$, $\mathbf{A}$ $n\times m$-es négyzetes mátrix.

Ebben az esetben a megoldás egyértelműen, legfeljebb n lépésben meghatározódik, hiszen $r(\mathbf{H}_n)=0$ és így a (2.9) kifejezésben $\mathbf{x}\equiv\mathbf{x}_n$, (a szorzások száma $1/2\cdot n^3$ nagyságrendű).

b) $r(\mathbf{A})=k$, $k<n$, $\mathbf{A}^T$ $m\times n$-es mátrix, $m<n$.

Ekkor a (2.1) lineáris egyenletrendszer alulhatározott. A *Moore—Penrose*-értelemben vett $\mathbf{x}^*$ megoldásra teljesülnie kell az alábbi két feltételnek:

$$\mathbf{A}^T\mathbf{x}^*-\mathbf{b}=\mathbf{0}$$

és

$$|\mathbf{x}^*|=\min.$$

Az általános sémából következő $\mathbf{x}_k$ megoldásra az első feltétel nyilvánvalóan teljesül. A 2. feltétel teljesülését a következőképpen látjuk be. Minthogy (2.10) szerint a megoldások

$$\mathbf{x}=\mathbf{x}_k+\mathbf{H}_k^T\mathbf{q}$$

alakban felírhatók

$$(\mathbf{x},\mathbf{x})=(\mathbf{x}_k+\mathbf{H}_k^T\mathbf{q},\mathbf{x}_k+\mathbf{H}_k^T\mathbf{q})=(\mathbf{x}_k,\mathbf{x}_k)+2(\mathbf{x}_k,\mathbf{H}_k^T\mathbf{q})+(\mathbf{H}_k^T\mathbf{q},\mathbf{H}_k^T\mathbf{q})>0.$$

Felhasználva a 2.1. lemmát és azt a tényt, hogy az $\mathbf{x}_k$ megoldás a $\mathbf{p}_1,\mathbf{p}_2,\ldots,\mathbf{p}_k$ vektorok lineáris kombinációjaként felírható, az $(\mathbf{x}_k,\mathbf{H}_k^T\mathbf{q})$ skalárszorzatra azt kapjuk, hogy

$$\mathbf{x}_k^T\mathbf{H}_k^T\mathbf{q}=\sum_{i=1}^{k}\alpha_i\mathbf{p}_i^T\mathbf{H}_k^T\mathbf{q}=\sum_{i=1}^{k}\alpha_i\mathbf{a}_i^T\mathbf{H}_{i-1}\mathbf{H}_k^T\mathbf{q}=\sum_{i=1}^{k}\alpha_i\mathbf{a}_i^T\mathbf{H}_k^T\mathbf{q}=0.$$

Tehát

$$(\mathbf{x},\mathbf{x})\geqq(\mathbf{x}_k,\mathbf{x}_k)$$

azaz $\mathbf{x}_k$ normál megoldás tehát

$$\mathbf{x}^*\equiv\mathbf{x}_k.$$

vagyis módszerosztályunk *Moore—Penrose*-értelemben vett megoldást szolgáltat.

4. A módszerosztály alkalmazása ritka együtthatós és speciális típusú lineáris egyenletrendszerek megoldására

A k_i, l_i, γ_i paraméterek, *lépésenként tetszőleges* megválasztása várhatóan megfelelő lehetőségeket biztosít ritka együtthatós, illetve bizonyos speciális struktúrájú lineáris egyenletrendszerek megoldására. Ezzel kapcsolatosan néhány észrevételt teszünk.

a) Ha a (2.5) összefüggésben $\gamma_i=0$ minden i-re, azaz

$$\mathbf{H}_{i+1}=\mathbf{H}_i-\frac{\mathbf{H}_i\mathbf{a}_{i+1}\mathbf{y}_{i+1}^T\mathbf{H}_i}{\mathbf{y}_{i+1}^T\mathbf{H}_i\mathbf{a}_{i+1}},$$

akkor a (2.4) kifejezésben szereplő k_i, l_i konstansokat célszerű úgy megválasztani, hogy az $\mathbf{y}_{i+1}$ vektor komponensei minél több helyen zérussá váljanak, mert így a diád képzése során a nem nulla elemek száma lényegesen csökken.

b) Ha a $\gamma_i = -1/p_{i+1}^T \mathbf{a}_{i+1}$ választással élünk minden lépésben, akkor a $\mathbf{H}_i$ mátrixsorozat minden eleme szimmetrikus

$$\mathbf{H}_{i+1} = \mathbf{H}_i - \frac{\mathbf{H}_i \mathbf{a}_{i+1} \mathbf{a}_{i+1}^T \mathbf{H}_i}{\mathbf{a}_{i+1}^T \mathbf{H}_i \mathbf{a}_{i+1}}.$$

(Ez elérhető $\gamma_i=0$, $k_i=0$, $l_i=-1$ választással is.) Ebben az esetben a mátrixok képzéseihez lépésenként $n(n+1)/2$ szorzásra van szükség.

c) A szimmetrikus $\mathbf{H}_i$ mátrixsorozatot tekintve (a gondolatmenet nem szimmetrikus sorozatra is átvihető) és figyelembe véve azt a tényt, hogy $\mathbf{H}_0=\mathbf{I}$, láthatjuk, hogy a $\mathbf{H}_i$ mátrixok telítődése a $\mathbf{H}_{i-1}$ telítettségétől és az $\mathbf{a}_{i+1}$ nem zérus elemeinek számától, valamint a kettő összevetésétől függ. Ez az összefüggés a következő példában világossá válik.

Ismeretes, hogy differenciasémát alkalmazva a

$$y''(x)+P(x)y'(x)+Q(x)y(x) = R(x)$$

másodrendű differenciálegyenletre, az $y(a)=y_0$, $y(b)=y_n$ peremértékek mellett, a feladat visszavezethető egy olyan lineáris egyenletrendszer megoldására, amelynek együttható mátrixa

$$\begin{pmatrix} -2 & 1 & \dots & 0 & 0 \\ 1 & -2 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & -2 & 1 \\ 0 & 0 & \dots & 1 & -2 \end{pmatrix}$$

alakú.

A fenti észrevételünket figyelembe véve, célszerű az egyenleteket a következőképpen átrendezni

$$\begin{pmatrix} -2 & 1 & 0 & \dots & & & & & & \\ 1 & -2 & 1 & 0 & \dots & & & & & \\ 0 & 0 & 0 & 1 & -2 & 1 & 0 & \dots & & \\ 0 & 0 & 0 & 0 & 1 & -2 & 1 & 0 & \dots & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -2 & 1 & 0 & \dots \end{pmatrix}$$

amely mátrix végén szerepelnek azok a sorok, amelyeket az elejéről kivettünk. Ebben az esetben ugyanis a $\mathbf{H}_i$ mátrixok nem fokozatosan telnek meg nem nulla elemekkel (a $\mathbf{D}_i$-k képzésében az $\mathbf{a}_i$-k szerepelnek), és így a kiszámítandó nem nulla elemek száma csökken.

Világos, hogy a fenti három észrevételt egy konkrét feladatnál általában együttesen kell alkalmazni és a megfelelő stratégia kiválasztásánál alapos megfontolásokra van szükség.

IRODALOM

[1] HUANG, H. Y., "A direct method for the general solution of a system of linear equations", *JOTA* **16** (1975) 429—445.

(Beérkezett: 1979. szeptember 28.)

ABAFFY JÓZSEF
MKKE MATEMATIKAI ÉS SZÁMÍTÁSTUDOMÁNYI INTÉZET
1093 BUDAPEST, DIMITROV TÉR 8.

A DIRECT METHOD CLASS FOR THE GENERAL SOLUTION OF SYSTEMS OF LINEAR EQUATIONS

J. ABAFFY

In this paper a class of direct solution methods for the systems of linear equations is given which yields the *Moore—Penrose* solution in case of undetermined systems of equations.

Alkalmazott Matematikai Lapok **5** *(1979)* 241—247

EGY MÓDSZER A HŐÁRAM BECSLÉSÉRE

ECSEDI ISTVÁN
Miskolc

A műszaki gyakorlatban nagy jelentősége van a hővezetés, hőátszármaztatás különböző típusainak. E tanulmány szilárd testekkel kapcsolatos hővezetési feladatot vizsgál. Az egyik közegből a másik közegbe hővezetéssel átszármaztatott hőmennyiség (hőáram) számértékére alsó és felső korlátokat ismertet. A hőáram pontos (szigorú) értéke a *Laplace*-féle parciális differenciálegyenlettel kapcsolatos *harmadik* kerületérték feladat megoldásának ismeretében adható csak meg. A *Laplace*-féle parciális differenciálegyenlet pontos megoldásának előállítása azonban igen gyakran nagy nehézségekbe ütközik, s effektíve sok esetben nem is adható meg.

A tanulmány által bizonyított (2.3), (3.4), (4.3) egyenlőtlenségi relációk alkalmazása nem igényli az (1.1), (1.2), (1.3), (1.4) egyenletek által kijelölt kerületérték feladat megoldását.

0. Fontosabb jelölések

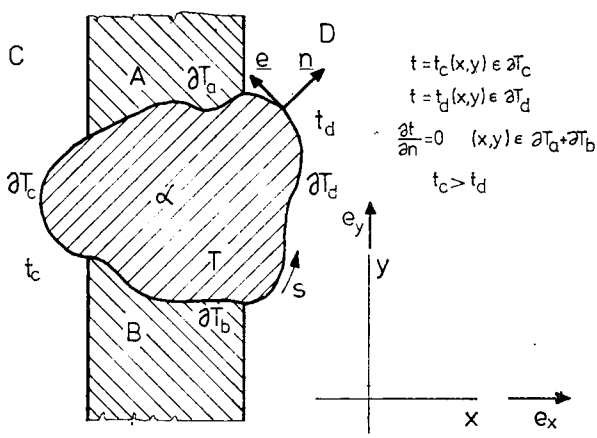
x, y derékszögű koordináták,
$\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z$ egységvektorok,
$t=t(x, y)$ hőmérséklet,
T x, y síkbeli korlátos, egyszeresen összefüggő tartomány,
$\partial T=\partial T_a+\partial T_b+\partial T_c+\partial T_d$ a T tartomány határa,
$\mathbf{n}$ ∂T határgörbe külső normális egységvektora,
s ∂T határgörbén értelmezett ívkoordináta,
Q_c, Q_d, q hőáramok,
λ „belső" hővezetési együttható,
$\nabla=\frac{\partial}{\partial x}\mathbf{e}_x+\frac{\partial}{\partial y}\mathbf{e}_y$ *Hamilton*-féle differenciáloperátor,
„$\times$" vektoriális szorzás jele,
„$\cdot$" skaláris szorzás jele,
$\Delta=\nabla\cdot\nabla=\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}$ *Laplace*-féle differenciáloperátor,
$\mathbf{e}=\mathbf{e}_z\times\mathbf{n}$ a ∂T határgörbe érintő egységvektora,
$\mathbf{e}_z=\mathbf{e}_x\times\mathbf{e}_y$.
Egyéb mennyiségeket, változókat a szöveg értelmezi.

1. Bevezetés

E tanulmány eredményei síkbeli stacionárius hővezetési problémákkal kapcsolatosak. A tanulmány által vizsgált modellt az 1. ábra szemlélteti. A C közegből a ∂T_c peremgörbén időegység alatt Q_c hőmennyiség áramlik az α szilárd testbe. Az α szilárd testből a ∂T_d peremgörbeszakaszon Q_d hőmennyiség távozik a D közegbe. Az A és B közegek tökéletesen hőszigetelők, így az α testből a $\partial T_a+\partial T_b$ peremgörbe szakaszon hő nem távozik, illetve nem lép be az α testbe.

A fentiekben vázolt modellben az α test szerepe a C közegből a D közegbe történő hőátszármaztatás (hőelvezetés).

A C közeg hőmérséklete t_c=állandó, a D közeg hőmérséklete t_d=állandó. Feltételezés szerint legyen $t_c > t_d$.



1. ábra
Egy síkbeli hővezetési modell

A hővezetés *Fourier*-féle elmélete szerint az α test $t=t(x, y)[(x, y)\in T]$ hőmérséklet mezejét a következő kerületérték feladattal hozhatjuk kapcsolatba ([1], [2]):

$$\Delta t = 0, \quad (x, y)\in T, \tag{1.1}$$

$$t = t_c, \quad (x, y)\in\partial T_c, \tag{1.2}$$

$$t = t_d, \quad (x, y)\in\partial T_d, \tag{1.3}$$

$$\frac{\partial t}{\partial n} = 0, \quad (x, y)\in\partial T_a+\partial T_b. \tag{1.4}$$

A C közegből az α testbe időegység alatt

$$Q_c = \lambda \int_{\partial T_c} \frac{\partial t}{\partial n}\, ds \tag{1.5}$$

hőmennyiség lép be. Az α testből a D közegbe pedig időegység alatt

$$Q_d = \lambda \int_{\partial T_d} \frac{\partial t}{\partial n} ds \tag{1.6}$$

hőmennyiség távozik. Könnyen kimutatható, hogy

$$Q_c + Q_d = 0, \tag{1.7}$$

$$q = Q_c > 0, \tag{1.8}$$

Az (1.1), (1.4) egyenletek alapján írhatjuk, hogy

$$\lambda \int_T \Delta t \, dT = \lambda \int_{\partial T} \frac{\partial t}{\partial n} ds = 0, \tag{1.9}$$

azaz

$$\lambda \int_{\partial T_c} \frac{\partial t}{\partial n} ds + \lambda \int_{\partial T_d} \frac{\partial t}{\partial n} ds = Q_c + Q_d = 0. \tag{1.10}$$

Másrészt viszont a

$$\lambda \int_{\partial T} t \frac{\partial t}{\partial n} ds = \lambda \int_T (\nabla t)^2 \, dT + \lambda \int_T t \, \Delta t \, dT \tag{1.11}$$

identitás alkalmazásával kapjuk, hogy

$$(t_c - t_d) q = \lambda \int_T (\nabla t)^2 \, dT. \tag{1.12}$$

Az (1.12) egyenletből már következik, hogy $t_c - t_d > 0$ esetén szükségképpen $q > 0$.

A q hőáramot az (1.1), (1.2), (1.3), (1.4) egyenletek által kijelölt kerületérték feladat $t = t(x, y)$ megoldásának az ismeretében a

$$q = \frac{\lambda}{t_c - t_d} \int_T (\nabla t)^2 \, dT \tag{1.13}$$

formula alapján tudjuk meghatározni.

E tanulmány tárgya olyan egyenlőtlenségi relációk levezetése, amelyek segítségével alsó- és felső korlátok képezhetők a q hőáram számára.

2. Felső korlát

Legyen $a = a(x, y)$ a $T + \partial T$ zárt tartományban folytonos a T tartományban folytonosan differenciálható olyan kétváltozós függvény, amely a ∂T_c és ∂T_d peremgörbe szakaszokon állandó értékű:

$$a(x, y) = A_c = \text{állandó} \quad (x, y) \in \partial T_c, \tag{2.1}$$

$$a(x, y) = A_d = \text{állandó} \quad (x, y) \in \partial T_d, \tag{2.2}$$

$$A_c \neq A_d.$$

Fennáll a

$$q \leqq \lambda \frac{t_c - t_d}{(A_c - A_d)^2} \int_T (\nabla a)^2 \, dT \tag{2.3}$$

egyenlőtlenségi reláció.

Bizonyítás: A *Schwarz*-egyenlőtlenség alapján írható, hogy

$$\int_T (\nabla t)^2 \, dT \int_T (\nabla a)^2 \, dT \geqq \Big(\int_T \nabla a \cdot \nabla t \, dT \Big)^2. \tag{2.4}$$

A szorzat függvény deriválási szabályának és a *Gauss*-féle integrálátalakítás tételének az együttes alkalmazásával belátható, hogy

$$\int_T \nabla t \cdot \nabla a \, dT = \int_T ((\nabla t) a) \cdot \nabla \, dT - \int_T a \Delta t \, dT = \tag{2.5}$$

$$= \int_{\partial T} \frac{\partial t}{\partial n} a \, ds = \frac{q}{\lambda} (A_c - A_d).$$

A (2.4) és (2.5) formulák kombinálásával a bizonyítandó (2.3) egyenlőtlenségi relációt nyerjük, ha tekintettel vagyunk a (1.13) formulára.

3. Egy Lemma

Lemma. Legyen $\mathbf{b} = b_x(x, y)\mathbf{e}_x + b_y(x, y)\mathbf{e}_y$ olyan nem azonosan zérus síkbeli vektormező, amely az alábbi feltételeknek tesz eleget:

$$\nabla \cdot \mathbf{b} = 0, \quad (x, y) \in T, \tag{3.1}$$

$$\mathbf{n} \cdot \mathbf{b} = 0, \quad (x, y) \in \partial T_a + \partial T_b. \tag{3.2}$$

Legyen továbbá

$$R = \int_{\partial T_c} \mathbf{b} \cdot \mathbf{n} ds. \tag{3.3}$$

Fennáll a

$$q \geqq \lambda (t_c - t_d) \frac{R^2}{\int_T \mathbf{b}^2 \, dT} \tag{3.4}$$

egyenlőtlenségi reláció.

Bizonyítás: A *Schwarz*-egyenlőtlenség alapján írható, hogy

$$\int_T (\nabla t)^2 \, dT \int_T \mathbf{b}^2 \, dT \geqq \Big(\int_T \mathbf{b} \cdot \nabla t \, dT \Big)^2. \tag{3.5}$$

A (3.5) egyenlőtlenség jobb oldalát átalakítjuk:

$$\int_T \mathbf{b} \cdot \nabla t \, dT = \int_T (\mathbf{b} t) \nabla \, dT - \int_T t (\nabla \cdot \mathbf{b}) \, dT = \tag{3.6}$$

$$= \int_{\partial T} (\mathbf{b} t) \cdot \mathbf{n} \, ds = t_c \int_{\partial T_c} \mathbf{b} \cdot \mathbf{n} \, ds + t_d \int_{\partial T_d} \mathbf{b} \cdot \mathbf{n} \, ds.$$

A (3.1), (3.2) egyenletekből a *Gauss*-féle integrál átalakítási tétel felhasználásával kapjuk a (3.7) egyenletet:

$$\int_T \nabla \cdot \mathbf{b}\, dT = \int_{\partial T} \mathbf{b} \cdot \mathbf{n}\, ds = \int_{\partial T_c} \mathbf{b} \cdot \mathbf{n}\, ds + \int_{\partial T_d} \mathbf{b} \cdot \mathbf{n}\, ds = 0. \tag{3.7}$$

A (3.3), (3.6), (3.7) egyenletek és a (3.5) egyenlőtlenség kombinálásával a bizonyítandó (3.4) egyenlőtlenségi relációt nyerjük.

4. Alsó korlát

Legyen $c=c(x, y)$ a $T+\partial T$ zárt tartományban folytonos a T tartományban differenciálható nem azonosan állandó olyan kétváltozós függvény, amely a ∂T_a és ∂T_b peremgörbéken állandó értékű:

$$c(x, y) = C_A, \quad (x, y) \in \partial T_a, \tag{4.1}$$

$$c(x, y) = C_B, \quad (x, y) \in \partial T_b. \tag{4.2}$$

Fennáll a

$$q \geqq \lambda (t_c - t_d) \frac{(C_A - C_B)^2}{\int_T (\nabla c)^2\, dT} \tag{4.3}$$

egyenlőtlenségi reláció.

Bizonyítás: Legyen

$$\mathbf{b} = \nabla c \times \mathbf{e}_z, \quad (\mathbf{e}_z = \mathbf{e}_x \times \mathbf{e}_y). \tag{4.4}$$

A (4.4) alakú $\mathbf{b}=\mathbf{b}(x, y)$ vektor tetszőleges differenciálható $c=c(x, y)$ kétváltozós skalár függvényt felvéve kielégíti a (3.1) differenciálegyenletet.

A (3.2) egyenlet szerint

$$\mathbf{n} \cdot \mathbf{b} = (\nabla c \times \mathbf{e}_z) \cdot \mathbf{n} = \nabla c \cdot (\mathbf{e}_z \times \mathbf{n}) = \tag{4.5}$$

$$= \nabla c \cdot \mathbf{e} = \frac{\partial c}{\partial s} = 0, \quad (x, y) \in \partial T_a + \partial T_b.$$

A (4.5) formulában $\mathbf{e}=\mathbf{e}_z \times \mathbf{n}$ a $\partial T_a+\partial T_b$ görbe érintő egységvektora (1. ábra). A (4.5) feltétel szerint a $c=c(x, y)$ függvény állandó értékű a ∂T_a és ∂T_b peremgörbe szakaszokon:

$$c(x, y) = C_A, \quad (x, y) \in \partial T_a, \tag{4.6}$$

$$c(x, y) = C_B, \quad (x, y) \in \partial T_b. \tag{4.7}$$

Az R mennyiségre az

$$R = \int_{\partial T_c} \mathbf{b} \cdot n\, ds = \int_{\partial T_c} \frac{\partial c}{\partial s}\, ds = C_A - C_B \tag{4.8}$$

eredményt tudjuk levezetni.

A (4.4) és (4.8) egyenleteknek a (3.4) egyenlőtlenségi relációba való helyettesítésével a bizonyítandó (4.3) egyenlőtlenséget nyerjük.

5. Megjegyzések az egyenlőtlenségi relációkhoz

Rövid diszkusszióval kimutatható, hogy a (2.3) relációban az egyenlőség jele csak akkor érvényes, ha az $a=a(x, y)$ függvény az alábbi peremérték feladat megoldása:

$$\Delta a = 0, \quad (x, y)\in T, \tag{5.1}$$

$$a = A_c = \text{állandó}, \quad (x, y)\in \partial T_c, \tag{5.2}$$

$$a = A_d = \text{állandó}, \quad (x, y)\in \partial T_d, \tag{5.3}$$

$$\frac{\partial a}{\partial n} = 0, \quad (x, y)\in \partial T_a + \partial T_b, \tag{5.4}$$

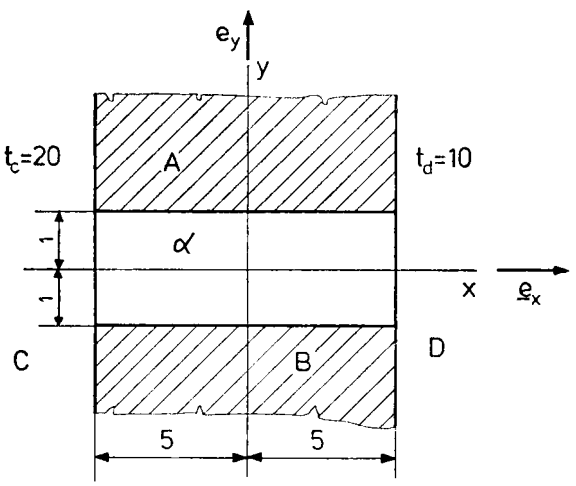
$$k(t_c - t_d) = A_c - A_d, \tag{5.5}$$

$$k = \text{állandó}, \quad k \neq 0.$$

A (4.3) egyenlőtlenségi relációban az egyenlőség jele pedig akkor érvényes, ha $c=c(x, y)$ a (5.1)—(5.5) egyenletek által meghatározott $a=a(x, y)$ harmonikus függvények egyikének a harmonikus társa.

6. Példa

Végezetül egy példán szemléltetjük a (2.3), (4.3) egyenlőtlenségi relációk alkalmazását. A számpéldákon a mértékegységek feltüntetésétől eltekintünk. A példa téglalap alakú T tartományra vonatkozik (2. ábra).



2. ábra
Hővezetés téglalap alakú tartományban

A következő adatokkal számolunk:

$$t_c = 20, \quad t_d = 10, \quad \lambda = 1.$$

A (2.3) egyenlőtlenségi relációba az $a=a(x, y)=(15-x)+1{,}2\,(x^2-25)y^2$ függvényt helyettesítve, a

$$q \leqq 2{,}04$$

felső korlátot kapjuk a hőáram számára.

A (4.3) egyenlőtlenségi relációban pedig a $c=c(x, y)=x(y^2-1)+1{,}4y$ alakú függvénnyel számolva, a

$$q \geqq 1{,}75$$

alsó korlátot nyerjük a hőáram számára.

IRODALOM

[1] CARSLAW, H. S. and JEAGER, J. C., *Heat Conduction in Solids* (2nd. ed., Oxford Univ. Press, London and New York, 1959).

[2] FRANK—MISES, *Die Differential und Integralgleichungen der Mechanik und Physik* (Braunschweig, 1930, Band II).

(Beérkezett: 1979. október 15.)

ECSEDI ISTVÁN
NEHÉZIPARI MŰSZAKI EGYETEM MECHANIKAI TANSZÉK
3515 MISKOLC, EGYETEMVÁROS

A METHOD TO ESTIMATION OF HEAT RATE

I. ECSEDI

The purpose of this paper is to establish upper and lower bounds for the heat rate. The author uses the *Schwarz* inequality to obtain bounds for the heat rate.

Alkalmazott Matematikai Lapok **5** *(1979)* 249—260

AZ ALAPANYAGOK KEVERÉSI ARÁNYÁNAK ÉS A TÁROLÓK NAGYSÁGÁNAK OPTIMALIZÁLÁSA ASZFALTKEVERŐ BERENDEZÉSEKRE

KELLE PÉTER
Budapest

A változó összetételű alapanyagok keverési arányának optimumát kvadratikus programozási algoritmussal határozzuk meg. A négy gyártásközi tároló együttes készletalakulására egy sztochasztikus modellt adunk, melyben a beérkezési folyamatokat *Wiener-folyamatok* közelítik, sztochasztikusan összefüggő paraméterekkel, a felhasználás pedig egyenletes. Adott hosszúságú időszakra a tárolók kiürülésének, illetve túlfolyásának a valószínűsége egy előírt szintet nem haladhat meg. A folyamatos anyagellátáshoz szükséges induló készletszint és az optimális tárolóméretek kialakításának kérdését megbízhatósági modellel írjuk le. A modell matematikai tulajdonságait, megoldásának eredményét ismertetjük és utalunk a megbízhatósági készletmodellekkel való kapcsolatra.

1. Bevezetés

Automatikus aszfaltkeverő berendezések anyagellátási problémáit vizsgáljuk. A változó összetételű alapanyagok optimális keverési arányát kvadratikus programozási algoritmussal határozzuk meg. A berendezések gyártásközi tárolóinak készletalakulását egy sztochasztikus modellel írjuk le. A fő feladat a folyamatos gyártáshoz szükséges induló készletszint és a tárolók nagyságának tervezése. A követelmény, hogy adott hosszúságú időszakra a kiürülés és a túlfolyás valószínűsége egy előírt szintet nem haladhat meg. A feladat megoldására adott, valószínűséggel korlátozott készletmodell alkalmas lehet más területeken is a biztonsági készletszint és az optimális raktárkapacitás meghatározására, ahol folyamatos, véletlen jellegű anyagbeérkezés történik. A számítógépes realizációt és a gyakorlati feladat numerikus eredményeit röviden ismertetjük. A feladat megfogalmazásában és a műszaki jellegű kérdésekben nyújtott segítségért köszönetet mondok DR. BODNÁR GÉZA főmérnöknek, az *Útépítő Tröszt* osztályvezetőjének.

2. Aszfaltkeverési arányok optimalizálása

Az *Útépítő Tröszt* telepein többféle aszfaltkeveréket állítanak elő nagy teljesítményű automata keverő berendezéseken. Az alapanyagként felhasznált kőzúzalékokat a különböző átmérőjű szemcsék súlyarányai jellemzik, melyet szemeloszlásnak neveznek. A k-adik alapanyag típusra legyen α_{ik} $(i=1, \ldots, m;\ k=1, \ldots, n)$ a ϱ_i átmérőnél kisebb szemcsék súlyaránya. A szemeloszlást az $\boldsymbol{\alpha}_k'=(\alpha_{1k}, \ldots, \alpha_{mk})$ vektor jellemzi, melyet az alapanyagok beérkezésekor időnként mérnek és regisztrálnak. A gyártás közben felhasznált alapanyagok szemeloszlása véletlen jelleggel változik, ezért az $\boldsymbol{\alpha}_k$ vektorokat valószínűségi változóknak tekintjük, melyek várható értékeit $\mathbf{a}_k'=(a_{1k}, \ldots, a_{mk})$ jelöli $k=1, \ldots, n$ esetén. A különböző típusú

alapanyagok egymástól függetlenül érkeznek, így az α_k valószínűségi vektorváltozók is sztochasztikusan függetlenek a különböző k értékekre.

A gyártott aszfalt az alapanyagok $x_1, \ldots, x_n$ arányú keveréke, adalékanyagok hozzáadásával. A keverék $\eta = \sum x_k \alpha_k$ szemeloszlásának szabvány által rögzített $\mathbf{u}' = (u_1, \ldots, u_m)$ és $\mathbf{v}' = (v_1, \ldots, v_m)$ alsó, illetve felső korlátok között kell lenni. Ezek a korlátok az aszfalt típusától függenek. Mivel α_k valószínűségi változó $k = 1, \ldots, n$ esetén, a korlátozó feltételek teljesülését csak bizonyos p valószínűséggel kívánhatjuk meg. Így az optimális keverési arány meghatározására a következő sztochasztikus lineáris programozási feladatot fogalmazhatjuk meg:

$$\min \mathbf{c}'\mathbf{x}$$

feltéve, hogy

$$P\left(\mathbf{u} \leqq \sum_k x_k \alpha_k \leqq \mathbf{v}\right) \geqq p, \tag{2.1}$$

$$\sum_k x_k = 1, \quad x_k \geqq 0, \quad k = 1, \ldots, n,$$

ahol $\mathbf{c}' = (c_1, \ldots, c_n)$ jelöli pl. az alapanyagok beszerzési árát.

A (2.1) feladat megoldása a sztochasztikus együttható mátrix miatt komoly nehézséget jelent. Az irodalomban csak speciális struktúrájú véletlen együttható mátrix esetén ismertek hatásos megoldási algoritmusok a (2.1) típusú feladatra. Elsősorban PRÉKOPA A. [7] cikkében közölt eredmények nyújtanak erre lehetőséget, ha az α_k eloszlása m dimenziós normális eloszlás, melynek $\mathbf{C}_k$ kovarianciamátrixa $\mathbf{C}_k = d_k \mathbf{C}$ alakú, ahol d_k konstans ($k = 1, \ldots, n$). Ebben az esetben a (2.1) feladat megoldása egy konvex programozási feladatra vezethető vissza [7] szerint. A vizsgált gyakorlati feladatra a fenti feltételek általában nem teljesülnek.

A felhasznált alapanyagok ára nem sokban különbözik, a szokásos költségminimalizálás helyett sokkal fontosabb cél a szabvány betartásának biztosítása. Ez a szemeloszlások nagy ingadozása miatt komoly gyakorlati nehézséget okoz. A (2.1) feladat átfogalmazható úgy, hogy a $\mathbf{c}'\mathbf{x}$ célfüggvény minimalizálása helyett a p értéket maximalizáljuk a súlyponti célkitűzésnek megfelelően. Ebben az esetben sem tudunk azonban az α_k valószínűségi vektorok tulajdonságai miatt hatásos megoldási algoritmust adni.

A sztochasztikus modell megoldási nehézségei miatt a következő determinisztikus feladat megoldását keressük

$$\min \sum_i \vartheta_i \left(\sum_k a_{ik} x_k - \frac{u_i + v_i}{2} \right)^2$$

feltéve, hogy

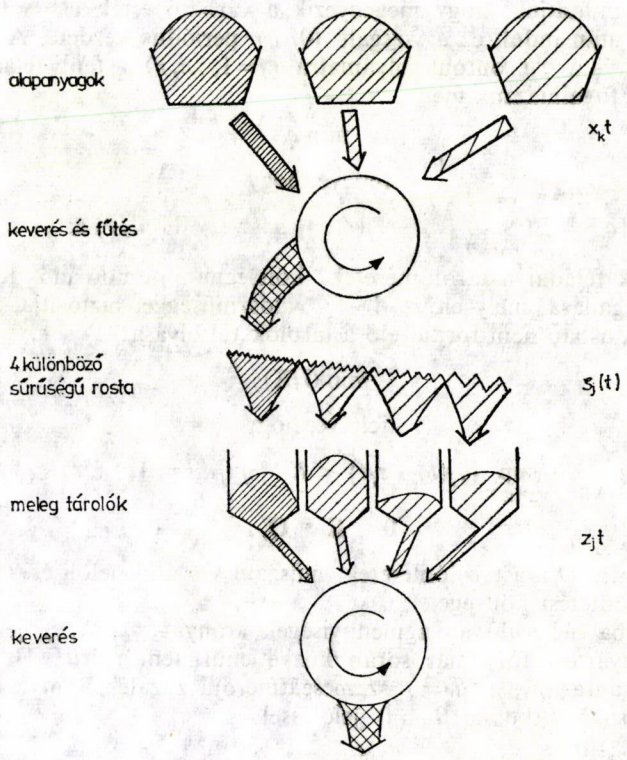
$$\mathbf{u} \leqq \sum_k x_k \mathbf{a}_k \leqq \mathbf{v}, \tag{2.2}$$

$$\sum_k x_k = 1, \quad x_k \geqq 0, \quad k = 1, \ldots, n; \quad i = 1, \ldots, m.$$

Itt a szemeloszlási szabvány középvonalától való súlyozott négyzetes eltérést minimalizáljuk. A lineáris feltételrendszer a keverék szemeloszlásának várható értékére vonatkozik, melyre a szabvány teljesülését feltétlenül megköveteljük. A súlyfaktorokat a $\vartheta_i = \sum_k D^2[\alpha_{ik}]$ ($i = 1, \ldots, m$) alakban választjuk, amit az α_k vektorok függetlensége indokol. A (2.2) kvadratikus programozási feladat megoldásáról az 5. pontban írunk.

3. A készletezési feladat és modellje

Először röviden ismertetjük az aszfaltkeverő berendezés gyártási folyamatának azt a részét, melyet egy sztochasztikus modellel írunk le. Az $x_1, \ldots, x_n$ arányban összekevert kőzúzalékokat r különböző sűrűségű rostán szétválasztják. A $\tau_{j-1} \leqq \leqq d < \tau_j$ szemcseátmérőjű zúzalék a j-edik melegített tárolóba kerül, ez a $\zeta_j(t)$ beérkezési folyamat $j=1, \ldots, r$-re, mely folyamatos és sztochasztikus a kőzúzalékokban levő szemcseátmérők súlyarányának véletlen jellegű változásai miatt. A tárolók output folyamata gyakorlatilag folytonos és egyenletes, rögzített $z_1, \ldots, z_r$ intenzitással. A készletezési feladat megértéséhez segítséget nyújthat a mellékelt 1. ábra, mely az anyagáramlást vázolja.



1. ábra
Az aszfaltkeverés technológiai folyamatának vizsgált részlete és az anyagáramlás folyamata

Az $x_1, \ldots, x_n$ keverési arányok és a tárolók $z_1, \ldots, z_r$ output intenzitásának rögzítése után a folyamat automatikusan zajlik mindaddig, míg a tárolók valamelyike kiürül, vagy megtelik és túlfolyik. Mindkét esetben le kell állítani a gyártási folyamatot, megvárni a feltöltődést vagy az x_i keverési arányokat kell megváltoztatni egy időre. Ez megzavarja a termelést (különösen a túlfolyás), így csak ritkán szabad előfordulnia. A tárolók kapacitását eredetileg a szabványnak megfelelő

kőzúzalékokra tervezték, de a beérkező alapanyagok szemloszlása gyakran lényegesen eltér ettől és így a fenti zavarok sűrűn előfordulnak. Ennek következtében szükséges a tárolók megnagyobbítása, bár ez a nagyságtól függően jelentős többletköltséget okoz. Másrészt a keverési eljárás nem kezdhető üres tárolókkal. Meg kell határozni az indulásnál szükséges minimális q feltöltési időt, mely a későbbi folyamatos ellátást is biztosítja.

A tárolók kapacitásának megnövelésére és a fenti q időtartam meghatározására valószínűséggel korlátozott készletmodellt dolgoztunk ki, mely együttes valószínűségi korlátokat tekint, a tárolók sztochasztikusan összefüggő input folyamata mellett.

Először azt a minimális q feltöltési időt keressük, mely előírt $1-\varepsilon_1$ valószínűséggel biztosítja a folyamatos anyagellátást egy adott T hosszúságú időtartamra. A q időtartam alatt beérkező $\zeta_j(q)$ mennyiséget a $z_j q$ értékkel közelítjük, melyről később megmutatjuk, hogy megegyezik a várható értékkel. A termelés a feltöltési időszak után indul, ez a vizsgált $(0, T)$ periódus kezdete. A $\zeta_j(t)$ beérkezési folyamatra és a $z_j t$ output folyamatra $(j=1, \ldots, r)$ a fenti feladatot a következő formában fogalmazzuk meg:

$$\min q$$

feltéve, hogy

$$g(q) = P\Big(\sup_{0 \leqq t \leqq T} \{z_j t - \zeta_j(t)\} < q z_j,\ j = 1, \ldots, r\Big) \geqq 1-\varepsilon_1. \tag{3.1}$$

A második feladat a tárolóméretek költségminimumot adó $\mathbf{K}'=(K_1, \ldots, K_r)$ vektorának megadása, mely előírt $1-\varepsilon_2$ valószínűséggel biztosítja, hogy a $(0, T)$ termelési periódusban nem fordul elő a tárolók túlfolyása:

$$\min G(\mathbf{K})$$

feltéve, hogy

$$H(\mathbf{K}) = P\Big(\sup_{0 \leqq t \leqq T} \{\zeta_j(t) - z_j t\} < K_j - q z_j,\ j = 1, \ldots, r\Big) \geqq 1-\varepsilon_2, \tag{3.2}$$

$$\mathbf{0} \leqq \mathbf{K} \leqq \mathbf{Q},$$

ahol $\mathbf{Q}'=(Q_1, \ldots, Q_r)$ a tárolóméretek műszaki korlátait jelöli és $G(\mathbf{K})$ a tárolók építési és működtetési költsége.

A tárolókba bekerülő anyagmennyiségek arányát $\gamma_1, \ldots, \gamma_r$ jelöli. Ezeket az arányokat a gyártási folyamat során közvetlenül nem mérik. Ha a felhasznált k-adik alapanyagra β_{jk} a $d<\tau_j$ szemcseátmérőjű zúzalék aránya (mely véletlenszerűen ingadozik), akkor a $\beta_{0k}=0$ jelöléssel

$$\gamma_j = \sum_{k=1}^{n} x_k(\beta_{jk} - \beta_{j-1,k}), \tag{3.3}$$

mely valószínűségi változó $j=1, \ldots, r$ esetén. A β_{jk} arányok realizációit sem mérik, de a minőségellenőrzés során megmérik és nyilvántartják az alapanyagoknak a ϱ_i átmérőjű rostákon áthulló α_{ik} súlyarányait $(i=1, \ldots, m; k=1, \ldots, n)$. A ϱ_i értékek különböznek a τ_j $(j=1, \ldots, r)$ értékektől. A rendelkezésre álló adatok is alátámasztják, hogy a kőzúzalékok szemeloszlásának súlyarányait a lognormális eloszlásfüggvény görbéje jól közelíti, ennek segítségével interpoláljuk β_{jk} realizációit.

A (3.3) összefüggés segítségével meghatározható az

$$F(\mathbf{y}) = P(\gamma_j < y_j,\ j = 1, \ldots, r) \tag{3.4}$$

r dimenziós együttes eloszlásfüggvény. Az output intenzitások értékét, melyet tetszés szerint rögzíthetünk, a

$$z_j = E[\gamma_j],\quad j = 1, \ldots, r \tag{3.5}$$

várható értékek alapján választjuk. Ez biztosítja a szabvány betartását megfelelő $x_1, \ldots, x_n$ keverési arányok esetén, továbbá azt, hogy a tárolókba bekerülő és onnan kivett anyagmennyiség várható értéke megegyezik.

Feltételezhetjük, hogy

$$E[\zeta_j(t)] = y_j t,\quad j = 1, \ldots, r \tag{3.6}$$

és

$$D^2[\zeta_j(t)] = \sigma_j^2 t,\quad j = 1, \ldots, r, \tag{3.7}$$

ahol y_j és σ_j konstans abban a rövid $(0, T)$ periódusban, melyre a folyamatos ellátást megkívánjuk. Itt $\mathbf{y}' = (y_1, \ldots, y_r)$ a $\gamma' = (\gamma_1, \ldots, \gamma_r)$ valószínűségi vektorváltozónak a realizációja. A $\zeta_j(t)$ folyamatnak olyan modelljét adjuk, mely gyakorlati szempontból megfelelő közelítést ad, és lehetővé teszi a (3.1) és (3.2) feladatok numerikus megoldását. A $\zeta_j(t)$ sztochasztikus folyamatról feltételezzük, hogy a $\gamma = \mathbf{y}$ feltétel mellett *homogén Wiener-folyamat* az y_j és σ_j paraméterekkel $j = 1, \ldots, r$ esetén, azaz szeparábilis sztochasztikus folyamat, melyre

$$P(\zeta_j(t) < u) = \Phi\left(\frac{u - y_j t}{\sigma_j \sqrt{t}}\right), \tag{3.8}$$

ahol $\Phi(u)$ a standard normális eloszlásfüggvényt jelöli.

4. A készletmodell matematikai tulajdonságai és megoldása

A $\zeta_j(t)$ folyamatra az előző pontban leírt feltételek lehetővé teszik G. BAXTER és M. DONSKER [1] következő tételének felhasználását:

4.1. Tétel: Ha $\xi(t)$ m és s paraméterű *szeparábilis homogén Wiener-folyamat*, akkor

$$P\left(\sup_{0 \leq t \leq T} \xi(t) < u\right) = P(\xi(T) < u) - e^{\frac{2mu}{s^2}} P(\xi(T) < -u). \tag{4.1}$$

A (3.1) feladat feltételi függvényében levő $\xi(t) = z_j t - \zeta_j(t)$ sztochasztikus folyamat teljesíti a 4.1. tétel feltételeit rögzített $\gamma_j = y_j$ esetén és paraméterei $m = z_j - y_j$ és $s = \sigma_j$, így

$$g_j(q|y_j) = P\left(\sup_{0 \leq t \leq T} \{z_j t - \zeta_j(t)\} < q z_j | \gamma_j = y_j\right) = \tag{4.2}$$

$$= \Phi\left(\frac{q z_j - (z_j - y_j)T}{\sigma_j \sqrt{T}}\right) - e^{\frac{2q z_j (z_j - y_j)}{\sigma_j^2}} \Phi\left(\frac{-q z_j - (z_j - y_j)T}{\sigma_j \sqrt{T}}\right).$$

Feltételezhetjük, hogy rögzített γ esetén a $\zeta_j(t)$ sztochasztikus folyamatok függetlenek, így

$$g(q) = \int_{R^r} \left[\prod_{j=1}^{r} g_j(q|y_j) \right] dF(\mathbf{y}). \tag{4.3}$$

A (4.3) függvény a q változó monoton növekvő függvénye, ezért a $g(q)=1-\varepsilon_1$ egyenletnek pontosan egy megoldása van $(0<\varepsilon_1<1)$ és ez a (3.1) feladat optimális megoldását adja.

A (3.2) feladat sztochasztikus feltételében szereplő

$$H(\mathbf{K}) = P\left(\sup_{0 \leqq t \leqq T} \{\zeta_j(t) - z_j t\} < K_j - q z_j, \quad j = 1, \ldots, r \right) \tag{4.4}$$

függvény a feladat megoldását a (3.1) feladat megoldásánál nehezebbé teszi. Ha a γ valószínűségi vektornak logaritmikusan konkáv $f(\mathbf{y})$ együttes sűrűségfüggvénye van, akkor $H(\mathbf{K})$ logaritmikus konkávitását is tudjuk igazolni. Az algoritmikus megoldást ez jelentősen megkönnyíti, hiszen ekkor a $\{\mathbf{K} | H(\mathbf{K}) \geqq 1-\varepsilon_1\}$ halmaz konvex rögzített $0<\varepsilon_1<1$ esetén. Felhasználjuk PRÉKOPA A. [6] következő tételét:

4.2. Tétel: Ha $f(\mathbf{x}, \mathbf{y})$ az $\mathbf{x}$ n komponensű és az $\mathbf{y}$ m komponensű vektor függvénye logaritmikusan konkáv R^{n+m}-ben, akkor az $\mathbf{x}$ változó

$$\int_{R^m} f(\mathbf{x}, \mathbf{y})\, d\mathbf{y}$$

függvénye logaritmikusan konkáv R^n-ben.

A 4.2. tételen alapul a $H(\mathbf{K})$ függvényre vonatkozó következő tétel:

4.3. Tétel: Legyen $\zeta_1(t), \ldots, \zeta_r(t)$ *szeparábilis homogén Wiener-folyamat* rögzített γ esetén. Ha γ valószínűségi vektorváltozó folytonos eloszlású és $f(\mathbf{y})$ együttes sűrűségfüggvénye logaritmikusan konkáv, továbbá q és $z_1, \ldots, z_r$ konstans, akkor

$$H(\mathbf{K}) = P\left(\sup_{0 \leqq t \leqq T} \{\zeta_j(t) - z_j t\} < K_j - q z_j, \; j = 1, \ldots, r \right)$$

a $\mathbf{K}' = (K_1, \ldots, K_r)$ logaritmikusan konkáv függvénye.

Bizonyítás: A 4.2. tételből következik, hogy rögzített t esetén $(0 \leqq t \leqq T)$ a

$$d(t, \mathbf{K}) = P\left(\zeta_j(t) < K_j - (q-t) z_j, \; j = 1, \ldots, r\right) = \tag{4.5}$$

$$= \int_{R^r} \left[\prod_{j=1}^{r} \Phi\left(\frac{K_j - (q-t) z_j - y_j t}{\sigma_j \sqrt{t}} \right) \right] f(\mathbf{y})\, d\mathbf{y}$$

függvény logaritmikusan konkáv a $\mathbf{K}$ vektorváltozó szerint. Ebből következik, hogy a $(0, T)$ intervallum minden $0<t_1<\ldots<t_N<T$ felosztására a

$$P\left(\zeta_j(t_i) - z_j t_i < K_j - q z_j, \; j = 1, \ldots, r, \; i = 1, \ldots, N\right) = \prod_{i=1}^{N} d(t_i, \mathbf{K}) \tag{4.6}$$

függvény szintén logaritmikusan konkáv. Tekintsük a $0<t_1^{(k)}<\ldots<t_{N_k}^{(k)}<T$ felosztásoknak egy olyan sorozatát, melyre

$$\lim_{k \to \infty} \max_i \{t_{i+1}^{(k)} - t_i^{(k)}\} = 0.$$

Ha $h_k(\mathbf{K})$ jelöli a (4.6) valószínűséget rögzített k esetén, akkor az minden k-ra logaritmikusan konkáv függvény. Ez a tulajdonság nem változik, ha a k szerinti határértéket vesszük, így

$$H(\mathbf{K}) = \lim_{k\to\infty} h_k(\mathbf{K})$$

szintén a $\mathbf{K}$ logaritmikusan konkáv függvénye, ami a 4.3. tétel állítása.

A gyakorlati feladat megoldásánál szétválasztjuk a (3.2) feladat együttes valószínűségi korlátját és a marginális eloszlásokra tett következő feltételeket tekintjük, $j=1, \ldots, r$ esetére:

$$b_j(K_j) = P\left(\sup_{0\leq t\leq T} \{\zeta_j(t) - z_j t\} < K_j - qz_j\right) \geqq 1-\delta_j, \tag{4.7}$$

ahol δ_j rögzített konstans $(0<\delta_j<1)$. A (4.1) összefüggés felhasználásával

$$b_j(K_j) = \int_0^\infty \left[\Phi\left(\frac{K_j - qz_j + (z_j - y_j)T}{\sigma_j\sqrt{T}}\right) - e^{\frac{2(y_j-z_j)(K_j-qz_j)}{\sigma_j^2}} \Phi\left(\frac{-K_j + qz_j + (z_j-y_j)T}{\sigma_j\sqrt{T}}\right)\right] dF_j(y_j), \tag{4.8}$$

ahol $F_j(y_j)$ $(j=1, \ldots, r)$ jelöli a γ valószínűségi vektorváltozó marginális eloszlásait és q értéke rögzített, ezt a (3.1) feladat megoldása adja. A (4.8) függvény monoton növekvő, ezért a (4.7) feltétel a K_j változóra egy egyszerű $L_j(\delta_j)$ alsó korlátot ad, így a $G(\mathbf{K})$ költségfüggvényt a

$$Q_j \geqq K_j \geqq L_j(\delta_j), \quad j = 1, \ldots, r$$

egyenlőtlenségek által definiált tartományon kell minimalizálni.

Több termékes készletmodellt, mely együttes valószínűségi korlátot vizsgál, először PRÉKOPA A. közölt [5] dolgozatában. A termékek (alapanyagok) beérkezési folyamatát véletlen jellegű diszkrét folyamattal modellezi, termékenként különböző paraméterekkel, de ezek függetlenségét feltételezve. A modell megoldási módszere PRÉKOPA A.—KELLE P. [8] dolgozatában szerepel. Esetünkben a beérkezések folyamatai egymással sztochasztikusan összefüggenek és folytonos sztochasztikus folyamatok. NÉMETH GY. [4] dolgozatában a folytonos beérkezést *homogén Wiener-folyamattal* modellezi, ugyanakkor a felhasználást is *Wiener-folyamatnak* tételezi fel. Eredménye azonban csak azon erős megszorítás mellett alkalmazható, ha a két folyamat intenzitása megyegyezik (esetünkben az $y_j=z_j$ feltétel mellett). Dolgozatunkban ezt a feltételt elhagyjuk és így általánosabban alkalmazható készletmodellt adunk. A modell több raktár (tároló) együttes induló készletszintjének [a (3.1) feladat] és optimális kapacitásának [a (3.2) feladat] meghatározását tűzi ki célul.

A PRÉKOPA A. [5] dolgozatában megfogalmazott több termékes, egy raktáras megbízhatósági készletmodell a termékeknek (alapanyagoknak) általunk modellezett folytonos, sztochasztikusan összefüggő beérkezési folyamata mellett a következő formában írható. Jelölje $M_1, \ldots, M_r$ a termékek induló készletszintjét, $\zeta_1(t), \ldots, \zeta_r(t)$ pedig a beérkezési folyamatát, melyet y_j, σ_j $(j=1, \ldots, r)$ paraméterű *homogén*

Wiener-folyamattal modellezünk (3.8) szerint, ahol $\mathbf{y}'=(y_1, \ldots, y_r)$ valószínűségi vektor $F(\mathbf{y})$ együttes eloszlásfüggvénnyel. Keressük a költségminimumot adó együttes induló készletszintet, mely a $z_1, \ldots, z_r$ intenzitású felhasználás szükségletét a $(0, T)$ időszakban adott $1-\varepsilon$ együttes valószínűséggel biztosítja:

$$\min G(\mathbf{M})$$

feltéve, hogy

$$g(\mathbf{M}) = P\Big(\sup_{0\leqq t\leqq T}\{z_j t-\zeta_j(t)\} < M_j,\ j=1, \ldots, r\Big) \geqq 1-\varepsilon, \tag{4.9}$$

$$\mathbf{0} \leqq \mathbf{M} \leqq \mathbf{Q}.$$

A (4.9) feladatnak a (3.2) feladattal való rokonsága azonnal szembetűnik. A $g(\mathbf{M})$ feltételi függvény konvexitására a 4.3. tételhez hasonló állítás igazolható, ha $\mathbf{y}$ együttes sűrűségfüggvénye logaritmikusan konkáv.

5. Számítógépes realizáció és numerikus eredmények

Az előző pontokban leírt gyakorlati problémák megoldására egy számítógépes programrendszert készítettünk, mely a leírt feladatok megoldási algoritmusain kívül bizonyos minőségellenőrzési, statisztikai és szimulációs eljárásokat is tartalmaz. Az egymáshoz kapcsolódó programokat a felhasznált módszerek szerint csoportosítva a következőkben tekintjük át.

1. *Statisztikai jellegű* programok a szemeloszlások vizsgálatára készültek. Ezek biztosítják a további programok alapadatait, de ezenkívül a minőségellenőrzési vizsgálatokban is felhasználják. Több évre visszamenőleg rendelkezésre állnak a felhasznált alapanyagok szemeloszlását jellemző α_{ik} mért értékei, melyek a k-adik alapanyagban a ϱ_i szemcseátmérőnél kisebb zúzalékok súlyarányai ($k=1, \ldots, 5$; $m=1, \ldots, 9$). A j-edik tárolóba ($j=1, \ldots, 4$) a $\tau_{j-1}\leqq d<\tau_j$ átmérőjű szemcsék kerülnek. A β_{jk} valószínűségi változó jelöli a $d<\tau_j$ szemcseátmérőjű zúzalék arányát. A τ_j értéket közrefogó két mérési hely, a ϱ_{i-1} és ϱ_i, valamint a megfelelő $\alpha_{i-1,k}$ és α_{ik} realizációinak felhasználásával lognormális eloszlást illesztünk m és s paraméterekkel, melyek értékeire (az anyagfajtára utaló k indexet elhagyva)

$$m = \frac{\log \varrho_{i-1}\Phi^{-1}(\alpha_i)-\log \varrho_i\Phi^{-1}(\alpha_{i-1})}{\Phi^{-1}(\alpha_i)-\Phi^{-1}(\alpha_{i-1})}$$

és

$$s = \frac{\log \varrho_i-\log \varrho_{i-1}}{\Phi^{-1}(\alpha_i)-\Phi^{-1}(\alpha_{i-1})}$$

adódik, ahol Φ^{-1} jelöli a standard normális eloszlásfüggvény inverzét. Az eloszlásfüggvénynek a τ_j helyen felvett értéke adja az interpolált β_j értéket:

$$\beta_j = \Phi\left(\frac{\Phi^{-1}(\alpha_i)(\log \tau_j-\log \varrho_{i-1})-\Phi^{-1}(\alpha_{i-1})(\log \tau_j-\log \varrho_i)}{\log \varrho_i-\log \varrho_{i-1}}\right). \tag{5.1}$$

A (3.4) szerint definiált $F(\mathbf{y})$ együttes eloszlást, valamint az $F_j(y_j)$ marginális eloszlásokat empirikus eloszlásokkal közelítettük, mert a mérési adatokhoz meg-

felelően illeszkedő, jól kezelhető folytonos eloszlást nem találtunk. A diszkrét eloszlások a (3.1) és (4.7) feladatok numerikus megoldását is egyszerűvé teszik.

2. *Kvadratikus programozási* algoritmusok nagyszámban ismertek, ezek közül *Beale módszerének* azt a változatát alkalmaztuk, melyet BERNAU H. [2] dolgozott ki. Ebben a változókra érvényes felső korlátokat (melyek technológiai okok miatt szükségesek) is figyelembe véve a lineáris programozási felső korlát technika megfelelő adaptálásával a [2] dolgozat szerzője jól működő eljárást és gyors programot dolgozott ki.

A (2.2) feladatot a gyártott 8 különféle aszfaltkeverékre oldottuk meg, mindegyik keverőtelep rendelkezésre álló adataira. Az eljárás gyorsasága lehetővé tenné az aktuális mérési adatok alapján az optimális keverési arány meghatározását mindegyik alapanyag-szállítmány esetén, melyek a gyakorlatban lényeges szemeloszlás-eltérést mutatnak. Ennek jelenleg a keverőtelepekkel való összeköttetés hiánya, az adatszolgáltatás lassúsága jelenti az akadályát. A keverési arányok meghatározása jelenleg grafikus közelítő eljárással történik, mely gyakran a szabvány betartását a keverék várható értékére sem biztosítja.

Példaként az ún. AB—8 típusú aszfaltkeverék előállításához szükséges keverési arányok számításának eredményeit közöljük. Az alábbi táblázat tartalmazza a felhasználható alapanyagok szemeloszlásának átlagos értékeit és szórásait, melyeket a minőségellenőrzés nyilvántartott mérési eredményeiből határoztunk meg. (Megjegyezzük, hogy ezen értékek egy része szintén az (5.1) szerint interpolált érték, mivel a minőségellenőrzési mérések nem mindegyik $i=1, \ldots, 9$ rostaméretre történnek meg.)

Alapanyag típusa a_i (mm)		NZ 0/5	NZ 5/12	NZ 12/20	Mészkőliszt	Homok
0,1	átlag:	0,080	0,010	0,010	0,720	0,020
	szórás:	0,042	0,002	0,001	0,092	0,009
0,2	átlag:	0,113	0,012	0,011	0,890	0,220
	szórás:	0,057	0,003	0,004	0,075	0,107
0,6	átlag:	0,224	0,022	0,016	0,980	0,580
	szórás:	0,082	0,009	0,005	0,004	0,135
2,0	átlag:	0,614	0,068	0,037	1,000	0,720
	szórás:	0,108	0,018	0,012	0,000	0,109
5,0	átlag:	0,950	0,240	0,103	1,000	0,950
	szórás:	0,017	0,083	0,053	0,000	0,031
8,0	átlag:	0,990	0,560	0,187	1,000	1,000
	szórás:	0,001	0,107	0,071	0,000	0,000
12,5	átlag:	1,000	0,950	0,330	1,000	1,000
	szórás:	0,000	0,012	0,131	0,000	0,000
20,0	átlag:	1,000	1,000	0,980	1,000	1,000
	szórás:	0,000	0,000	0,008	0,000	0,000
25,0	átlag:	1,000	1,000	0,999	1,000	1,000
	szórás:	0,000	0,000	0,000	0,000	0,000

Korlátozó feltételként szerepel, hogy a mészkőliszt aránya legfeljebb 8 százalék legyen, továbbá, hogy a 0,1 és 0,2 mm közötti átmérőjű szemcsék legfeljebb 50 százaléka lehet homokból származó. Ez utóbbi feltétel a keverék szabvány szemeloszlásának átlaga alapján az $x_5 \leqq 0{,}15$ korlátozást jelenti. A feltétel teljesülését a kiszámított keverési arányok alapján megvalósuló szemeloszlásra ellenőrizni kell, és esetleg más korlát felvételével az eljárást újra kell futtatni. A (2.2) kvadratikus programozási feladat az $\mathbf{x}' = (0{,}585;\ 0{,}214;\ 0{,}0;\ 0{,}08;\ 0{,}121)$ vektort adja, ezt tekintjük az optimális keverési aránynak. Itt a homok keverési arányára tett feltétel teljesül és ez nem jelent aktív korlátozást. A következő táblázat tartalmazza a keverék szemeloszlását (a várható értékeket), összehasonlítva a szabvány alsó és felső korlátjával, valamint ezek átlagával (a szabvány középvonalával).

mm	Keverék	Szabvány min.	Szabvány max.	Középvonal
0,1	0,109	0,10	0,12	0,110
0,2	0,167	0,12	0,20	0,160
0,6	0,296	0,25	0,35	0,300
2,0	0,541	0,45	0,65	0,550
5,0	0,802	0,70	0,85	0,775
8,0	0,900	0,85	1,00	0,925
12,5	0,982	0,95	1,00	0,975
20,0	1,000	1,00	1,00	1,000
25,0	1,000	1,00	1,00	1,000

3. *A készletmodell megoldása* mindegyik aszfaltkeverék-típusra külön-külön szükséges. Az induló q feltöltési idő típusonként különbözik. A szükséges tároló méretek is változnak, ezek tárolónkénti maximumára kell tervezni, hogy a túlfolyás valószínűsége mindegyik keverék típusra az adott korlát alatt legyen. A (3.2) feladat célfüggvényeként a tárolók összkapacitásának minimuma szerepelt, továbbá a $\mathbf{Q}$ kapacitáskorlát vektor sem jelentett aktív korlátozást. Így a (4.7) feltételt egyenlőséggel kielégítő K_j értékeket el lehetett fogadni optimális megoldásnak.

Példaként vázoljuk az AB—8 típusú aszfaltkeverékre nyert számítási eredményeket. A j-edik tárolóba $(j=1, 2, 3, 4)$ érkező anyagmennyiséget a (3.8) *homogén Wiener-folyamattal* modellezzük, melynek az y_j paramétere a γ_j valószínűségi változó realizációja. A rostákon áthulló maximális szemcseátmérők (a τ_j értékek) 3,63; 9,06; 12,7 és 25,38 mm. Jelölje a γ_j várható értékét μ_j, szórását s_j és U_{ij} a $\mu_j - 2{,}5\, s_j$, $\mu_j - 1{,}5\, s_j$, $\mu_j - 0{,}5\, s_j$, $\mu_j + 0{,}5\, s_j$, $\mu_j + 1{,}5\, s_j$, $\mu_j + 2{,}5\, s_j$ végpontokkal definiált intervallumokat, illetve v_{ij} ezek középpontját $(i=1, \ldots, 5;\ j=1, \ldots, 4)$. Meghatározzuk a $\gamma_1 \in U_{i_1 1}, \ldots, \gamma_4 \in U_{i_4 4}$ feltételeket együttesen kielégítő γ mintaelemek számát, ezt $k_{i_1, \ldots, i_4}$ jelöli, ahol i_j $(j=1, \ldots, 4)$ az 1, 2, ..., 5 számok mindegyikén végigfut. A (4.3) integrált a

$$\sum_{i_1=1}^{5} \cdots \sum_{i_4=1}^{5} \left[\prod_{j=1}^{4} g_j(q|v_{i_j j}) \right] k_{i_1, \ldots, i_4} / N \tag{5.2}$$

összeggel közelítjük, ahol N jelöli az összes mintaelem számát. Az output intenzitások vektora (3.5) alapján $\mathbf{z}' = (0{,}697;\ 0{,}237;\ 0{,}054;\ 0{,}012)$. A beérkezési folyamatok szórására a következő becslést adjuk: $\boldsymbol{\sigma}' = (0{,}043;\ 0{,}023;\ 0{,}005;\ 0{,}001)$. Ennek meghatározásáról a továbbiakban még írunk. A zavartalan működési pe-

riódus minimális hosszára a szakemberek 1 órás időszakot tartanak szükségesnek, mely idő alatt közel 100 tonna aszfaltkeverék készül el. A folyamatos gyártás valószínűségének minimális értékét 0,8-nak javasolják. Az egyes tárolók folyamatos anyagellátására ezt a valószínűségi korlátot vesszük (a marginális valószínűségek korlátja), de az együttes valószínűségi korlátra az $1-\varepsilon_1=0{,}75$ értéket javasoljuk. Az alábbi táblázatba foglaljuk, hogy a különböző q értékek (induló feltöltési idő, órában) milyen együttes valószínűséggel biztosítják a folyamatos anyagellátást:

q	0,14	0,16	0,18	0,20	0,22	0,24	0,26	0,28	0,30	0,32
$1-\varepsilon_1$	0,333	0,381	0,664	0,712	0,750	0,781	0,806	0,821	0,845	0,861

A műszaki szakemberekkel egyetértésben a $q=0{,}22$ óra induló várakozási időt tekintjük optimálisnak, melyhez a kiürülés szempontjából kritikus 3. tároló folyamatos anyagellátásának a valószínűsége 0,81, míg a többi tárolónál ez az érték 1-hez közeli.

A túlfolyás szempontjából az 1. tároló méretezése a kritikus az említett AB—8 típusú aszfaltkeverék esetén. Az alábbi táblázatban a 100 t/ó intenzitású gyártás esetén a különböző tárolóméretekhez tartozó értékek azt a valószínűséget fejezik ki, hogy legalább egy órás időtartam alatt nincs túlfolyás, ha az induló várakozási idő $q=0{,}22$ óra.

K_1 (tonna)	16,8	18,5	20,3	22,0	23,7	25,5	27,2
$1-\delta_1$	0,240	0,497	0,702	0,843	0,927	0,970	0,989

A javasolt $1-\delta_1=0{,}9$ valószínűséghez tartozó minimális tárolóméret 23,1 tonna. A négy tároló szükséges nagysága AB—8 típusú aszfaltkeverék gyártásához $\mathbf{K}'=(23{,}1;\ 9{,}2;\ 2{,}3;\ 1{,}1)$. Más aszfaltkeverékeknél ezek az értékek különböznek. A számításoknál a (4.8) integrált az (5.2) közelítő formula megfelelő egy dimenziós változata helyettesítette, itt azonban a γ_j empirikus sűrűségfüggvényét az (5.2)-ben szereplő 5 részintervallum helyett 11 részintervallumra történt felosztással közelítettük.

Fölmerült a kérdés, hogy a τ_j $(j=1, \ldots, 4)$ rostméretek változtatása esetén hogyan módosul a keverési arány és a szükséges tárolókapacitás. A műszakilag indokolt változatokra elvégeztük a fenti számításokat. Nagyobb termelési kapacitású gépeket is vizsgáltunk. Ezek eredményeivel, továbbá egyéb műszaki jellegű kérdésekkel és a részletes numerikus eredményekkel a BODNÁR G.—KELLE P. [3] cikkben foglalkozunk.

4. *Szimulációs eljárással* vizsgáltuk meg azt a kérdést, hogy a tárolók túlfolyása, illetve kiürülése esetén adott módon megváltoztatva az alapanyagok keverési arányát, mikor várható a következő hasonló jellegű zavar. A sztochasztikus szimulációs eljárás alkalmas a korábban megoldott sztochasztikus feladatok eredményeinek ellenőrzésére is.

Komoly gyakorlati nehézséget jelent a modellezett $\zeta_j(t)$ *Wiener-folyamatok* σ_j $(j=1, 2, 3, 4)$ paraméterének meghatározása. A minőségellenőrzés mérései

ritkán történnek ahhoz, hogy ebből jó becsléseket tudjunk adni. Az egyik keverőtelepen rendelkezésre állnak megfelelő számban a mérési adatok, ezt a műszaki szakértők véleményével kiegészítve használtuk fel a paraméterek rögzítésére. Az anyagáramlás időbeli lefolyását a rendelkezésre álló adatokkal és a modellezett folyamattal szimulálva összehasonlítottuk a kiürülési, illetve túlfolyási időpontokat, és ez alapján javítottuk a paraméterek becslését.

A programok Fortran nyelven készültek és az MTA SZTAKI CDC 3300-as típusú gépén futottak. Az eredmények felhasználását az *Útépítő Tröszt Műszaki Osztályán* végzik.

IRODALOM

[1] BAXTER, G. and DONSKER, M. D., "On the distribution of the supremum functional for processes with stationary independent increments", *Trans. Amer. Math. Soc.* **85** (1957) 73—87.
[2] BERNAU, H. „Felső korlát technikák a kvadratikus programozáshoz", *Alkalmazott Matematikai Lapok* **3** (1977) 161—170.
[3] BODNÁR, G. und KELLE, P., „Die optimale Grösse der Warmsilos von Asphaltmischanlagen", *Das stationäre Mischwerk. Der bituminöse Strassenbau* (megjelenés alatt).
[4] NÉMETH, GY., „Sztochasztikus készletmodellekkel kapcsolatos vizsgálatok", *MTA III. Oszt. Közleményei* **16** (1971) 133—135.
[5] PRÉKOPA, A., "Stochastic programming models for inventory control and water storage problems", Colloquia Mathematica Societatis János Bolyai 7. Inventory Control and Water Storage Győr, 1971. (Bolyai J. Math. Soc. and North Holland Publ. Comp. Budapest, 1973) 229—246.
[6] PRÉKOPA, A., "On logarithmic concave measures and functions", *Acta Scientiarum Mathematicarum* **34** (1973) 335—343.
[7] PRÉKOPA, A., "Programming under probabilistic constraints with a random technology matrix", *Mathematische Operationsforschung u. Statistik* **5** (1974) 109—116.
[8] PRÉKOPA, A. és KELLE, P., „Sztochasztikus programozáson alapuló megbízhatósági készletmodellek", *Alkalmazott Matematikai Lapok* **2** (1976) 1—16.

(Beérkezett: 1979. október 12.)

KELLE PÉTER
MTA SZÁMÍTÁSTECHNIKAI ÉS AUTOMATIZÁLÁSI KUTATÓ INTÉZETE
1502 BUDAPEST XI., KENDE U. 13—17.

OPTIMIZATION OF MIXTURE RATE AND DEPOT CAPACITIES FOR ASPHALT MIXERS

P. KELLE

The optimal mixture rate of raw materials of varying compositions is determined by quadratic programming. A stochastic inventory control model is given for the 4 hot depots of the asphalt mixers. The input processes are approximated by *Wiener processes* with stochastically dependent parameters, the outputs are linear. The probability of stockout and owerflow of depots, in a given time period, may not exceed a prescribed level. A reliability type model is formulated for the initial stock level and for the optimal depot capacities. The mathematical properties and the solution of the model is discussed. We give reference to the connection with reliability type inventory models.

Alkalmazott Matematikai Lapok **5** *(1979)* 261—268

EGY KVÁZI-BELSŐ PONT ELJÁRÁS LINEÁRIS ÉS NEMLINEÁRIS FELTÉTELEKET TARTALMAZÓ NEMLINEÁRIS PROGRAMOZÁSI FELADATOK MEGOLDÁSÁRA

WILLIAM F. AMAR AIGBE
Budapest

Ebben a dolgozatban egy olyan belső pont eljárással foglalkozunk, amely különbözik a SUMT belső pont algoritmusaitól, mégpedig abban, hogy csak a nem-lineáris feltételekkel büntetünk. Részletesen megvizsgáljuk e kvázi-belső pont módszer és az optimalitás összefüggését, és bebizonyítjuk az algoritmus konvergenciáját.

1. Bevezetés

Mint tudjuk, a SUMT (*Sequential Unconstrained Minimization Techniques*) módszerek olyan nemlineáris programozási algoritmusok, amelyek a feladat megoldását feltétel nélküli minimalizálások sorozatára vezetik vissza. Ebben a cikkben olyan kvázi-belső pont módszert konstruáltunk, amely lineáris feltételek melletti minimalizálások sorozatára vezeti vissza a feladatot. Más szóval, ha adva van egy nemlineáris programozási feladat egyenlőtlenséges feltételekkel, a feltételek közül csak a nemlineárisokat büntetjük meg. Így egy lineáris feltételekkel korlátozott nemlineáris programozási feladatot kapunk, amelyet *Ritter konjugált irányos módszere* segítségével oldunk meg.

A feladat megfogalmazása

Tekintsük a következő nemlineáris programozási feladatot

$$\text{(I)} \qquad \begin{aligned} &\operatorname{Min} f(\mathbf{x}) \\ &g_i(\mathbf{x}) \leqq 0, \quad i = 1, \ldots, r, \\ &\mathbf{Ax} \leqq \mathbf{b}, \\ &\mathbf{x} \in E^n, \end{aligned}$$

ahol E^n az *n-dimenziós Euklideszi teret* jelöli. Feltesszük, hogy az $f(\mathbf{x})$, $g_i(\mathbf{x})$, $i=1, \ldots, r$ konvex függvények és hogy folytonosan differenciálhatók E^n-en.

Legyen $\mathbf{A}$ egy $m \times n$-es mátrix és $\mathbf{b} \in E^m$.

Definiáljuk az

$$S_1 = \{\mathbf{x} | g_i(\mathbf{x}) \leqq 0;\ i = 1, 2, \ldots, r\}, \tag{1.1}$$

$$S_{10} = \{\mathbf{x} | g_i(\mathbf{x}) < 0;\ i = 1, 2, \ldots, r\}, \tag{1.2}$$

$$S_2 = \{\mathbf{x} | \mathbf{Ax} \leqq \mathbf{b}\} \tag{1.3}$$

és az

$$S = (S_1 \cap S_2) \in E^n$$

halmazokat. Tehát S a megengedett tartománya az (I) feladatnak. Legyen $S_0 = S_{10} \cap S_2$. Feltesszük, hogy S_0 nem üres és S kompakt halmaz. Feltételeink mellett az $f(\mathbf{x})$ függvény S-en felveszi az infimumát.

A feladathoz konstruálunk egy az S_0-on értelmezett $F(\mathbf{x}, \beta)$ segédfüggvényt és helyettesítjük az (I) feladatot a következő segédfeladatok sorozatával:

(II_k) $$\operatorname{Min} F(\mathbf{x}, \beta^k),$$

$$\mathbf{Ax} \leqq \mathbf{b},$$

$$\mathbf{x} \in S_{10},$$

ahol $F(\mathbf{x}, \beta^k) = f(\mathbf{x}) + \beta^k B(g(\mathbf{x}))$, $\beta^k > 0$, $k = 0, 1, 2, \ldots$, itt $B(g(\mathbf{x}))$ az S_0-on értelmezett valós, folytonos függvény $\{\beta^k\}$ egy pozitív, monoton fogyva nullához konvergáló sorozat, $k \to \infty$-re.

$F(\mathbf{x}, \beta)$ függvény tulajdonságai

1. $F(\mathbf{x}, \beta)$ folytonosan differenciálható S_0-on minden $\beta > 0$-ra.
2. Ha $\mathbf{x}^* \in S_0$, akkor $\lim_{\beta^k \to 0} F(\mathbf{x}^*, \beta^k) = f(\mathbf{x}^*)$.
3. A $B(\mathbf{y}): R^r \to R$ függvénynek olyannak kell lennie, hogy ha $\mathbf{x}^N \in S_0$ és $\mathbf{x}^N \to \hat{\mathbf{x}} \in (S - S_{10})$, $N \to \infty$-re (azaz $\hat{\mathbf{x}}$ egy határpont), akkor minden rögzített $\beta > 0$-ra, $F(\mathbf{x}^N, \beta) \to +\infty$.

E tulajdonságokból nyilvánvaló, hogy az $F(\mathbf{x}, \beta)$ függvény S_0-on felveszi az infinumát rögzített β mellett.

A módszer

Az alapgondolat a következő:

Az (I) feladathoz szerkesztünk $F(\mathbf{x}, \beta^k)$ segédfüggvényeket, amelyeket minimalizálunk $k = 0, 1, 2, \ldots$-re, $\mathbf{Ax} \leqq \mathbf{b}$ lineáris feltételekkel; az (II_k)-t megoldjuk. $\mathbf{x}_0 \in S_0$ pontból kiindulva, valamilyen β^0 mellett minimalizáljuk a (II_0) feladatot. A megoldás β^0-tól függ. Jelöljük a (II_0) megoldását $\mathbf{x}_0^*(\beta^0)$-lal. $F(\mathbf{x}, \beta^0)$ tulajdonságai miatt $\mathbf{x}_0^*(\beta^0) \in S_0$. Az így kapott minimum pontból kiindulva valamilyen $\beta^1 < \beta^0$ mellett újra minimalizáljuk a (II_1)-et. Az eljárást folytatva, mindig a (II_k)-t minimalizáljuk egy $\beta^k < \beta^{k-1}$ mellett, ahol $\{\beta^k\} \to 0$; a kapott $\mathbf{x}_{k-1}^*(\beta^{k-1})$ pontból kiindulva. Megmutatható, hogy $\mathbf{x}_k^*(\beta^k)$ sorozat minden torlódási pontja megoldása (I)-nek. A kvázibelső pont módszer elnevezés onnan ered, hogy nem a megengedett tartomány belső pontjaival, hanem S_0 pontjaival dolgozunk.

A gyakrabban használt B függvények a következők:

$$B_1(\mathbf{g}(\mathbf{x})) = -\sum_{i=1}^{r} \ln(-g_i(\mathbf{x})). \tag{1.4}$$

$$B_2(\mathbf{g}(\mathbf{x})) = -\sum_{i=1}^{r} \frac{1}{g_i(\mathbf{x})}, \tag{1.5}$$

$$B_3(\mathbf{g}(\mathbf{x})) = \sum_{i=1}^{r} \frac{1}{[g_i(\mathbf{x})]^2}, \tag{1.6}$$

$$B_4(\mathbf{g}(\mathbf{x})) = -\sum_{i=1}^{r} \frac{1}{\operatorname{Min}[0, g_i(\mathbf{x})]}. \tag{1.7}$$

2. Optimalitási kritériumok

Legyen $\mathbf{x}^*$ az (I) feladat optimális megoldása, és definiáljuk az aktív indexek halmazát a következőképpen

$$I_g(\mathbf{x}^*) = \{i | g_i(\mathbf{x}^*) = 0\}$$

és

$$I_A(\mathbf{x}^*) = \{j | \mathbf{a}_j^T \mathbf{x} = b_j\}.$$

Megköveteljük a következő regularitási kritérium teljesülését:

(2.1) Az $\mathbf{x}^*$ pontban a $\nabla g_i(\mathbf{x}^*)$, $i \in I_g(\mathbf{x}^*)$ és az $\mathbf{a}_j$, $j \in I_A(\mathbf{x}^*)$ vektorok lineárisan függetlenek.

Ebből az (I) feladatra, már következik, hogy ez $\mathbf{x}^*$ pontban teljesülnek a *Kuhn—Tucker szükséges optimalitási feltételek*. Ezt mondja ki a következő tétel.

2.1. Tétel. Tegyük fel, hogy $\mathbf{x}^*$ az (I) feladat optimális megoldása. Ha az (I) feladatra teljesül a (2.1) regularitási kritérium, akkor $\mathbf{x}^*$-hoz léteznek olyan $\lambda_i^* \geqq 0$ és $\mu_j^* \geqq 0$ *Lagrange-szorzók*, hogy fennáll a

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^{r} \lambda_i^* \nabla g_i(\mathbf{x}^*) + \sum_{j=1}^{m} \mu_j^* \mathbf{a}_j = \mathbf{0}, \tag{2.2}$$

$$\lambda_i^* g_i(\mathbf{x}^*) = 0, \quad i = 1, \ldots, r, \tag{2.3}$$

$$\mu_j^* \mathbf{a}_j^T \mathbf{x}^* = 0, \quad j = 1, \ldots, m \tag{2.4}$$

egyenletrendszer.

A (2.2) feltételek a *Lagrange-féle szükséges feltételek*, a (2.3) és (2.4) feltételek elnevezése pedig: *komplementaritási feltételek*.

Bevezetve a

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\mathbf{x}) + \sum_{i=1}^{r} \lambda_i g_i(\mathbf{x}) + \sum_{j=1}^{m} \mu_j (\mathbf{a}_j^T \mathbf{x} - b_j) \tag{2.5}$$

jelölést, az ún. *Lagrange-féle szükséges feltétel* így is írható:

$$\nabla L(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \mathbf{0}, \tag{2.6}$$

ahol λ^* egy r-dimenziós vektort és μ^* egy m-dimenziós vektort jelöl, melynek i-edik komponense λ_i^*, illetve j-edik komponense μ_j^*.

3. A kvázi-belső pont módszer és az optimalitás összefüggése

Az (1.4) barrier-függvényt felhasználva, az (I) feladatból a következőt kapjuk:

$$(\mathrm{II}_k) \qquad \mathrm{Min}\left\{f(\mathbf{x})-\beta^k\sum_{i=1}^{r}\ln\left(-g_i(\mathbf{x})\right)\right\},$$

$$\mathbf{a}_j^T\mathbf{x}\leqq b_j,\quad j=1,\ldots,m,$$

$$\mathbf{x}\in S_{10}.$$

Tegyük fel, hogy $\mathbf{x}_k^*(\beta^k)$ a (II_k) részfeladatnak a megoldása. Mivel ez (II_k)-nak egy *Kuhn—Tucker-pontja*, ezért érvényes a következő: $\exists\boldsymbol{\mu}\geqq\mathbf{0}$,

$$(3.1) \qquad \nabla f\left(\mathbf{x}_k^*(\beta)\right)-\sum_{i=1}^{r}\frac{\beta^k}{g_i\left(\mathbf{x}_k^*(\beta^k)\right)}\nabla g_i\left(\mathbf{x}_k^*(\beta^k)\right)+\sum_{j=1}^{m}\mu_j(\beta^k)\mathbf{a}_j=\mathbf{0}.$$

Legyen

$$(3.2) \qquad \lambda_i(\beta^k)=-\frac{\beta^k}{g_i\left(\mathbf{x}_k^*(\beta^k)\right)},\quad i=1,2,\ldots,r,$$

ekkor

$$(3.3) \qquad \nabla f\left(\mathbf{x}_k^*(\beta^k)\right)+\sum_{i=1}^{r}\lambda_i(\beta^k)\nabla g_i\left(\mathbf{x}_k^*(\beta^k)\right)+\sum_{j=1}^{m}\mu_j(\beta^k)\mathbf{a}_j=\mathbf{0}.$$

Most tekintsük az (I) feladathoz tartozó (2.3) és (2.4) komplementaritási feltételeket. Perturbáljuk a (2.3) feltételrendszert, azaz írjunk 0 helyett $-\beta$ t, ahol $\beta>0$. Ekkor a következő perturbált feltételrendszert kapjuk:

$$(3.4) \qquad \nabla f(\mathbf{x})+\sum_{i=1}^{r}\lambda_i\nabla g_i(\mathbf{x})+\sum_{j=1}^{m}\mu_j\mathbf{a}_j=\mathbf{0},$$

$$(3.5) \qquad \lambda_i g_i(\mathbf{x})=-\beta,\quad i=1,\ldots,r,$$

$$(3.6) \qquad \mu_j\mathbf{a}_j^T\mathbf{x}=0,\quad j=1,\ldots,m,$$

$$\lambda_i\geqq 0,$$

$$\mu_j\geqq 0.$$

Tegyük fel, hogy ez az egyenletrendszer $\mathbf{x}$-re, $\boldsymbol{\lambda}$-ra és $\boldsymbol{\mu}$-re megoldható és a megoldásokat jelöljük $\mathbf{x}(\beta)$-, $\lambda(\beta)$- és $\mu(\beta)$-val. Továbbá tegyük fel, hogy $\mathbf{x}(\beta)$ megengedett pont. (3.5)-ből fejezzük ki λ_i-t és így kapjuk a

$$\nabla f\left(\mathbf{x}(\beta)\right)-\sum_{i=1}^{r}\frac{\beta}{g_i\left(\mathbf{x}(\beta)\right)}\nabla g_i(\mathbf{x})\sum_{j=1}^{m}\mu_j(\beta)\mathbf{a}_j=\mathbf{0}$$

egyenletet.

Ennek az egyenletnek a bal oldala nem más, mint a fenti barrier-függvény gradiense az $\mathbf{x}(\beta)$ helyen. Ebből adódik, hogy a *Kuhn—Tucker-féle elsőrendű opti-*

malitási kritériumban szereplő *Lagrange-szorzók* a barrier-módszer alkalmazásakor automatikusan adódnak. Ezt a tényt a következő tételben fogalmazzuk meg.

3.1. Tétel: Legyen $\mathbf{x}_k^*(\beta^k) \to \mathbf{x}^*$, ahol $\mathbf{x}^*$ az (I) feladat megoldása. Az f, g_i, $i=1, \ldots, r$ függvények legyenek folytonosan differenciálhatók és az aktív feltételek gradiensei az $\mathbf{x}_k^*(\beta^k)$ és $\mathbf{x}^*$ helyen lineárisan függetlenek. Akkor, ha $\beta^k \to 0$,

$$\lambda(\beta^k) \to \lambda(0) = \lambda^*$$

és

$$\mu(\beta^k) \to \mu(0) = \mu^*,$$

ahol (λ^*, μ^*) az $\mathbf{x}^*$ ponthoz tartozó *Lagrange-vektorpárt* jelöli az (I) feladatnál.

Bizonyítás: A bizonyítás megértéséhez először tegyük fel, hogy egyenlőségek állnak a lineáris feltételeknél a (II_k) feladatban. Tegyük fel, hogy $\beta^k \to 0$. Ha $i \notin I_g(\mathbf{x}^*)$, akkor

$$-\frac{\beta^k}{g_i(\mathbf{x}_k^*(\beta^k))} \to 0.$$

Most megmutatjuk, hogy $\mathbf{x}(\beta^k) \to \mathbf{x}^*$ esetén, $\lambda(\beta^k) \to \lambda^*$ és $\mu(\beta^k) \to \mu^*$, továbbá $\mathbf{x}^*, (\lambda^*, \mu^*)$ kielégítik a *Kuhn—Tucker-feltételeket.* Ez pedig konvex feladat esetén az optimalitás elégséges feltétele.

(3.3)-ból

$$-\nabla f(\mathbf{x}_k^*(\beta^k)) = \sum_{i \in I_g(\mathbf{x})} \lambda_i(\beta^k) \nabla g_i(\mathbf{x}_k^*(\beta^k)) + \sum_{j=1}^{m} \mu_j(\beta^k)\mathbf{a}_j + o_k(1). \tag{3.7}$$

Legyen $\vartheta_k = o_k(1)$ és az egyszerűség kedvéért tegyük fel, hogy a nemlineáris feltételek közül az első s darab feltétel aktív, akkor (3.7) mátrix formába írva:

$$-\begin{bmatrix} \nabla_{x_1} f(\mathbf{x}_k^*(\beta^k)) \\ \vdots \\ \nabla_{x_n} f(\mathbf{x}_k^*(\beta^k)) \end{bmatrix} = \begin{bmatrix} \nabla_{x_1} g_1(\mathbf{x}_k^*(\beta^k)) \ldots \nabla_{x_1} g_s(\mathbf{x}_k^*(\beta^k)) a_{11} \ldots a_{1m} \\ \vdots \\ \nabla_{x_n} g_1(\mathbf{x}_k^*(\beta^k)) \ldots \nabla_{x_n} g_s(\mathbf{x}_k^*(\beta^k)) a_{n1} \ldots a_{nm} \end{bmatrix} \begin{bmatrix} \lambda_1(\beta^k) \\ \vdots \\ \lambda_s(\beta^k) \\ \mu_1(\beta^k) \\ \vdots \\ \mu_m(\beta^k) \end{bmatrix} + \vartheta_k. \tag{3.8}$$

Jelöljük az $n \times (m+s)$ méretű mátrixot $\mathbf{H}(\beta^k)$-val és legyen

$$\eta^T(\beta^k) = (\lambda_1(\beta^k), \ldots, \lambda_s(\beta^k), \mu_1(\beta^k), \ldots, \mu_m(\beta^k).$$

Tehát (3.8)-ból kapjuk a

$$-\nabla_{\mathbf{x}} f(\mathbf{x}_k^*(\beta^k)) = \mathbf{H}(\beta^k)\eta(\beta^k) + \vartheta_k \tag{3.9}$$

egyenletet, amely

$$-\hat{\nabla} f((\mathbf{x}_k^*(\beta^k)) = \hat{\mathbf{H}}(\beta^k)\eta(\beta^k) + \vartheta_k \tag{3.10}$$

alakban írható, ha $\hat{\mathbf{H}}(\beta^k)$-val az $s+m$ méretű nemszinguláris négyzetes mátrixot, illetve $\hat{\nabla}_{\mathbf{x}}(\mathbf{x}_k^*(\beta^k))$-val a megfelelő dimenziójú vektort jelöljük. Mivel a gradiensek folytonosak, kiválasztható olyan részmátrix, hogy elég nagy k_0-ra minden $k > k_0$

esetén a $\hat{\mathbf{H}}(\beta^k)$ mátrix nemszinguláris lesz az $\mathbf{x}^*$ közelében.

Ha $\mathbf{x}_k^*(\beta^k) \to \mathbf{x}^*$, akkor mivel az f és g_i, $i=1, \ldots, s$ gradiensei folytonosak, a határátmenet után, a függetlenség miatt, (3.10) a következő alakban írható:

$$(3.11) \qquad -\begin{bmatrix} \nabla_{x_1} f(\mathbf{x}^*) \\ \vdots \\ \nabla_{x_n} f(\mathbf{x}^*) \end{bmatrix} = \begin{bmatrix} \nabla_{x_1} g_1(\mathbf{x}^*) \ldots \nabla_{x_1} g_s(\mathbf{x}^*) a_{11} \ldots a_{1m} \\ \vdots \\ \nabla_{x_n} g_1(\mathbf{x}^*) \ldots \nabla_{x_1} g_s(\mathbf{x}^*) a_{n1} \ldots a_{nm} \end{bmatrix} \begin{bmatrix} \lambda_1(0) \\ \vdots \\ \lambda_s(0) \\ \mu_1(0) \\ \vdots \\ \mu_m(0) \end{bmatrix}.$$

Mint tudjuk, az (I) feladat *Lagrange-függvénye*

$$(3.12) \qquad L(\mathbf{x}, \lambda, \mu) = f(\mathbf{x}) + \sum_{i=1}^{r} \lambda_i \nabla g_i(\mathbf{x}) + \sum_{j=1}^{m} \mu_j (\mathbf{a}_j^T \mathbf{x}_1 - b_j)$$

és ennek gradiense

$$(3.13) \qquad \nabla L(\mathbf{x}^*, \lambda^*, \mu^*) = \nabla f(\mathbf{x}^*) + \sum_{i=1}^{r} \lambda_1^* \nabla g_i(\mathbf{x}^*) + \sum_{j=1}^{m} \mu_j^* \mathbf{a}_j = 0.$$

A (3.11) és (3.13) egyenlőségeket összehasonlítva, a gradiensek lineáris függetlensége miatt

$$\eta^T(0) = (\lambda^*, \mu^*).$$

Ha most a (II_k) lineáris feltételekben egyenlőtlenségek állnak, akkor a lineáris függetlenség (amely egyben a *Lagrange-szorzók* egyértelműségét is jelenti) miatt a bizonyítás ugyanúgy megy.

Vegyük észre, hogy bár $\lambda_i(\beta^k) \geqq 0$, az (I) feladatra a *Kuhn—Tucker elégséges feltételek* nem teljesülnek az $\mathbf{x}_k^*(\beta^k)$ pontban, mivel

$$\lambda_i(\beta^k) g_i(\mathbf{x}_k^*(\beta^k)) = \frac{-\beta^k}{g_i(\mathbf{x}_k^*(\beta^k))} \cdot g_i(\mathbf{x}_k^*(\beta^k)) = -\beta^k < 0.$$

Ebből a barrier függvény érdekes primál-duál tulajdonságait kaphatjuk. Lásd FIACCO, MCCORMICK (1968) és AVRIEL (1976). A fenti tételből az is látszik, hogy a *Lagrange-szorzók* becsülhetők a (3.2) módon definiált λ_i számokból, amelyek nagy szerepet játszanak az aktív-stratégia választásánál.

Konvergencia feltételek:

1. f konvex és folytonosan differenciálható függvény E^n-ben.
2. g_i, $i=1, \ldots, r$ folytonosan differenciálható és kvázi-konvex E^n-ben. Továbbá, legyen $S_{10} = S_1^0$, ahol S_1^0 az S_1 belsejét jelöli.
3. S kompakt halmaz.
4. $B(\mathbf{x}) \geqq 0$, B folytonos és $B(\mathbf{x}) \to +\infty$, ha $\mathbf{x} \to \hat{\mathbf{x}}$, ahol $\hat{\mathbf{x}}$ S_1 határpontja.
5. $\liminf_{k \to \infty} F(\mathbf{x}^k, \beta^k) \geqq f(\mathbf{x}^*)$, ahol x^k megoldása (II_k)-nek.

3.1. Lemma: A

$$\operatorname*{Min}_{\mathbf{x} \in S_0} F(\mathbf{x}, \beta^k)$$

feladatnak a fenti feltételek fennállása esetén mindig létezik megoldása.

Bizonyítás: Legyen

$$L_0 = \{\mathbf{x} | F(\mathbf{x}, \beta^k) \leqq F(\mathbf{x}^0, \beta^k),\ \mathbf{x} \in S_0\},$$

ahol $\mathbf{x}^0$ egy kezdő pont. Ez nyilván egy kompakt részhalmaza S_0-nak. Tehát a

$$\operatorname*{Min}_{\mathbf{x} \in L_0} f(\mathbf{x}, \beta^k)$$

feladatnak létezik megoldása és minden megoldása egyben az eredeti feladat optimális megoldása is.

A 4. feltevés nem minden barrier-függvényre teljesül, de az alábbi gondolatmenet kis módosításával az állítás azokra az esetekre is megmutatható.

3.2. Tétel: Tegyük fel, hogy a fenti feltevések teljesülnek. Akkor az $\mathbf{x}^k$ sorozaminden torlódási pontja megoldása (I)-nek.

Bizonyítás: Egyszerűség kedvéért tegyük fel, hogy $\mathbf{x}^k \to \tilde{\mathbf{x}}$. Legyen D az optimális megoldások konvex kompakt halmaza, továbbá legyen σ az (I) feladat optimális célfüggvényértéke. Tegyük fel, hogy $\tilde{\mathbf{x}} \notin D$. Egy olyan $\mathbf{x}'$ pontot akarunk találni, amely eleme az $S_{10} \cap S_2$ halmaznak és $f(\tilde{\mathbf{x}}) > f(\mathbf{x}') > \sigma$.

Első eset: Ha $S_{10} \cap S_2 \subset D$, akkor mivel f folytonos és $S_{10} \cap S_2$ konvex, ezért $D = S$. Tehát f konstans ezen a halmazon, ami ellentmond annak a feltet vésünknek, hogy $\tilde{\mathbf{x}} \notin D$.

Második eset: Van olyan $\hat{\mathbf{x}} \in S_{10} \cap S_2$, amelyre $\hat{\mathbf{x}} \notin D$ teljesül. Legyen $\hat{\hat{\mathbf{x}}} \in D$. Akkor $0 \leqq \lambda < 1$-re: $\lambda \hat{\hat{\mathbf{x}}} + (1-\lambda)\hat{\mathbf{x}} \in S_{10} \cap S_2$. Mivel $f(\hat{\hat{\mathbf{x}}}) = \sigma$ és $f(\hat{\mathbf{x}}) > \sigma$, ezért az f függvény folytonossága miatt találhatunk egy fenti tulajdonságú $\mathbf{x}'$ pontot. Így

$$\liminf_{k \to \infty} F(\mathbf{x}^k, \beta^k) \geqq f(\tilde{\mathbf{x}}) > f(\mathbf{x}') = \lim_{k \to \infty} F(\mathbf{x}', \beta^k).$$

Elég nagy k esetén szigorú egyenlőtlenséggel állunk szemben:

$$F(\mathbf{x}^k, \beta^k) > F(\mathbf{x}', \beta^k),$$

ami ellentmond annak, hogy $\mathbf{x}^k$ a (II_k) feladat optimális megoldása.

A módszerre FORTRAN-program készült, és azt a MTA CDC 3300-as gépén leteszteltük. Példaként a következő kisméretű feladatot [2] teszteltük:

$$\operatorname{Min}(x_1 + x_2)$$

$$G(\mathbf{x}) = P\begin{Bmatrix} 2x_1 + x_2 - 6 \geqq \beta_1 \\ x_1 + 8x_2 - \delta \geqq \beta_2 \end{Bmatrix} \geqq 0{,}8,$$

$$x_1 + 4x_2 \geqq 4,$$

$$3x_1 + x_2 \geqq 3$$

$$x_1 \geqq 0, \quad x_2 \geqq 0,$$

ahol β_1, β_2 együttes eloszlása normális, várható értékük 0, szórásuk 1 és a korrelációs együttható 0,2. A feladat megoldásánál 30 másodperc alatt találtuk meg az optimumot.

A segédfeladatok megoldására *Ritter konjugált irányos módszerét* választottuk [6], amelyre folytonosan differenciálható célfüggvény és lineáris feltételek mellett a konvergencia sebessége szuperlineáris. Az algoritmus leírását az olvasó megtalálhatja [6]-ban.

IRODALOM

[1] AVRIEL, M., *Nonlinear Pogramming-Analysis and Methods* (Prentice-Hall, Inc. — New Jersey, 1976).
[2] DEÁK, I., „Egy sztochasztikus programozási modell számítógépes kiértékelése", *MTA Számítástechnikai Központja, Közlemények* **9** (1972) 33—49.
[3] FIACCO, A. V. and MCCORMICK, G. P., *Nonlinear Programming: SUMT* (Wiley, New York, London, 1968).
[4] GERENCSÉR, L., „Nemlineáris programozási feladatok megoldása szekvenciális módszerekkel", *MTA—SZTAKI Tanulmányok* **49** (1976).
[5] PRÉKOPA, A., „Sztochasztikus rendszerek optimálizálási problémáiról", doktori értekezés, MTA, Budapest, 1970.
[6] RITTER, K., "A method of Conjugate directions for linearly constrained nonlinear programming problems", *SIAM J. Num. Analysis* **12** (1975) 273—303.

(Beérkezett: 1979. november 15.)

WILLIAM F. AMAR AIGBE
MTA SZÁMÍTÁSTECHNIKAI ÉS AUTOMATIZÁLÁSI KUTATÓ INTÉZET
1250 BUDAPEST URI U. 49.

A BARRIER METHOD FOR SOLVING NONLINEAR PROGRAMMING PROBLEMS

W. F. AMAR AIGBE

In this article a "barrier" method is dealt with. This method is different from that of the SUMT — method in the sense that a general nonlinear programming problem is transformed into sequential linearly constrained nonlinear programming problem.

Alkalmazott Matematikai Lapok 5 (1979) 269—276

GLOBÁLIS MINIMUM MEGHATÁROZÁSA KIZÁRÁSOS MÓDSZERREL

KÁLOVICS FERENC
Miskolc

Az egyenletrendszerek (egyenletek) numerikus megoldásával foglalkozó irodalomban találhatók ún. kizárási tételek, amelyek felhasználhatók az előre adott, korlátos D tartományba eső összes megoldás megkeresésére [1, 2, 4, 5]. Ebben a dolgozatban egy általánosabb kizárási tételből kiindulva, a fentiektől eltérő alkalmazási lehetőséget mutatunk be.

1. Egy kizárási tétel

1.1. Tétel. Az $f: D \subset R^n \to R^m$ függvényre teljesüljön, hogy

$$\|f(x)-f(a)\|_2 \leqq K_1\|x-a\|_1+K_2\|x-a\|_1^2, \tag{1.1}$$

ahol $x, a \in D$; $K_1=K_1(a)$, $K_2=K_2(a)>0$; $\|\cdot\|_1$ és $\|\cdot\|_2$ pedig tetszőleges R^n-, ill. R^m-beli vektornorma. Akkor az f függvény $\alpha \in D$ zérushelyeire teljesül, hogy

$$\|\alpha-a\|_1 \geqq \sqrt{\frac{\|f(a)\|_2}{K_2}+\left(\frac{K_1}{2K_2}\right)^2}-\frac{K_1}{2K_2} \tag{1.2}$$

tetszőleges $a \in D$ pontban.

Bizonyítás. Az $x=\alpha$ választással az (1.1) egyenlőtlenség a

$$K_2\|\alpha-a\|_1^2+K_1\|\alpha-a\|_1-\|f(a)\|_2 \geqq 0$$

alakra hozható. $\|\alpha-a\|_1$-ra nézve ez egy másodfokú egyenlőtlenség. $K_2>0$ és (1.1) alapján $K_1 \geqq 0$, tehát a

$$K_2\|\alpha-a\|_1^2+K_1\|\alpha-a\|_1-\|f(a)\|_2=0$$

egyenletnek pontosan egy nem-negatív megoldása van, amely

$$\|\alpha-a\|_1=\sqrt{\frac{\|f(a)\|_2}{K_2}+\left(\frac{K_1}{2K_2}\right)^2}-\frac{K_1}{2K_2}.$$

Ez pedig éppen az (1.2) teljesülését jelenti.

Megjegyzések. 1. Ha az $f(\not\equiv \text{konst.})$ függvényre (1.1) helyett a *Lipschitz-folytonosságot* írjuk csak elő ($K_2=0$), akkor

$$\|\alpha-a\|_1 \geqq \frac{\|f(a)\|_2}{K_1}, \tag{1.3}$$

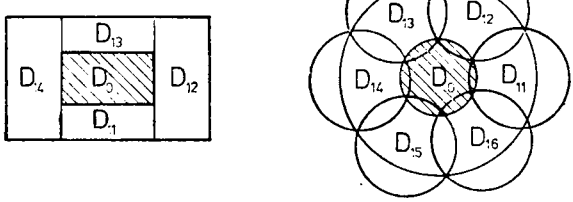
s ez — természetesen — megegyezik az (1.2)-ből $K_2 \to 0$ határátmenettel nyerhető eredménnyel.

2. Ha az (1.1) feltétel a K_1, K_2 és K_1', K_2' értékpárra is teljesül és

$$K_1\|x-a\|_1+K_2\|x-a\|_1^2 \leqq K_1'\|x-a\|_1+K_2'\|x-a\|_1^2,$$

akkor a K_1', K_2' párral az (1.2)-ből kapott eredmény nem javul a K_1, K_2 párral számítotthoz képest.

3. Egyenletek és egyenletrendszerek megoldásánál a D tartományt véges tégla-, ill. gömb-tartománynak célszerű választani. Ha az f függvény ezen D-n kétszer folytonosan differenciálható (*Frechet-értelemben*), akkor biztosan létezik (1.1)-nek megfelelő K_1 és K_2. A tégla-, ill. gömbtartomány középpontját választva az „a" pontnak, az (1.2) ill. (1.3) alapján (a megfelelő normát használva) kizárható — a gyökök létezésére nézve — egy „a" középpontú kocka (tégla)-, ill. gömbtartomány. A „maradék tartományt" az eredetinél kisebb átmérőjű tégla-, ill. gömbtartományokkal „ügyesen" lefedve, majd ismételve a fenti eljárást, a D-beli gyököket véges számú lépésben megközelíthetjük előre adott pontossággal [1]. Az 1. ábra az [1]-ben javasolt lefedés kétdimenziós esetét, ill. az [5]-ben használt lefedést mutatja.



1. ábra

4. Az 1.1. tételben szereplő feltétel mellé további kikötéseket véve, megadhatók hibabecslő-formulák a kizárásos módszerekhez is [1], de ezek értéke két okból is minimális:

— a kizárásos módszerek előnye más, többváltozós esetben is használható, egyenletmegoldó eljárásokkal [3] szemben éppen az, hogy olyankor is használhatók, amikor nem teljesülnek az egyéb eljárások konvergenciáját biztosító feltételek, vagy azok vizsgálata nehézkes;

— a kizárásos módszereket addig célszerű használni, míg egy gyorsabb, de csak lokálisan konvergens módszerhez megfelelő közelítést nyerünk.

2. Alkalmazás globális minimum keresésére

Legyen adott a

$$\frac{-r \leqq x_i \leqq r;\ i = 1, 2, \ldots, n}{f(x_1, x_2, \ldots, x_n) \to \min.} \tag{2.1}$$

szélsőérték-feladat, ahol f nem-negatív értékű, kétszer folytonosan differenciálható függvény a

$$D = \{x^T = (x_1, x_2, \ldots, x_n) \in R^n: \ -r \leqq x_i \leqq r; \ i = 1, 2, \ldots, n\}$$

kockán.

A feladatot felfoghatjuk úgy is, hogy az

$$f(x_1, x_2, \ldots, x_n) - m = 0 \tag{2.2}$$

egyenletnek a

$$\tilde{D} = \{\tilde{x}^T = (x_1, x_2, \ldots, x_n, m) \in R^{n+1}: \ -r \leqq x_i \leqq r; \ m \geqq 0\}$$

téglában („négyzetes oszlopban") a legkisebb $n+1$-edik koordinátájú $\tilde{\alpha} \in \tilde{D}$ megoldását (megoldásait) kell keresni. Ezen kereséshez felhasználhatjuk az (1.2), ill. (1.3) összefüggéseket, ha megadjuk a megfelelő K_1, ill. K_2 értékeket.

A továbbiakban legyen f kétszer folytonosan differenciálható és jelölje $f'(x), f''(x)$ és $\|\cdot\|$ a következőket

$$f'(x) = (\partial_1 f(x), \partial_2 f(x), \ldots, \partial_n f(x))$$

$$f''(x) = \begin{bmatrix} \partial_1\partial_1 f(x) \ldots \partial_n\partial_1 f(x) \\ \vdots \qquad\qquad \vdots \\ \partial_1\partial_n f(x) \ldots \partial_n\partial_n f(x) \end{bmatrix}$$

$$\|C\| = \max_i \sum_k |c_{ik}|,$$

ahol $C = [c_{ik}]$ valós, véges mátrix.

2.1. Tétel. A $g(x_1, x_2, \ldots, x_n, m) = f(x_1, x_2, \ldots, x_n) - m$ függvényre teljesül, hogy

$$|g(\tilde{x}) - g(\tilde{a})| \leqq \left(1 + \sup_{y \in D} \|f'(y)\|\right) \|\tilde{x} - \tilde{a}\|,$$

(2.3) illetve

$$|g(\tilde{x}) - g(\tilde{a}) \leqq \left(1 + \|f'(a)\|\right) \|\tilde{x} - \tilde{a}\| + \frac{n}{2} \sup_{y \in D} \|f''(y)\| \, \|\tilde{x} - \tilde{a}\|^2,$$

ahol $\tilde{x}, \tilde{a} \in \tilde{D} \subset R^{n+1}$; $x, a \in D \subset R^n$ pedig az $\tilde{x}$, ill. $\tilde{a}$ vektorokból az $n+1$-edik koordináta elhagyásával keletkező vektorok.

Bizonyítás. A $g: \tilde{D} \subset R^{n+1} \to R^1$ függvény kétszer folytonosan differenciálható, tehát

$$g(\tilde{x}) - g(\tilde{a}) = g'(\tilde{\xi})(\tilde{x} - \tilde{a}),$$

(2.4) illetve

$$g(\tilde{x}) - g(\tilde{a}) = g'(\tilde{a})(\tilde{x} - \tilde{a}) + \frac{1}{2}(\tilde{x} - \tilde{a})^T g''(\tilde{\eta})(\tilde{x} - \tilde{a}),$$

ahol $\tilde{x}, \tilde{a}, \tilde{\xi}, \tilde{\eta} \in \tilde{D}$.

Másrészt

$$g(\tilde{x}) = f(x) - m,$$

tehát

$$\|g'(\tilde{x})\| = \|f'(x)\| + 1.$$

Most vegyük a (2.4) egyenletek mindkét oldalának $\|\cdot\|$ normáját és a jobb oldalon használjuk fel a norma két alaptulajdonságát! Akkor:

$$|g(\tilde{x})-g(\tilde{a})| \leqq (1+\|f'(\xi)\|)\|\tilde{x}-\tilde{a}\|,$$

illetve

$$|g(\tilde{x})-g(\tilde{a})| \leqq (1+\|f'(a)\|)\cdot\|\tilde{x}-\tilde{a}\|+\frac{1}{2}\|(\tilde{x}-\tilde{a})^T g''(\tilde{\eta})\|\,\|\tilde{x}-\tilde{a}\|.$$

Ha figyelembe vesszük még, hogy

$$\|(\tilde{x}-\tilde{a})^T g''(\tilde{\eta})\| = \|(x-a)^T f''(\eta)\| \leqq \|(x-a)^T\|\cdot\|f''(\eta)\| \leqq n\|\tilde{x}-\tilde{a}\|\|f''(\eta)\|,$$

akkor már nyilvánvaló a (2.3) egyenlőtlenségek teljesülése.

A továbbiakban feltételezzük, hogy a (2.3)-ban szereplő

$$1+\sup_{y\in D}\|f'(y)\|, \quad \text{ill.} \quad \frac{n}{2}\sup_{y\in D}\|f''(y)\|$$

„globális jellemzőknek" ismerjük egy (pozitív értékű) felső korlátját (K_1, K_2).

Ilyen felső korlátok ismeretében — az 1.1. és 2.1. tételek alapján — az $\tilde{a}\in\tilde{D}$ pontban a kizárható kocka sugara („a kocka élei párhuzamosak a koordinátatengelyekkel"):

$$\varrho_1(\tilde{a}) = \frac{|f(a)-m|}{K_1},$$

(2.5) vagy

$$\varrho_2(\tilde{a}) = \sqrt{\frac{|f(a)-m|}{K_2}+\left(\frac{\|f'(a)\|+1}{2K_2}\right)^2}-\frac{\|f'(a)\|+1}{2K_2}.$$

A (2.2) feladat megoldását ezek után a következő lépésekkel írhatjuk le:

1. Kezdjük el a megoldás (megoldások) keresését a

$$D_0 = \{(x_1, x_2, \ldots, x_n, 0):\ -r \leqq x_i \leqq r;\ i = 1, 2, \ldots, n\}$$

n-dimenziós kockában. Azaz határozzuk meg (2.5) alapján az $\tilde{a}=(0, 0, \ldots, 0)\in R^{n+1}$ pontban a kizárható kocka sugarát $(\varrho_0=(\varrho_1)_0$ vagy $(\varrho_2)_0)$. Ha ez az érték $<r$, akkor osszuk fel a kockát 2^n számú, $0{,}5r$ sugarú kockára. Ezek középpontjai:

$$\left((-1)^{k_1}\frac{r}{4}, (-1)^{k_2}\frac{r}{4}, \ldots, (-1)^{k_n}\frac{r}{4}\right),$$

ahol $k_i=0$ vagy 1. Amelyik kocka most sem lesz kizárható, azt osszuk újra fel. Ha $m=0$-ra nincs megoldása a feladatnak, akkor (2.5) alapján létezik olyan finom felosztás, amelynél már minden kocka kizárható. Ha valamely felosztás („teljes lefedés") alkalmával a kiszámolt sugárértékek minimuma $m_0(\neq 0)$, akkor a (2.2) feladatnak $m<m_0$-ra nincs megoldása, tehát az f függvény minimuma ennél nagyobb.

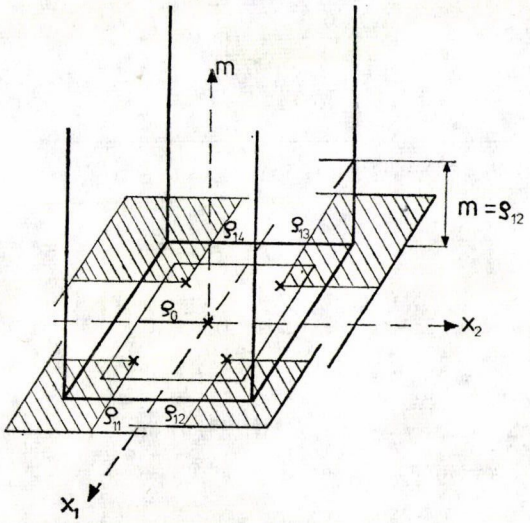
2. Ha $m_0\neq 0$, akkor növeljük m értékét m_0-ra és folytassuk a keresést a

$$D_1 = \{(x_1, x_2, \ldots, x_n, m_0):\ -r \leqq x_i \leqq r;\ i = 1, 2\ \ldots, n\}$$

$n+1$ dimenziós hasábban a korábbi módon.

3. A fenti típusú lépéseket ismételve, $m_i \leqq \varepsilon_1 (f(x) - m < \varepsilon_2)$ helyzetben a keresést befejezzük, vagy más módszerre térünk át.

Az $n=2$ esetre a 2. ábra szemlélteti az eljárást.



2. ábra

Megjegyzés. Az optimumszámítási feladatok egy része könnyen visszavezethető (2.1) típusú feladatra. Pl. az

$$A_{11} \leqq a_{11}x_1 + a_{12}x_2 \leqq A_{12}$$

$$\underline{A_{21} \leqq a_{21}x_1 + a_{22}x_2 \leqq A_{22}}$$

$$f(x_1, x_2) \to \min.$$

feladat az

$$a_{11}x_1 + a_{12}x_2 = u_1$$

$$a_{21}x_1 + a_{22}x_2 = u_2$$

új változókkal (ha $a_{11}a_{22} \neq a_{12}a_{21}$) egy lineáris transzformáció után (2.1) típusú feladatra vezethető vissza úgy, hogy a változók száma nem is nő.

3. Egy numerikus példa

Adottak az

$$\mathbf{r}_1(t) = \{\cos t;\ \cos t;\ 0{,}6 + \sin t\} \quad \frac{\pi}{6} \leqq t \leqq \frac{\pi}{3}$$

$$\mathbf{r}_2(u) = \left\{\cos u;\ \sin u;\ \frac{1}{2}u\right\} \quad \frac{\pi}{6} \leqq u \leqq \frac{\pi}{3}$$

térbeli görbék (ellipszis-, ill. csavarvonal ív).

A következőkben megmutatjuk, hogyan használható fel a bevezetett eljárás a két görbe (euklideszi) távolságának meghatározására. Célunk tehát a

$$\frac{\pi}{6} \leqq t \leqq \frac{\pi}{3}$$

$$\frac{\pi}{6} \leqq u \leqq \frac{\pi}{3}$$

$$f(t, u) = 0{,}25u^2 - 0{,}6u + 2{,}36 - 2\cos t(\sin u + \cos u) + \cos^2 t + (1{,}2 - u)\sin t \to \min.$$

globális szélsőérték-feladat megoldása.

Mivel

$$f'(t, u) =$$

$$= \{2\sin t(\sin u + \cos u) - \sin 2t + (1{,}2 - u)\cos t;\ 0{,}5u - 0{,}6 + 2\cos t(\sin u - \cos u) - \sin t\},$$

illetve

$$f''(t, u) = \begin{bmatrix} 2\cos t(\sin u + \cos u) - 2\cos 2t - (1{,}2 - u)\sin t; & 2\sin t(\cos u - \sin u) - \cos t \\ 2\sin t(\cos u - \sin u) - \cos t; & 0{,}5 + 2\cos t(\cos u + \sin u) \end{bmatrix}$$

ezért $\frac{\pi}{6} \leqq t, u \leqq \frac{\pi}{3}$ esetén

$$1 + \|f'\| < 7{,}20 = K_1; \quad \|f''\| < 6{,}16 = K_2.$$

A megadott K_1, K_2 érték durva $\left(\text{a } \left[0; \frac{\pi}{2}\right] \text{ intervallumra is érvényes}\right)$ becslésből származik, de gépi úton még ilyen K_1, ill. K_2 érték meghatározása is nehézkes lehet. Az eljárás további része viszont könnyen programozható. Ahhoz, hogy formálisan is a (2.1) feladat alakjához jussunk, a

$$t = \frac{\pi}{4} + \tau \qquad -\frac{\pi}{12} \leqq \tau \leqq \frac{\pi}{12}$$

$$u = \frac{\pi}{4} + \vartheta \qquad -\frac{\pi}{12} \leqq \vartheta \leqq \frac{\pi}{12}$$

összefüggésekkel új változókat kellene bevezetni. Mivel most csak néhány lépést végzünk el a számolásból, így a transzformációnak nincs meg a számítástechnikai előnye, ezért mellőzzük.

Legyen először $m=0$. A (2.5) képletek alapján

$$\varrho_1\left(\frac{\pi}{4}; \frac{\pi}{4}; 0\right) = \frac{0{,}836}{7{,}2} = 0{,}116; \quad \varrho_2\left(\frac{\pi}{4}; \frac{\pi}{4}; 0\right) = 0{,}190$$

Mivel $\varrho_1, \varrho_2 < \frac{\pi}{12} = 0{,}262$, ezért az „alapnégyzetet" felosztjuk négy darab $\frac{\pi}{24}$

sugarú négyzetre. A (2.5) alapján

$$\varrho_1\left(\frac{5\pi}{24}; \frac{5\pi}{24}; 0\right) = \frac{0{,}811}{7{,}2} = 0{,}112; \quad \varrho_2\left(\frac{5\pi}{24}; \frac{5\pi}{24}; 0\right) = 0{,}181$$

$$\varrho_1\left(\frac{7\pi}{24}; \frac{5\pi}{24}; 0\right) = \frac{1{,}170}{7{,}2} = 0{,}162; \quad \varrho_2\left(\frac{7\pi}{24}; \frac{5\pi}{24}; 0\right) = 0{,}222$$

$$\varrho_1\left(\frac{7\pi}{24}; \frac{7\pi}{24}; 0\right) = \frac{0{,}908}{7{,}2} = 0{,}126; \quad \varrho_2\left(\frac{7\pi}{24}; \frac{5\pi}{24}; 0\right) = 0{,}207$$

$$\varrho_1\left(\frac{5\pi}{24}; \frac{7\pi}{24}; 0\right) = \frac{0{,}597}{7{,}2} = 0{,}082; \quad \varrho_2\left(\frac{5\pi}{24}; \frac{7\pi}{24}; 0\right) = 0{,}171.$$

A kedvezőbb ϱ_2 értékeket véve alapul, $\varrho_2 > \frac{\pi}{24}$ minden esetben, tehát az

$$m < 0{,}171$$

értékekre nincs a feladatnak megoldása.

Most legyen $m = 0{,}171$. A $\left(\frac{\pi}{4}; \frac{\pi}{4}\ 0{,}171\right)$ pontban felesleges a ϱ_1, ϱ_2 értékeket meghatározni, hisz ϱ_1 és ϱ_2 az m-nek monoton csökkenő függvényei. Az

$$\left(\frac{5\pi}{24}; \frac{5\pi}{24}; 0{,}171\right); \quad \left(\frac{7\pi}{24}; \frac{5\pi}{24}; 0{,}171\right); \quad \left(\frac{7\pi}{24}; \frac{7\pi}{24}; 0{,}171\right); \quad \left(\frac{5\pi}{24}; \frac{7\pi}{24}; 0{,}171\right)$$

pontokban pedig minimális számolással meghatározhatjuk a ϱ_1, ϱ_2 értékeket, mert az előbbi $f(t, u)$, $\|f'(t, u)\|$ értékek felhasználhatók. Ezekben a pontokban rendre

$$\varrho_1 = 0{,}088 \quad \varrho_2 = 0{,}149$$

$$\varrho_1 = 0{,}138 \quad \varrho_2 = 0{,}196$$

$$\varrho_1 = 0{,}102 \quad \varrho_2 = 0{,}174$$

$$\varrho_1 = 0{,}059 \quad \varrho_2 = 0{,}131.$$

A $\varrho_2 > \frac{\pi}{24}$ egyenlőtlenség most is teljesül minden esetben, tehát

$$m \geqq 0{,}171 + 0{,}131 = 0{,}302.$$

Az eddigi számolás során az előforduló legkisebb függvényérték:

$$f(t, u) = 0{,}597,$$

tehát

$$0{,}302 \leqq m \leqq 0{,}597. \tag{3.1}$$

A következő lépésben ($m = 0{,}302$), amint a fenti ϱ_2 értékek alapján is várható, lesz olyan négyzet, amelyet tovább kell bontani — pl. az $\left(\frac{5\pi}{24}; \frac{7\pi}{24}; 0{,}302\right)$ közép-

pontú, és lesz olyan is, amely továbbra is „megfelelő méretű" — pl. a $\left(\frac{7\pi}{24}; \frac{5\pi}{24}; 0{,}302\right)$ középpontú.

Megjegyezzük még, hogy az eredeti feladat a $d = \min \sqrt{f(t, u)}$ euklideszi távolság keresésére vonatkozott, s erre (3.1) alapján a

$$0{,}549 = \sqrt{0{,}302} \leqq d \leqq \sqrt{0{,}597} = 0{,}773$$

egyenlőtlenség írható fel.

IRODALOM

[1] KÁLOVICS, F., „Nemlineáris egyenletrendszerek megoldása érintőparaboloid-módszerrel" *NME Közleményei (Miskolc) IV. Sorozat, Természettudományok* **23** (1977) 19—33.
[2] KÁLOVICS, F., „Kizárási tételek többváltozós polinom zérushelyeinek kereséséhez", *NME Közleményei (Miskolc) IV. Sorozat, Természettudományok* (sajtó alatt).
[3] ORTEGA, J. M. and RHEINBOLDT, W. C., *Iterative Solutions of Nonlinear Equations in Several Variables* (Academik Press, 1970).
[4] SZABÓ, Z., „Über gleichungslösende Iterationen ohne Divergenzpunkt, I—II", *Publ. Math. Debreceniensis* **20** (1973) 223—233, **21** (1974) 285—293.
[5] TÓTH, B., „Polinom összes zérushelyének érintőparaboloid módszerrel való meghatározása", *NME Közleményei (Miskolc) IV. Sorozat, Természettudományok* (sajtó alatt).

(Beérkezett: 1979. április 18.)

KÁLOVICS FERENC
NME MATEMATIKAI INTÉZETE
3515 MISKOLC, EGYETEMVÁROS

DETERMINATION OF THE GLOBAL MINIMUM BY THE METHOD OF EXCLUSIONS

F. KÁLOVICS

In the literature dealing with the numerical solution of systems of equations (equations) there are exclusion theorems suitable for finding all the solutions falling within a limited domain D which is given beforhand. In this paper, starting from a more general exclusion theorem a new application is presented.

Alkalmazott Matematikai Lapok 5 *(1979)* 277—288

A PEGAZUS MÓDSZEREK NEMLINEÁRIS EGYENLETEK MEGOLDÁSÁRA

KALMÁR JÁNOS
Sopron

A *Pegazus módszer* — amely nevét onnan kapta, hogy először egy „Pagasus" számítógép szubrutinkönyvtárában bukkantak rá —, a BIT folyóirat 12. kötetében lett publikálva 1972-ben. Ezt követte 1973-ban a módosított *Pegazus módszer* bemutatása — szintén a BIT-ben —, mely az eredeti módszer algoritmusát fejlesztette tovább. Mindkét cikkben az ismertetett hibaanalízis jóformán csak az eredményekre szorítkozott, ezért a számításokat DR. MÓRICZ FERENC docens úr biztatására megismételtem, s a kapott új eredmények ösztönöztek a *Pegazus módszerek* újabb publikálására. MÓRICZ FERENC úrnak ezúton is köszönöm a cikk megírásához nyújtott sokoldalú segítségét.

1. A gyökközelítő iterációs módszerek általános tulajdonságai

Mindenekelőtt TRAUB [3] könyve alapján ismertetem az iterációs függvények elméletével kapcsolatos azon eredményeket, melyeket a hibaanalízis és a rend meghatározása folyamán erősen kihasználunk.

Egy gyökközelítő algoritmus iterációs függvénye általában a következő alakú:

$$x_{i+1} = \Phi(x_i, \ldots, x_{i-n}), \tag{1.1}$$

ami azt jelenti, hogy az új közelítést a módszer a megelőző $n+1$ közelítés alapján számítja, az eljárás $n+1$ lépcsős. A levezetett hibaképletek általános alakja ekkor

$$e_{i+1} \approx A \prod_{j=0}^{n} e_{i-j}^{l_j},$$

azaz

$$\lim_{|e_i| \to 0} \frac{e_{i+1}}{\prod_{j=0}^{n} e_{i-j}^{l_j}} = A, \tag{1.2}$$

ahol $e_k = x_k - \alpha$, és α az $f(x)=0$ egyenlet keresett gyöke.

Egy bizonyos iterációs módszert p-edrendűnek akkor nevezünk, ha a megoldandó egyenletek valamely osztályára (ez esetben egyszeres gyökök esetén) teljesül a következő összefüggés:

$$\lim_{|e_i| \to 0} \frac{|e_{i+1}|}{|e_i|^p} = K, \tag{1.3}$$

ahol $K \neq 0$ az aszimptotikus hibakonstans.

Egy Φ módszer numerikus hatékonyságát p rendje és d információszükséglete (d =az iterációs lépésenként szükségessé váló új függvényérték-számítások száma) segítségével a következőképpen definiáljuk:

$$\text{EFF}(\Phi) = \sqrt[d]{p}. \tag{1.4}$$

Az (1.2) hibaképletű módszer differenciaegyenletéhez rendeljük hozzá az $x^{n+1} - \sum_{j=0}^{n} l_j x^{n-j}$ polinomot, amit ezentúl karakterisztikus egyenletnek hívunk. Ekkor TRAUB eredményeinek némi általánosításával kimondható a következő

Tétel. Adott Φ iterációs módszer rendje a hibaképletéhez rendelt karakterisztikus egyenlet 1-nél nagyobb valós gyöke, aszimptotikus hibakonstansa pedig

$$K = A^{\frac{p-1}{q-1}}, \quad \text{ahol} \quad q = \sum_{j=0}^{n} l_j. \tag{1.5}$$

2. A Pegazus módszerek bemutatása

A *Pegazus módszerek* tulajdonképpen a húrmódszer legújabb variánsai, melyek a húrmódszer egyszeres gyökökre vonatkozó stabiltását a többlépcsős módszerek *gyors konvergenciájával* egyesítik. Most pedig lássuk az iteráció lefolyását!

Tudjuk, hogy a húrmódszer algoritmusa az új közelítést mindig a gyök különböző oldalára eső közelítéseken át húzott húr és az x tengely metszéspontjaként definiálja. Ez az algoritmus biztosítja, hogy az iteráció nem pattanhat el, mint időnként a szelőmódszernél, viszont a konvergencia igen lassú, csak elsőrendű. Például konvex függvény esetén az egyik alappont végig helyben marad, az iteráció csak a másik oldalról közeledik a gyökhöz. Ezt az algoritmust először az *Illionis-módszer* fejlesztette tovább azáltal, hogy a stabil alappontban a függvényértéket felezte, ami a konvergenciát valóban felgyorsította.

A *Pegazus módszerek* ehhez hasonlóan a húrmódszer algoritmusát azzal javítják tovább, hogy a stabil alappontban a függvényértéket egy, az előző közelítésektől függő számmal szorozzák. Az 1. ábra _._._._ jelölése mutatja, hogyan módosítja a *Pegazus módszer* a húrmódszer algoritmusát:

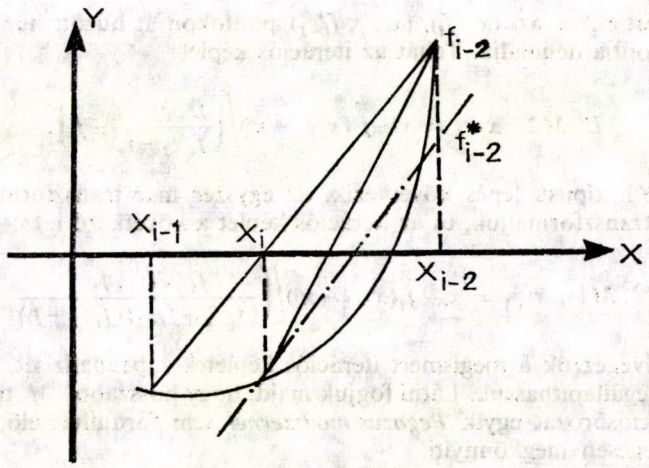
Jelöljük U-val azt a lépést, ahol x_{i+1}-et a húrmódszer algoritmusában az $(x_i, f_i), (x_{i-1}, f_{i-1})$ pontokon át húzott húr és az x tengely metszéspontja definiálja — ez a lépéstípus a *Pegazus módszereknél* is akkor fordul elő, ha $f_i f_{i-1} < 0$ teljesül. Ennek iterációs képlete a már jól ismert szelőképlet:

$$x_{i+1} = x_i - f_i(x_{i-1} - x_i)/(f_{i-1} - f_i), \tag{2.2}$$

melynek hibaképlete

$$e_{i+1} \approx \frac{c_2}{c_1} e_i e_{i-1}. \tag{2.3}$$

Ha a fenti előjelfeltétel nem teljesül, de előtte U típusú lépést hajtottunk végre, a húrmódszer a következő húrt az $(x_i, f_i), (x_{i-2}, f_{i-2})$ pontokon át húzza meg, a



1. ábra

Pegazus módszerek pedig egy M_1 típusú lépést hajtanak végre az (x_i, f_i), (x_{i-2}, f^*_{i-2}) pontra támaszkodva, ahol f^*_{i-2} a következőképpen transzformált függvényértéket jelöli (ami a párhuzamos szelők tétele alapján egyszerűen megszerkeszthető):

$$f^*_{i-2} = f_{i-2} f_{i-1}/(f_{i-1}+f_i). \tag{2.4}$$

Ekkor tehát az U típusú lépést követő $M1$ típusú lépés iterációs képlete a következő lesz:

$$U, M1 \quad x_{i+1} = x_i - f_i(x_{i-2}-x_i)\Big/\left(\frac{f_{i-2}f_{i-1}}{f_{i-1}+f_i} - f_i\right). \tag{2.5}$$

Ha az előjelfeltétel értelmében az első $M1$ típusú lépés után is a stabil (most már x_{i-3}) alappontra támaszkodik az iteráció, akkor az előbb már transzformált függvényértéket az előző képletnek megfelelően ismét transzformálni kell. Eredményül a következő iterációs képlet adódik:

$$U, M1, M1 \quad x_{i+1} = x_i - f_i(x_{i-3}-x_i)\Big/\left(\frac{f_{i-3}f_{i-2}f_{i-1}}{(f_{i-2}+f_{i-1})(f_{i-1}+f_i)} - f_i\right). \tag{2.6}$$

A most vázolt iterációsorozatok még mindkét *Pegazus módszernél* előfordulhatnak, de a *módosított Pegazus módszernél* olyan újabb lépéstípus is előfordulhat — ezt a továbbiakban $M2$-vel jelöljük —, amit a következőképpen definiálunk:

Ha U típusú lépést a húrmódszer algoritmusa szerint U típusú lépés követ, a *módosított Pegazus módszer* ekkor az f_{i-1} függvényértéket a következőképpen transzformálja:

$$f^*_{i-1} = f_{i-1} f_{i-2}/(f_{i-2}+f_i). \tag{2.7}$$

Az új közelítést ekkor az (x_i, f_i), (x_{i-1}, f_{i-1}^*) pontokon át húzott húr és az x tengely metszéspontja definiálja. Tehát az iterációs képlet:

$$U, M2 \quad x_{i+1} = x_i - f_i(x_{i-1} - x_i) \Big/ \left(\frac{f_{i-1} f_{i-2}}{f_{i-2} + f_i} - f_i \right). \tag{2.8}$$

Ha ezután $M1$ típusú lépés következik, az egyszer már transzformált függvényértéket ismét transzformáljuk, és az iterációs képlet a következő lesz:

$$U, M2, M1 \quad x_{i+1} = x_i - f_i(x_{i-2} - x_i) \Big/ \left(\frac{f_{i-3} f_{i-2} f_{i-1}}{(f_{i-3} + f_{i-1})(f_{i-1} + f_i)} - f_i \right). \tag{2.9}$$

Ezután elvégezzük a megismert iterációs képletek hibaanalízisét, hogy a módszer rendjét megállapíthassuk. Látni fogjuk majd, hogy hosszabb, M típusú lépésekből álló iterációsorozat egyik *Pegazus módszernél* sem fordulhat elő, ami a hibaanalízist lényegesen megkönnyíti.

Megjegyezzük még, hogy a *módosított Pegazus módszernél* megismert $M2$ típusú lépés csak U típusú lépést követhet, s utána akár U, akár $M1$ típusú lépés következhet.

3. A Pegazus módszerek hibaanalízise

Először az $U, M1$ iterációsorozat hibaanalízisét végezzük el. A már látott (2.5) iterációs képletet átrendezve kapjuk, hogy

$$e_{i+1} = \frac{A}{B} = \frac{f_{i-1}(e_i f_{i-2} e_{i-2}) - e_{i-2} f_i^2}{f_{i-2} f_{i-1} - f_i f_{i-1} - f_i^2}. \tag{3.1}$$

A következőkben kihasználjuk még az alábbi *Taylor-sorfejtéseket:*

$$\begin{aligned} f_{i-2} &= e_{i-2}(c_1 + c_2 e_{i-2} + c_3 e_{i-2} + O(e_{i-1}^3)), \\ f_{i-1} &= e_{i-1}(c_1 + O(e_{i-1})), \\ f_i &= e_i(c_1 + c_2 e_i + c_3 e_i^2 + O(e_i^3)), \\ f_i^2 &= e_i^2 c_1(c_1 + 2c_2 e_i + O(e_i^2)). \end{aligned} \tag{3.2}$$

Helyettesítsük be (3.1) számlálójába a fenti sorfejtéseket:

$$\begin{aligned} A = e_{i-1}(c_1 + O(e_{i-1}))e_i e_{i-2}\{c_2(e_{i-2} - e_i) + c_3(e_{i-2}^2 - e_i^2) + O(e_{i-2}^3) + O(e_i^3)\} - \\ - e_{i-2} e_i c_1(c_1 + 2c_2 e_i + O(e_i^2)). \end{aligned} \tag{3.3}$$

(2.3)-at kihasználva kapjuk, hogy

$$\begin{aligned} A &\approx e_i e_{i-1} e_{i-2}^2 \{(c_1 + O(e_{i-1}))(c_2 + c_3 e_{i-2} + O(e_{i-1})) - c_2(c_1 + O(e_{i-1} e_{i-2}))\} = \\ &= e_i e_{i-1} e_{i-2}^3 \left(c_1 c_3 + O\left(\frac{e_{i-1}}{e_{i-2}}\right) + O(e_{i-1}) \right). \end{aligned} \tag{3.4}$$

Hasonlóképpen a nevezőbe behelyettesítve:

$$(3.5)\qquad B = e_{i-1}(c_1+O(e_{i-1}))(c_1(e_{i-2}-e_i)+O(e_{i-2}^2)+O(e_i^2))-e_i^2c_1(c_1+O(e_i)).$$

Megint kihasználva (2.3)-at, kapjuk:

$$(3.6)B\quad \approx e_{i-1}e_{i-2}(c_1+O(e_{i-1}))(c_1+O(e_{i-1}))+O(e_{i-1}^2e_{i-2}^2) = e_{i-1}e_{i-2}(c_1^2+O(e_{i-1})).$$

Feltéve, hogy az $O\left(\frac{e_{i-1}}{e_{i-2}}\right)$ tag elhanyagolható (vagyis minden lépésben „sokat" finomodik a közelítés), képezhetjük az A/B hányadost:

$$(3.7)\qquad e_{i+1} \approx \frac{c_3}{c_1} e_i e_{i-2}^2.$$

Hasonlóan az $U, M2$ típusú sorozatnál is levezethető, hogy ha

$$(3.8)\qquad e_{i+1} = \frac{A'}{B'},$$

akkor

$$(3.9)\ A' = e_{i-2}(c_1+c_2e_{i-2}+O(e_{i-2}^2))e_ie_{i-1}(c_2(e_{i-1}-e_i)+c_3(e_{i-1}^2-e_i^2)+O(e_i^3)+$$
$$+O(e_{i-1}^3))-e_{i-1}e_i^2c_1(c_1+2c_2e_i+O(e_i^2)),$$

$$(3.10)\ B' = e_{i-2}(c_1+c_2e_{i-2}+O(e_{i-2}^2))(c_1(e_{i-1}-e_i)+O(e_{i-1}^2)+O(e_i^2))-e_i^2c_1(c_1+O(e_i))$$

Ugyanis mindenütt csak az $i-2$ és $i-1$ indexeket kellett felcserélni (mivel az iterációs képletek is csak ebben különböznek). Ezután használjuk ki (2.3)-at:

$$(3.11)\quad A' \approx e_ie_{i-1}^2e_{i-2}\left\{(c_1+c_2e_{i-2}+O(e_{i-2}))\left(c_2+c_3e_{i-1}+O(e_{i-1}^2)-\frac{c_2^2}{c_1}e_{i-2}\right)-\right.$$
$$\left.-c_2(c_1+O(e_{i-1}e_{i-2}))\right\},$$

mert $O(e_{i-1})=O(e_{i-2})$. Továbbá

$$(3.12)\qquad B' \approx e_{i-1}e_{i-2}(c_1+O(e_{i-2})).$$

Tehát

$$(3.13)\qquad e_{i+1} \approx \frac{c_3}{c_1} e_i e_{i-1}^2.$$

Most vizsgáljuk meg az $U, M1, M1$ iterációsorozatot, amelynek iterációs képlete (2.6). Ezt rendezve kapjuk, hogy

$$(3.14)\qquad x_{i+1} = \frac{x_if_{i-1}f_{i-2}f_{i-3}-x_{i-3}f_i(f_{i-1}+f_i)(f_{i-3}+f_{i-1})}{f_{i-1}f_{i-2}f_{i-3}-f_i(f_{i-1}+f_i)(f_{i-2}+f_{i-1})}.$$

Tehát a hibaképlet

$$(3.15)\qquad e_{i+1} = \frac{A_1}{B_1} = \frac{e_if_{i-1}f_{i-2}f_{i-3}-e_{i-3}f_i(f_{i-1}+f_i)(f_{i-2}+f_{i-1})}{f_{i-1}f_{i-2}f_{i-3}-f_i(f_{i-1}+f_i)(f_{i-2}+f_{i-1})}.$$

Helyettesítsük be a számlálóba a megfelelő sorfejtéseket:

$$A_1 = f_{i-1}f_{i-2}(e_i f_{i-3} - e_{i-3}f_i) - e_{i-3}f_i(f_{i-1}^2 + f_i f_{i-2} + f_i f_{i-1}) =$$
$$= e_i e_{i-1} e_{i-2} e_{i-3}(c_1 + O(e_{i-1}))(c_1 + c_2 e_{i-2} + O(e_{i-1}^2))(c_2(e_{i-3} - e_i) +$$
$$+ c_3(e_{i-3}^2 - e_i^2) + O(e_{i-3}^2)) - e_i e_{i-3}(c_1 + O(e_i))(e_{i-1}^2(c_1^2 + O(e_{i-1})) +$$
$$+ e_i e_{i-2}(c_1 + O(e_i))(c_1 + c_2 e_{i-2} + O(e_{i-2}^2)) + O(e_i e_{i-1})). \tag{3.16}$$

(2.3)-at és (3.7)-et kihasználva kapjuk, hogy

$$A_1 \approx e_i e_{i-1} e_{i-2} e_{i-3}^2(c_1^2 c_2 + c_1 c_2^2 e_{i-2} + c_1 c_3 e_{i-3} + O(e_{i-1}) + O(e_{i-2}^2) +$$
$$+ O(e_{i-3}^2)) - e_i e_{i-1} e_{i-2} e_{i-3}^2(c_1 + O(e_i))[c_2(c_1 + O(e_{i-1})) +$$
$$+ c_3 e_{i-3}(1 + O(e_i))(c_1 + c_2 e_{i-2} + O(e_{i-2}^2)) + O(e_{i-3}^2)] =$$
$$= e_i e_{i-1} e_{i-2}^2 e_{i-3}^2(c_1 c_2^2 + O(e_{i-3})). \tag{3.17}$$

Helyettesítsük be a nevezőbe is a megfelelő sorfejtéseket:

$$B_1 = e_{i-1}e_{i-2}(c_1 + O(e_{i-1}))(c_1 + O(e_{i-2}))(c_1(e_{i-3} - e_i) +$$
$$+ O(e_{i-3}^2) + O(e_i^2) + O(e_i e_{i-1}^2)). \tag{3.18}$$

Ismét felhasználva a (2.3) és (3.7) hibaformulákat, kapjuk, hogy

$$B_1 \approx e_{i-1}e_{i-2}e_{i-3}(c_1^2 + O(e_{i-3})). \tag{3.19}$$

Tehát

$$e_{i+1} \approx \frac{c_2}{c_1} e_i e_{i-1}. \tag{3.20}$$

Állapítsuk most meg az U, $M2$, $M1$ iterációsorozat hibaképletét. Az U, $M1$, $M1$ iterációsorozat iterációs képletével összehasonlítva megállapíthatjuk, hogy a két iterációs képlet indexcserével egymásba átvihető, sőt a levezetés közben felhasználandó (2.3) és (3.13) hibaképletekből ugyanazon indexcserével jutunk el a (2.3) és (3.7) hibaképlethez, amiket az előző levezetésben már felhasználtunk. Emiatt a levezetést nem kell megismételni, csak az előző eredményben kell a megfelelő indexcserét végrehajtani, ami most hatástalan marad, mert (3.20)-ban a kritikus indexek nem is szerepelnek.

Mielőtt hosszabb iterációsorozatok hibaanalízisét elvégeznénk, előjelvizsgálatot végzünk, előfordulhat-e a fentieknél hosszabb, M típusú lépés után U típusú lépést nem tartalmazó iterációsorozat? Ennek megállapításában hasznos segítőtársunk lesz az előjeltáblázat, mely azon alapszik, hogy $f_i e_i$ előjele minden i-re állandó, tehát a függvényértékek előjelvizsgálatát visszavezethetjük a hibasorozat előjelvizsgálatára. Hibaképleteink, vagyis (2.3), (3.7), (3.13) és (3.20) alapján tudjuk, hogy e_{i+1} előjele a korábbi hibáknak és a c_2/c_1, illetve c_3/c_1 együtthatóknak az előjelétől függ. Ennek megfelelően az előjeltáblázatnak 4 blokkja van, hogy a kétfajta együttható összes előjelkombinációját megvizsgálhasssuk. Az előjelvizsgálat blokkonként két ágon fut a kezdeti hibák előjelkombinációinak megfelelően. Egy ágon addig vizsgálom az új közelítések előjeleit, míg két egymást követő új közelítés hibája különböző előjelű nem lesz, ezután U típusú lépés következik,

és a blokkban az előjelvizsgálat az egyik ág elején folytatódhat, amit nyíllal jelölünk. Minket azon ágak érdekelnek elsősorban, ahol ciklus alakul ki, mert az egy ciklusba eső iterációs lépések sorozatát egy makrolépésnek tekintve, már homogén lépéssorozatot kapunk, amelynek rendje és hatékonysága meghatározható.

Az előjeltáblázatból mindenekelőtt leolvasható, hogy a már megvizsgáltaknál hosszabb iterációsorozatok valóban nem fordulhatnak elő, a hibaanalízist nem kell tovább folytatni. Emellett minden blokkban egy olyan ciklust kapunk, amelybe a másik ág is bekapcsolódik. Érdekes, hogy mindkét módszernél kétfajta ciklus különböztethető meg, amelyek bekövetkezése csak c_3/c_1 előjelétől függ.

4. A Pegazus módszerek hatékonyságvizsgálata

Az előjeltáblázatból látszik, hogy a *Pegazus módszernél* U, U, $M1$, ill. U, U, $M1$, $M1$ típusú iterációs ciklusok fordulnak elő. A hibaképleteket felhasználva kapjuk, hogy

$$(4.1) \qquad U \qquad e_{i+1} \approx e_i e_{i-1} \frac{c_2}{c_1},$$

$$(4.2) \qquad U, U \qquad e_{i+2} \approx e_{i+1} e_i \frac{c_2}{c_1} \approx e_i^2 e_{i-1} \left(\frac{c_2}{c_1}\right)^2,$$

$$(4.3) \qquad U, U, M1 \qquad e_{i+3} \approx e_{i+2} e_i^2 \frac{c_3}{c_1} \approx e_i^4 e_{i-1} \frac{c_3}{c_1} \left(\frac{c_2}{c_1}\right)^2.$$

Tehát

$$(4.4) \qquad e_{i+3} \approx e_i^4 e_{i-3}^2 e_{i-6}^2 \dots e_{i-3k}^2 \frac{c_3}{c_1} \left(\frac{c_2}{c_1}\right)^{2(k+1)} e_{i-3k-1}.$$

Legyen $d_{i-j} = e_{i-3j}$, akkor

$$(4.5) \qquad d_{i+1} \approx d_i^4 d_{i-1}^2 \dots d_{i-k}^2 \frac{c_3}{c_1} \left(\frac{c_2}{c_1}\right)^{2(k+1)} e_{i-3k-1}.$$

A hibaképlethez rendelt karakterisztikus egyenlet:

$$(4.6) \qquad \begin{aligned} x^{k+1} - 4x^k - 2(x^{k-1} + x^{k-2} + \dots + 1) &= 0, \\ x^{k+1} - 4x^k - 2(x^k - 1)/(x-1) &= 0, \\ (x^{k+2} - 5x^{k+1} + 2x^k + 2)/(x-1) &= 0. \end{aligned}$$

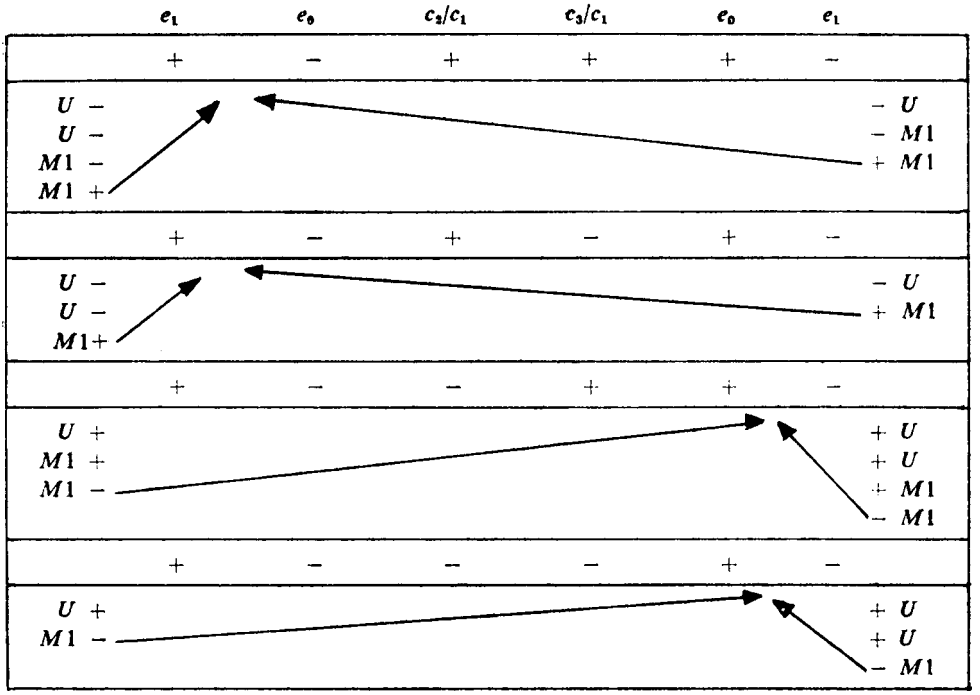
A számláló gyökét keressük. $x > 1$-re és elég nagy k-ra (elég sok iterációs lépés után) a konstans tag aszimptotikusan elhanyagolható:

$$(4.7) \qquad x^2 - 5x + 2 = 0, \quad p = (5 + \sqrt{17})/2.$$

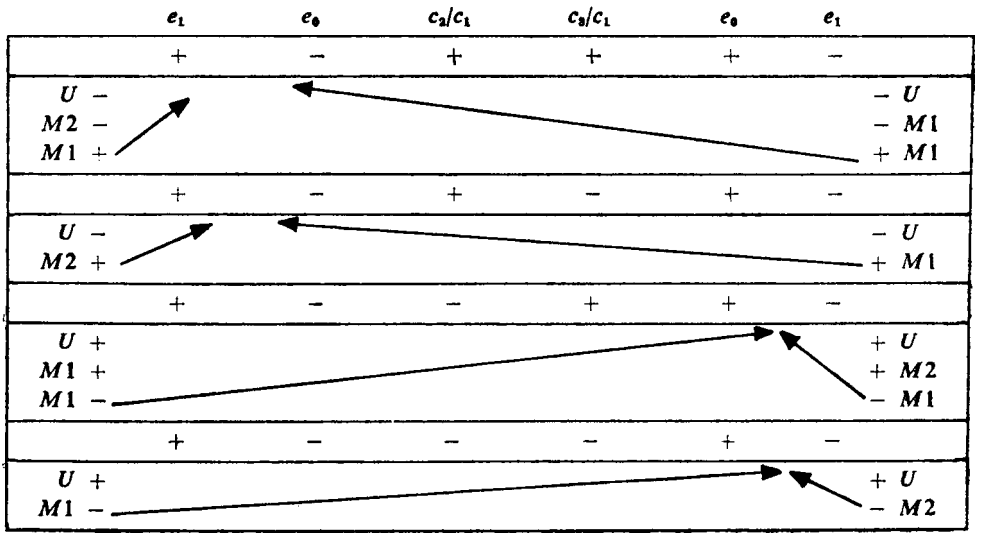
Tehát ebben az esetben a módszer rendje 4,56, numerikus hatékonysága pedig $\sqrt[3]{4{,}56} = 1{,}658$.

Előjeltáblázat

Pegazus-módszer



Módosított Pegazus-módszer



Az $U, U, M1, M1$ típusú iterációsorozat esetén:

$$U, U, M1, M1 \quad e_{i+4} \approx e_{i+3}e_{i+2}\frac{c_2}{c_1} \approx e_i^6 e_{i-1}^2 \frac{c_3}{c_1}\left(\frac{c_3}{c_1}\right)^5. \tag{4.8}$$

Tehát

$$e_{i+4} \approx e_i^6 e_{i-4}^3 e_{i-8}^3 \dots e_{i-4k}^3 \left(\frac{c_3}{c_1}\right)^{2k+1}\left(\frac{c_2}{c_1}\right)^{4k+5} e_{i-4k-1}. \tag{4.9}$$

Legyen $d_{i-j}=e_{i-4j}$, akkor:

$$d_{i+1} \approx d_i^6 d_{i-1}^3 \dots d_{i-k}^3 \left(\frac{c_3}{c_1}\right)^{2k+1}\left(\frac{c_2}{c_1}\right)^{4k+5} e_{i-4k-1}. \tag{4.10}$$

A hibaképlethez rendelt karakterisztikus egyenlet:

$$\begin{aligned} x^{k+1}-6x^k-8(x^{k-1}+\dots+1) &= 0, \\ (x^{k+2}-7x^{k+1}-2x^k+8)/(x-1) &= 0. \end{aligned} \tag{4.11}$$

A fentiekhez hasonló módon eljárva, a redukált egyenlet:

$$x^2-7x-2=0, \quad p=(7+\sqrt{57})/2. \tag{4.12}$$

Tehát a módszer rendje most 7,275, numerikus hatékonysága pedig $\sqrt[4]{7{,}275}=1{,}64$.

Szintén az előjeltáblázatból látható, hogy a *módosított Pegazus módszernél* $U, M2$, ill. $U, M2, M1$ típusú iterációs ciklusok fordulhatnak elő. A (2.31), (3.13) és (3.20) hibaképleteket felhasználva kapjuk:

$$U \qquad e_{i+1} \approx e_i e_{i-1}\frac{c_2}{c_1}, \tag{4.13}$$

$$U, M2 \quad e_{i+2} \approx e_{i+1}e_i^2\frac{c_3}{c_1} \approx e_i^3 e_{i-1}\frac{c_3}{c_1}\frac{c_2}{c_1}. \tag{4.14}$$

Tehát

$$e_{i+2} \approx e_i^3 e_{i-2}e_{i-4} \dots e_{i-2k}\frac{c_3}{c_1}\left(\frac{c_2}{c_1}\right)^{k+1} e_{i-2k-1}. \tag{4.15}$$

Legyen $d_{i-j}=e_{i-2j}$, akkor:

$$d_{i+1} \approx d_i^3 d_{i-1} \dots d_{i-k}\frac{c_3}{c_1}\left(\frac{c_2}{c_1}\right)^{k+1} e_{i-2k-1}. \tag{4.16}$$

A hibaképlethez rendelt karakterisztikus egyenlet:

$$\begin{aligned} x^{k+1}-3x^k-(x^{k-1}+\dots+1) &= 0, \\ (x^{k+2}-4x^{k+1}+2x^k+1)/(x-1) &= 0. \end{aligned} \tag{4.17}$$

A redukált egyenlet:

$$x^2-4x+2=0, \quad p=2+\sqrt{2}. \tag{4.18}$$

Tehát a módszer rendje most 3,414, hatékonysága $\sqrt{3{,}414}=1{,}84$.

Az $U, M2, M1$ típusú iterációs ciklus esetén:

$$U, M2, M1 \quad e_{i+3} \approx e_{i+2} e_{i+1} \frac{c_2}{c_1} \approx e_i^4 e_{i-1}^2 \frac{c_3}{c_1} \left(\frac{c_2}{c_1}\right)^3. \tag{4.19}$$

Tehát

$$e_{i+3} \approx e_i^4 e_{i-3}^6 e_{i-6}^6 \dots e_{i-3k}^6 \left(\frac{c_3}{c_1}\right)^{2k+1} \left(\frac{c_2}{c_1}\right)^{2k+3} e_{i-3k-1}. \tag{4.20}$$

Legyen $d_{i-j} = e_{i-3j}$, akkor:

$$d_{i+1} \approx d_i^4 d_{i-1}^6 \dots d_{i-k}^6 \left(\frac{c_3}{c_1}\right)^{2k+1} \left(\frac{c_2}{c_1}\right)^{2k+3} e_{i-3k-1}. \tag{4.21}$$

A hibaképlethez rendelt karakterisztikus egyenlet:

$$\begin{aligned} x^{k+1} - 4x^k - 6(x^{k-1} + \dots + 1) &= 0 \\ (x^{k+2} - 5x^{k+1} - 2x^k + 6)/(x-1) &= 0. \end{aligned} \tag{4.22}$$

A redukált egyenlet:

$$x^2 - 5x - 2 = 0, \quad p = (5 + \sqrt{33})/2. \tag{4.23}$$

Tehát a módszer rendje most 5,37, hatékonysága pedig $\sqrt[3]{5{,}37} = 1{,}75$.

Visszatérve a BIT-ben megjelent publikációkra, az eredeti *Pegazus módszernél* ismertetett hibaképletek eltérnek eredményeimtől, ugyanis az $U, M1$ iterációsorozat hibaképlete ott tévesen

$$e_{i+1} \approx \frac{c_2}{c_1} e_i e_{i-2}^2. \tag{4.24}$$

Tehát az aszimptotikus hibakonstansban van eltérés, az $U, U, M1, M1$ iterációsorozat hibaképletében pedig az indexekben van eltérés az általam levezetett (3.20) hibaképlettől:

$$e_{i+1} \approx \frac{c_2}{c_1} e_i e_{i-2}. \tag{4.25}$$

Ezen hibaképletek mellett a lehetséges iterációs ciklusok kiválasztása leegyszerűsödött (hiszen csak egyetlen együtthatót kellett vizsgálniok), és eredményül az adódott, hogy csak az $U, U, M1, M1$ iterációs ciklus lehetséges. Ezen iterációs ciklus mellett a rend, ill. numerikus hatékonyság érdekes módon megegyezik az általam levezetettekkel, annak ellenére, hogy a megfelelő hibaképletek különböztek. Viszont az általam megállapított másik, $U, U, M1$ iterációs ciklus az eredeti publikáció szerint nem fordulhat elő.

A *módosított Pegazus módszert* bemutató cikk már felhasználta az előző publikáció eredményeit, és a vázolt hibaanalízis későbbi eredményei valószínűleg emiatt különböznek az általam levezetett hibaképletektől. A cikk szerint lehetséges iterációs ciklusok is eltérnek a nálam megállapítottaktól — az eredeti publikáció egyik ciklusa mindenképpen hibás, mivel két U típusú lépés követi egymást benne. Ezután már az sem meglepő, hogy a számolt rendek alacsonyabbak, mint a most megállapítottak.

5. A számítógépes teszt kiértékelése

A közölt módszereket összehasonlító teszt segítségével ki is próbáltam, és bebizonyosodott, hogy az elméletileg számolt rendeknek megfelelően a *Pegazus módszerek* még a szelőmódszernél is gyorsabb konvergenciát biztosítanak gyakorlatilag azonos számításigény mellett (iterációs lépésenként legfeljebb 1—1 szorzással, osztással, ill. összeadással kell többet végrehajtani).

Mivel az iteráció az egyes tesztfüggvények esetén általában 5—10 lépés alatt eljutott a gyökhöz, ezért az elméletileg megadott ciklusok ilyen rövid lépésszám mellett nem alakultak ki (de általában az elméletileg kimutatott iterációsorozatok követték egymást). Ellenben ha az iteráció 20—30 lépéses volt, ez azt jelentette, hogy a kezdőbecsléseket a gyököktől túl távol, vagy más, „szerencsétlen" módon választottam meg, ami miatt adott esetben a hibaképletben egyébként elhanyagolt tagok domináns szerepet játszanak. Például az $x^{10}-1=0$ egyenlet esetén az iterációt a 0,5; 2 kezdőbecslésből indítva az iteráció első U típusú lépését 8 darab $M1$ típusú lépés követte, mert nem voltunk elég közel a gyökhöz (a függvény $x>1$ esetén nagyon gyorsan nő). A *Pegazus módszer* a fenti esetben 2 darab U, U, $M1$, $M1$ iterációsorozattal fejeződött be, míg a *módosított Pegazus módszert* alkalmazva a fenti probléma megoldására az iteráció 3 darab U, $M2$ iterációsorozattal zárult. Tehát a gyök közelében a ciklus kialakulása már észlelhető.

Számítási eredmények az $x^{10}-1=0$ *egyenletre*

Pegazus módszer			*Módosított Pegazus módszer*		
U	1	0,501 463 414 684 292 60	U	1	0,501 463 414 684 292 60
U	2	0,502 925 359 296 471 60	$M2$	2	0,504 384 411 830 016 80
$M1$	3	0,505 843 418 829 366 30	$M1$	3	0,510 203 152 174 673 08
$M1$	4	0,511 656 295 243 992 71	$M1$	4	0,521 748 203 406 812 29
$M1$	5	0,523 189 658 347 014 71	$M1$	5	0,544 473 112 401 761 91
$M1$	6	0,545 891 329 007 995 70	$M1$	6	0,588 495 690 178 717 99
$M1$	7	0,589 867 763 420 812 42	$M1$	7	0,671 055 728 005 294 80
$M1$	8	0,672 334 930 954 176 81	$M1$	8	0,815 267 628 225 299 17
$M1$	9	0,816 340 110 799 975 49	$M1$	9	1,015 707 980 340 703 39
$M1$	10	1,016 117 388 777 518 94	U	10	0,983 168 632 569 569 49
U	11	0,982 874 317 331 300 45	$M2$	11	1,000 151 581 701 450 16
U	12	0,998 758 124 273 488 90	U	12	0,999 988 165 337 880 13
$M1$	13	0,999 995 825 272 022 44	$M2$	13	1,000 000 000 245 350 74
$M1$	14	1,000 000 022 243 434 33	U	14	1,000 000 000 099 986 84
U	15	1,000 000 000 099 580 29	$M2$	15	1,000 000 000 000 000 00
U	16	1,000 000 000 099 999 90	$M1$	16	1,000 000 000 000 000 00
$M1$	17	1,000 000 000 000 000 00			
$M1$	18	1,000 000 000 000 000 00			

IRODALOM

[1] DOWEL, M. and JARRAT, P., "The 'Pegasus' method for computing the root of an equation", *BIT* **12** (1972) 503—508.
[2] KING, R. F., "An improved Pegasus method for root finding", *BIT* **13** (1973) 423—427.
[3] TRAUB, J. F., *Iterative Methods for the Solution of Equations* (Englewood Cliffs, N. J., 1964).

(Beérkezett: 1979. június 28.)

KALMÁR JÁNOS
MTA GEODÉZIAI ÉS GEOFIZIKAI KUTATÓ INTÉZET
9400 SOPRON, MÚZEUM U. 6—8.

THE PEGASUS METHODS FOR THE SOLUTION OF NONLINEAR EQUATIONS

J. KALMÁR

The paper after introducing in the topic of iterative root-determining methods presents the up-to-date versions of chord methods and some new results concerning their efficiency. The experiences computer applications are also discussed.

Alkalmazott Matematikai Lapok 5 *(1979)* 289—293

SZIMMETRIKUS VÉLETLEN (0, 1) MÁTRIX SPEKTRUMÁRÓL

JUHÁSZ FERENC
Budapest

A dolgozatban megmutatjuk, hogy egy szimmetrikus véletlen (0, 1) mátrix legnagyobb sajátértéke n rendű, míg a többi sajátérték tetszőleges $\varepsilon>0$ esetén $o(n^{\frac{1}{2}+\varepsilon})$ rendű mértékben. A [7] dolgozat a jelen munka egy korábbi változata.

1. Bevezetés

Véletlen mátrixok spektrumának határeloszlása az ún. *Wigner-féle félkör törvény* [8]. Ennek ismeretében állíthatjuk, hogy a sajátértékek tetszőlegesen nagy hányada $\sqrt{n}$ rendű.

Ebben a dolgozatban megmutatjuk, hogy egy szimmetrikus véletlen (0, 1) mátrixnak egyetlen n rendű sajátértéke van, a többi sajátérték ettől élesen elválik: $o(n^{\frac{1}{2}+\varepsilon})$ rendű.

2. A legnagyobb sajátérték vizsgálata

2.1. Tétel. Legyen $\mathbf{A}=(a_{ij})$ olyan $n\times n$ méretű szimmetrikus (0, 1) mátrix, amelyben az egyesek sűrűsége d ($\mathbf{A}$ dn^2 egyest tartalmaz). Ekkor az $\mathbf{A}$ mátrix λ_1 legnagyobb sajátértékére

$$dn \leqq \lambda_1 \leqq \sqrt{d}\, n.$$

Bizonyítás. Legyen $\mathbf{e}_n=(1, \ldots, 1)$ az n dimenziós csupa egyesből álló vektor. Ekkor az $\mathbf{A}$ mátrix λ_i, $i=1, \ldots, n$ sajátértékeire

$$\lambda_i \geqq \frac{(\mathbf{e}_n, \mathbf{A}\mathbf{e}_n)}{(\mathbf{e}_n, \mathbf{e}_n)} = \frac{dn^2}{n}, \quad i = 1, \ldots, n$$

és

$$\sum_{i=1}^{n} \lambda_i^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}^2 = dn^2.$$

Ezekből következik a tétel állítása.

1. Következmény. Ha az $\mathbf{A}$ mátrix mérete végtelenhez tart oly módon, hogy az egyesek sűrűsége $d\geqq K>0$; akkor a legnagyobb sajátérték rendje n.

2.2. Tétel. Legyen az $\mathbf{A}_n=(a_{ij})$ olyan $n\times n$ méretű szimmetrikus mátrix, amelynek elemei $i>j$ esetén teljesen független valószínűségi változók, $P(a_{ij}=1)=p$, $P(a_{ij}=0)=q=1-p$. Tegyük fel, hogy a főátlóban $a_{ii}\equiv 0$. Ha $\lambda_1=\lambda_1(n)$ az $\mathbf{A}_n$ legnagyobb sajátértéke, akkor

$$\lim_{n\to\infty}\frac{\lambda_1}{n}=p \text{ mértékben.}$$

Bizonyítás. A *Perron—Frobenius-tétel* [4] alapján

$$\min_{1\leqq i\leqq n}\sum_{j=1}^{n}a_{ij}\leqq\lambda_1\leqq\max_{1\leqq i\leqq n}\sum_{j=1}^{n}a_{ij}.$$

A *Csebisev-egyenlőtlenség* éles alakja [5] szerint van olyan $K>0$ szám, hogy minden i indexre

$$P\left(\left|\frac{1}{n}\sum_{j=1}a_{ij}-p\right|\geqq\delta\right)\leqq\exp(-Kn).$$

Innen

$$P\left(\max_{1\leqq i\leqq n}\left|\frac{1}{n}\sum_{j=1}^{n}a_{ij}-p\right|\geqq\delta\right)\leqq nP\left(\left|\frac{1}{n}\sum a_{ij}-p\right|\geqq\delta\right)\leqq n\exp(-Kn)=o(1).$$

3. A legnagyobbtól különböző sajátértékek vizsgálata

A *Wigner-tétel* [8] LUDWIG ARNOLDtól származó általánosítása [2, 3] a sajátértékek eloszlását jellemzi.

3.1. Tétel (Wigner) [2, 3]. Legyen az $\mathbf{A}_n=(a_{ij})$ olyan $n\times n$ méretű szimmetrikus mátrix, amelynek elemei $i\geqq j$ esetén teljesen független valószínűségi változók. Tegyük fel, hogy a főátlóban a_{ii} eloszlásfüggvénye G, a_{ij} eloszlásfüggvénye $i>j$ esetén H, amelyre $\int x^2dH(x)=\sigma^2<\infty$. Legyen az $\frac{1}{\sqrt{n}}\mathbf{A}_n$ mátrix sajátértékeinek tapasztalati eloszlásfüggvénye

$$F_n(x)=\frac{1}{n}\sum_{\lambda_i<x\sqrt{n}}1.$$

Ekkor tetszőleges x esetén

$$\lim_{n\to\infty}F_n(x)=\int_{-\infty}^{x}f(x)\,dx \text{ mértékben,}$$

ahol

$$f(x)=\begin{cases}\frac{1}{2\pi\sigma^2}\sqrt{4\sigma^2-x^2}, & \text{ha } |x|<2\sigma,\\ 0, & \text{különben.}\end{cases}$$

További általánosítások találhatók a [6] könyvben. A tétel megengedi, hogy valahány ($o(n)$ számú) sajátérték n rendű legyen. Megmutatjuk, hogy valójában nem ez

a helyzet: a legnagyobbtól különböző sajátértékek közel $\sqrt{n}$ rendben növekszenek csak. Ez történik minden szimmetrikus véletlen mátrixszal, ahol az elemek várható értéke nem nulla.

3.1. Definíció. Az $f(n)$ valószínűségi változó sorozatról azt mondjuk, hogy $O(n^\alpha)$ mértékben, ha minden $\delta>0$ számhoz van olyan n_0 és K szám, hogy $n>n_0$ esetén $P\left(\left|\frac{f(n)}{n^\alpha}\right|>K\right)<\delta$.

3.2. Definíció. Az $f(n)$ valószínűségi változó sorozatról azt mondjuk, hogy $o(n^\alpha)$ mértékben, ha minden $\delta>0$ és $K>0$ számhoz van olyan n_0 szám, hogy $n>n_0$ esetén $P\left(\left|\frac{f(n)}{n^\alpha}\right|>K\right)<\delta$.

3.2. Tétel. Legyen az $\mathbf{A}_n=(a_{ij})$ mátrix olyan, mint a 2.2. tételben. Ha $\lambda_2=\lambda_2(n)$ az $\mathbf{A}_n$ második legnagyobb sajátértéke, akkor tetszőleges $\varepsilon>0$ esetén $\lambda_2=o\left(n^{\frac{1}{2}+\varepsilon}\right)$ mértékben.

Bizonyítás. Legyen P_n az $\mathbf{e}_n=(1,\ldots,1)$ vektorral párhuzamos merőleges vetítés. A *Courant—Fischer-tétel* [4] szerint az $\mathbf{A}_n$ mátrix második legnagyobb sajátértékére

$$\lambda_2=\min_{(y,y)=1}\max_{(x,y)=0}\frac{(\mathbf{x},\mathbf{Ax})}{(\mathbf{x},\mathbf{x})}\leqq\|\mathbf{P}_n\mathbf{A}_n\mathbf{P}_n\|,$$

ahol $\|\cdot\|$ jelöli a mátrix maximális sajátértékét. Minthogy a $\mathbf{P}_n=(p_{ij})$ mátrixra $p_{ij}=\delta_{ij}-\frac{1}{n}$ (δ_{ij} a *Kronecker-féle szimbólum*), ezért könnyű látni, hogy a $\mathbf{P}_n\mathbf{A}_n\mathbf{P}_n=$ $=\mathbf{B}_n=(b_{ij})$ mátrixra $b_{ij}=a_{ij}-d_i-d_j+d$, ahol d_i és d jelöli az egyesek sűrűségét az i indexű sorban, illetve az egész mátrixban. Írjuk mátrixunkat a következő alakba: $\mathbf{B}_n=\mathbf{C}_n+\mathbf{R}_n$, ahol $\mathbf{C}_n=(c_{ij})$, $c_{ij}=a_{ij}-p$ és $\mathbf{R}_n=(r_{ij})$, $r_{ij}=d-d_j-$ $-d-p$. Ekkor

$$\|\mathbf{B}_n\|\leqq\|\mathbf{C}_n\|+\|\mathbf{R}_n\|.$$

Legyen $\|\mathbf{R}_n\|_2^2=\sum_{i=1}^{n}\sum_{j=1}^{n}r_{ij}^2$. Minthogy

$$d_i+d_j-d=\sum_{k=1}^{n}\sum_{l=k+1}^{n}h_{kl}a_{kl},$$

ahol

$$h_{kl}=\begin{cases}-\frac{1}{n^2}, & \text{ha } k\neq i,j \text{ és } l\neq i,j,\\ \frac{2}{n}-\frac{1}{n^2}, & \text{ha } k=i \text{ és } l=j,\\ \frac{1}{n}-\frac{1}{n^2}, & \text{különben,}\end{cases}$$

ezért, ha $D(\cdot)$ jelöli a szórást, akkor

$$D^2(d_i+d_j-d) \leqq pq\left(n^2\frac{1}{n^4}+\left(\frac{2}{n}-\frac{1}{n^2}\right)^2+2n\left(\frac{1}{n}-\frac{1}{n^2}\right)^2\right) \leqq \frac{3pq}{n}.$$

Innen a várható értékre

$$E\|\mathbf{R}_n\|_2^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} Er_{ij}^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} D^2(d_i+d_j-d) \leqq n^2\frac{3pq}{n},$$

azaz $E\|\mathbf{R}_n\|_2 \leqq \sqrt{3pqn}$.

A *Markov-egyenlőtlenség* [5] szerint

$$P(\|\mathbf{R}_n\| \geqq L\sqrt{3pqn}) \leqq P(\|\mathbf{R}_n\|_2 \geqq L\sqrt{3pqn}) \leqq \frac{1}{L},$$

ahol $\|\mathbf{R}_n\|$ jelöli az $\mathbf{R}_n$ mátrix maximális sajátértékét.

A $\|\mathbf{C}_n\|$ mennyiség becslésére szolgál a következő tétel.

3.3. Tétel. Legyen $\mathbf{A}_n=(a_{ij})$ olyan mint a *Wigner-tételben* és tegyük fel, hogy az elemek várható értéke nulla: $\int x dH(x)=0$. Ha ezenkívül H összes momentuma véges, akkor tetszőleges $\varepsilon>0$ esetén az $\mathbf{A}_n$ mátrix $\lambda=\lambda(n)$ maximális sajátértékére $\lim_{n\to\infty} P(|\lambda|\geqq n^{\frac{1}{2}+\varepsilon})=0$, azaz $\lambda=o\left(n^{\frac{1}{2}+\varepsilon}\right)$ mértékben.

Bizonyítás. A [8] és [1] dolgozatban olvasható, hogy a várható értékre az $E\lambda^{2l}\leqq E\sum_{i=1}^{n}\lambda_i^{2l}=O(n^{l+1})$ egyenlőtlenség teljesül. Ebből a *Markov-egyenlőtlenség* szerint

$$P(|\lambda| \geqq n^{\frac{1}{2}+\varepsilon}) \leqq n^{-(l+2l\varepsilon)}E\lambda^{2l} = O(n^{-(2l\varepsilon-1)}) = o(1),$$

ha l elég nagy volt.

1. Megjegyzés. Ha a 3.3. tétel feltételei közül a magasabb momentumok létezését elhagyjuk a $\lambda=o(n)$ mértékben állítás még teljesül.

3.4. Tétel. Legyen az $\mathbf{A}_n=(a_{ij})$ mátrix olyan, mint a 2.2. tételben. Ha $\lambda_n=\lambda_n(n)$ az $\mathbf{A}_n$ legkisebb sajátértéke, akkor tetszőleges $\varepsilon>0$ esetén $\lambda_n=o\left(n^{\frac{1}{2}+\varepsilon}\right)$ mértékben.

Bizonyítás: Írjunk az $\mathbf{e}_n=(1,\ldots,1)$ vektor, mint tengely köré egy $\omega<\frac{\pi}{4}$ szögű kettős kúpot. E kúpon kívül helyezkedik el az $\mathbf{A}_n$ mátrixnak legalább $(n-1)$ sajátvektora. Legyen $\mathbf{v}=(v_i)$ egy ilyen sajátvektor, μ a hozzá tartozó sajátérték, $\|\mathbf{v}\|^2=\sum_{i=1}^{n} v_i^2=1$. Bontsuk fel a $\mathbf{v}$ vektort a következő módon: $\mathbf{v}=c\mathbf{e}_n+\mathbf{P}_n\mathbf{v}$, ahol $\mathbf{P}_n$ az $\mathbf{e}_n$ vektorral párhuzamos merőleges vetítés. Mivel $\mathbf{v}$ a kúpon kívül helyezkedik el, ezért $c=O\left(\frac{1}{\sqrt{n}}\right)$.

Az $\mathbf{A}_n\mathbf{e}_n$ vektornak az $\mathbf{e}_n$ irányú egyenestől való távolsága $\|\mathbf{P}_n\mathbf{A}_n\mathbf{e}_n\|$, ezért $\|\mathbf{P}_n\mathbf{A}_n\mathbf{e}_n\| = \|\mathbf{A}_n\mathbf{e}_n - np\mathbf{e}_n\|$. A várható értékre

$$E\|\mathbf{A}_n\mathbf{e}_n - np\mathbf{e}_n\|^2 \leqq \sum_{i=1}^{n} D^2\left(\sum_{j=1}^{n} a_{ij}\right) = pqn^2,$$

azaz $E\|\mathbf{A}_n\mathbf{e}_n - np\mathbf{e}_n\| \leqq \sqrt{pq}\,n$. A *Markov-egyenlőtlenség* szerint

$$P\left(\|\mathbf{P}_n\mathbf{A}_n\mathbf{e}_n\| \geqq L\sqrt{pq}\,n\right) \leqq P\left(\|\mathbf{A}_n\mathbf{e}_n - np\mathbf{e}_n\| \geqq L\sqrt{pq}\,n\right) \leqq \frac{1}{L}.$$

Ezért felhasználva, hogy $\|\mathbf{P}_n\mathbf{A}_n\mathbf{P}_n\mathbf{v}\| = o\left(n^{\frac{1}{2}+\varepsilon}\right)$ mértékben,

$$\|\mathbf{P}_n\mathbf{A}_n\mathbf{v}\| \leqq |c|\,\|\mathbf{P}_n\mathbf{A}_n\mathbf{e}_n\| + \|\mathbf{P}_n\mathbf{A}_n\mathbf{P}_n\mathbf{v}\| = o\left(n^{\frac{1}{2}+\varepsilon}\right) \quad \text{mértékben.}$$

Innen

$$|\mu| = \frac{\|\mathbf{A}_n\mathbf{v}\|}{\|\mathbf{v}\|} = \frac{\|\mathbf{P}_n\mathbf{A}_n\mathbf{v}\|}{\|\mathbf{P}_n\mathbf{v}\|} \leqq \frac{\|\mathbf{P}_n\mathbf{A}_n\mathbf{v}\|}{\cos\omega} = o\left(n^{\frac{1}{2}+\varepsilon}\right) \quad \text{mértékben.}$$

Minthogy ez legalább $(n-1)$ sajátértékre igaz, következésképpen a legkisebbre is.

IRODALOM

[1] ARNOLD, L., "On the asymptotic distribution of the eigenvalues of random matrices", *J. Math. Analysis Appl.* **20** (1967) 262—268.

[2] ARNOLD, L., "On Wigner's semicircle law for the eigenvalues of random matrices", *Z. Wahrscheinlichkeitstheorie verw. Geb.* **19** (1971) 191—198.

[3] ARNOLD, L., "Deterministic version of Wigner's semicircle law for the distribution of matrix eigenvalues", *Linear Algebra and its Appl.* **13** (1976) 185—199.

[4] BELLMAN, R., *Introduction to Matrix Analysis* (McGraw-Hill Book Company, Inc., New York—Toronto—London, 1960).

[5] FELLER, W., *An Introduction to Probability Theory and its Applications* (Vol. I., John Wiley and Sons, New York, 1957).

[6] GIRKO, V. L., *Véletlen mátrixok* (Vüsa Skola, Kijev, 1975) oroszul.

[7] JUHÁSZ, F., "On the Spectrum of a Random Graph", Proceedings of the Colloquium of Algebraic Methods in Graph Theory, Szeged, 1978.

[8] WIGNER, E. P., "Characteristic vectors of bordered matrices with infinite dimensions", *Ann. of Math.* **62** (1955) 548—564.

(Beérkezett: 1979. július 6.)

JUHÁSZ FERENC
MTA SZÁMÍTÁSTECHNIKAI ÉS AUTOMATIZÁLÁSI KUTATÓ INTÉZET
1250 BUDAPEST, URI U. 49.

ON THE SPECTRUM OF A SYMMETRIC RANDOM (0, 1) MATRIX

F. JUHÁSZ

In the present paper we deal with the asymptotic behaviour of the spectrum of a symmetric random (0, 1) matrix. It will be shown that the largest eigenvalue is of order n, and there is a sudden drop: the other eigenvalues are $o\left(n^{\frac{1}{2}+\varepsilon}\right)$. The paper [7] is an earlier version of this one.

Alkalmazott Matematikai Lapok **5** *(1979)* 295—302

FERDE ELOSZLÁSÚ ADATSOROK SZIMMETRIKUSSÁ TÉTELE HATVÁNYOZÁSSAL

RÉTHÁTI LÁSZLÓ
Budapest

A természetes adat- és idősorok aszimmetriája megnehezíti a különböző valószínűséghez tartozó értékek számítását és az események előrejelzését. Mivel a normális eloszlás feltételezésével jelentős hibát követhetünk el, a különböző típusú eloszlásfüggvényekre végzett illeszkedésvizsgálatok pedig nagyon időigényesek, az alkalmazott matematika területén régi törekvés az ezen hátrányokat kiküszöbölő eljárások kutatása. A szerző által kidolgozott módszer az adatsorok olyan transzformációján alapul, amely $\beta=0$ ferdeségi együtthatóval rendelkező sor előállítását teszi lehetővé.

1. Bevezetés

A fizikai (műszaki) tartalmú paraméterek mérése vagy megfigyelése során kapott adat- és idősorok az esetek többségében aszimmetrikusak. Ennek oka általában az, hogy a vizsgált mennyiségnek — alsó vagy felső — korlátja van: a havi csapadékösszeg nem lehet negatív, a víz alatt fekvő porózus közeg telítettsége nem lehet 100%-nál nagyobb stb.

A fizikai tartalmú sorok előállításának célja az anyag (vagy folyamat) jellemzése, a szerkezetek méretezése, illetve az előrejelzés. Ezek a feladatok könnyen megoldhatók, ha a rendelkezésünkre álló sorozat normális eloszlású. Ellenkező esetben két utat követhetünk: *a*) meghatározzuk az adott sorra legjobban illeszkedő eloszlástípust, vagy *b*) valamely erre alkalmas módszerrel[1] transzformációt hajtunk végre. A következőkben ismertetett eljárás az utóbbi megoldási lehetőségek egyik változata, amelynek az a lényege, hogy az új valószínűségi változó (η) az eredeti ξ változónak valamely — az általunk megszabott követelményt kielégítő — hatványa ($\eta=\xi^a$).

2. A transzformáció hatásának általános vizsgálata

Az aszimmetria mértékéül válasszuk a

$$\beta=\frac{n\sqrt{n-1}\sum_{i=1}^{n}(\eta_i-\bar{\eta})^3}{(n-2)\left[\sum_{i=1}^{n}(\eta_i-\bar{\eta})^2\right]^{3/2}} \tag{2.1}$$

[1] Például BETHLAHMY közelítő grafikus eljárásával *(Proc. of the ASCE,* **103** (1977)*)*.

kifejezéssel definiált ferdeségi együtthatót.

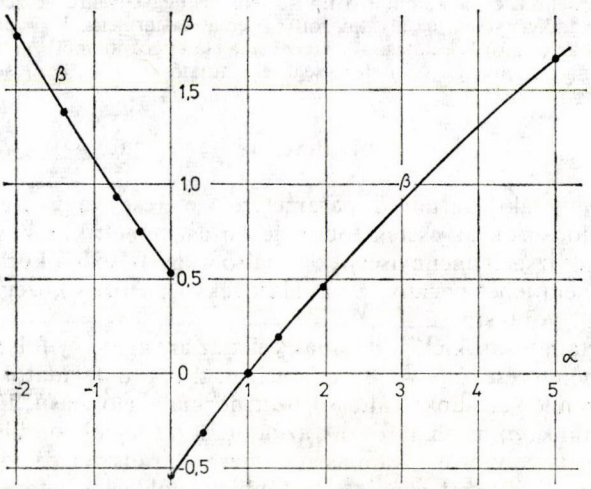
Mivel az $\eta=\xi^{\alpha}$ transzformáció célja olyan adatsor előállítása, amelyre $\beta=0$, első lépésként tájékozódnunk kell a (2.1) alatti kifejezés tulajdonságairól. Az ennek érdekében végzett vizsgálatok egyik eredményét mutatja be az 1. ábra, amelyet úgy nyertünk, hogy egy normális eloszlású diszkrét sor tagjait különböző α hatványokra emeltük. Az itt látható szakadásos függvény két pontját ismerjük:

a) mivel a sor eredetileg szimmetrikus volt, az $\alpha=1$ helyen $\beta=0$;

b) ha $\alpha\to 0$, a

$$\lim_{\alpha\to 0}\frac{x^{\alpha}-1}{\alpha}\to \ln x \quad (x>0)$$

határérték-tétel segítségével bizonyítható, hogy $|\beta|\to|\beta_L|$, ahol β_L az $\eta=\ln\xi$ transzformációhoz tartozó ferdeségi együttható.



1. ábra

A β ferdeségi együttható α függvényében, egy mesterséges normális sorra

Figyelembe véve a görbe alakját, valamint az előző bekezdésben említetteket, azzal a heurisztikus feltételezéssel élhetünk, hogy a (2.1) alatti $\beta(\alpha)$ függvény az

$$f(\alpha)\approx\beta(\alpha)=\pm(1-\alpha)\beta_L e^{m\alpha} \tag{2.2}$$

egyenlettel közelíthető.

3. A természetes adatsorok vizsgálata

Elemezzünk a következőkben néhány — fizikai tartalommal rendelkező — idősort. Válasszuk ki erre a célra az ország 11 meteorológiai állomásán 1937 és 1975 között mért éves csapadékösszegek idősorait. Az 1. táblázat szerint — a battonyai mérőállomás kivételével — az idősor elemei a *Wald—Wolfowitz-próba* szerint függetlenek ($p_W=0{,}112-0{,}882$), a β ferdeségi együttható pedig pozitív.

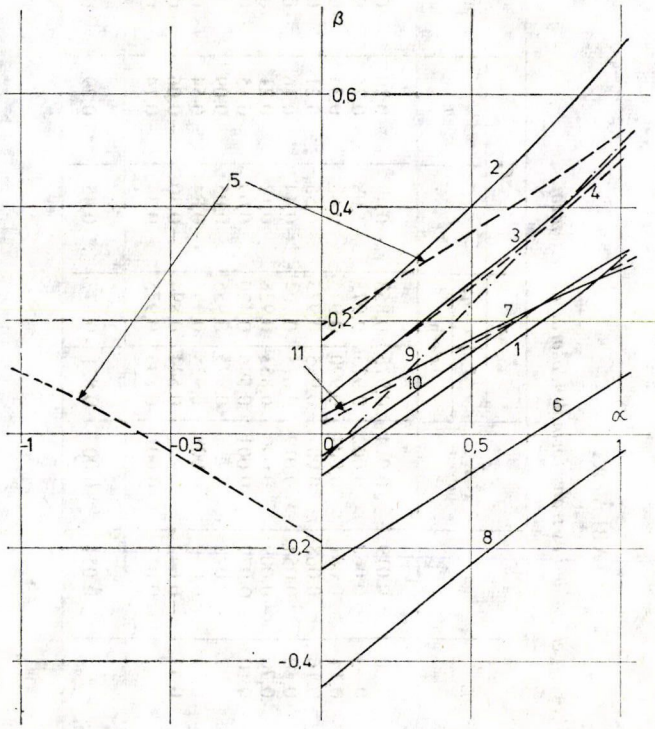
1. TÁBLÁZAT

Éves csapadékösszegek eredeti és transzformált idősorainak jellemzői

Sz.	Mérőállomás helye	p_W	β_1	α_{kr}	m	β az $\alpha=\alpha_{kr}$ helyen	p_K az helyen		E_r^F		E_r^A	
							$\alpha=1$	$\alpha=\alpha_{kr}$	ξ	η	ξ	η
1.	Ásotthalom	0,486	0,307	0,20	0,019	−0,014	0,975	0,853	−0,49	−0,30	−0,24	0,00
2.	Cibakháza	0,112	0,679	−0,33	0,060	0,002	0,975	0,975	−1,08	−0,75	−0,13	0,21
3.	Kondoros	0,300	0,498	−0,14	0,081	0,004	0,990	0,975	−0,68	−0,33	−0,23	0,10
4.	Karcag	0,310	0,481	−0,16	0,068	0,004	0,955	0,975	−0,55	−0,21	−0,38	−0,06
5.	Túrkeve	0,485	0,533	−0,60	0,054	0,019	0,678	0,704	−0,44	−0,16	−0,59	0,13
6.	Téglás	0,567	0,101	0,70	−0,009	0,000	0,853	0,928	0,12	0,21	−0,23	−0,13
7.	Karcsa	0,762	0,292	−0,10	−0,079	−0,001	0,596	0,420	0,09	0,45	−0,24	0,06
8.	Battonya	0,019	−0,028	1,07	−0,078		0,975	0,955	−0,05	−0,09	0,35	0,34
9.	Vásárosnamény	0,571	0,511	0,09	0,059	0,008	0,928	0,995	−0,84	−0,54	−0,06	0,09
10.	Tiszalök	0,882	0,318	0,11	−0,042		0,544	0,894	−0,30	−0,06	−0,18	0,09
11.	Dévaványa	0,287	0,299	−0,07	0,042	0,001	0,894	0,975	−0,34	−0,14	−0,40	−0,07
Abszolút értékek átlaga			0,368		0,054	0,007	0,851	0,877	0,45	0,29	0,28	0,12

3.1. A $\beta(\alpha)$ összefüggés közelítő alakja

A csapadéksorok ferdeségi együtthatójának a hatványkitevővel való változását a 2. ábrán látható görbesereg érzékelteti. A görbék meglepően sima lefutásúak, annak ellenére, hogy a sorok statisztikai szerkezetét véletlen jellegű tényezők alakították ki.



2. ábra
A csapadék idősorára meghatározott $f(\alpha)$ összefüggés

Az alaki hasonlóságból arra következtethetünk, hogy a $\beta(\alpha)$ kapcsolat itt is a (2.2) alatti egyenlettel közelíthető meg. Legyen az $\alpha=1$ helyen (vagyis az eredeti sorra) $\beta=\beta_1$, akkor az analógia alapján a következő egyenlet írható fel:

$$f(\alpha) \approx \beta(\alpha) = (1-b\alpha)\beta_L e^{m\alpha}, \tag{3.1}$$

ahol b annak az $\alpha=\alpha_{kr}$ hatványkitevőnek a reciproka, amelyre a sor ferdeségi együtthatója $\beta=0$.

A (3.1) egyenletben szereplő b és m értékeket a következőképpen célszerű meghatároznunk. A sor elemeiből négyzetgyököt vonva számítható $\beta_{0,5}$, majd a logaritmikus sorból β_L. Ekkor a következő két egyenletet írhatjuk fel:

$$\beta_{0,5} = \left(1-\frac{0,5}{\alpha_{kr}}\right)\beta_L e^{0,5m}$$

és

$$\beta_1 = \left(1 - \frac{1}{\alpha_{kr}}\right) \beta_L e^m.$$

A második egyenletből:

$$e^m = \frac{\beta_1}{\beta_L} \frac{\alpha_{kr}}{\alpha_{kr} - 1}; \tag{3.2}$$

ezt az első egyenletbe helyettesítve α_{kr}-re vegyes másodfokú egyenletet kapunk, melynek megoldása:

$$\alpha_{kr} = 0{,}5 \pm \sqrt{\frac{A+1}{4A}}, \quad \text{ahol} \quad A = \frac{\beta_{0,5}^2}{\beta_1 \cdot \beta_L} - 1. \tag{3.3}$$

Az adott idősort vizsgálva úgy kell eljárnunk, hogy első lépésként a (3.3) egyenletből meghatározzuk α_{kr} értékét (a két gyök közül a valódit a $\beta_1 - \beta_{0,5} - \beta_L$ értékhármas előjelei alapján kiválasztva), majd ezt a (3.2) egyenletbe helyettesítve számítjuk az m állandót.

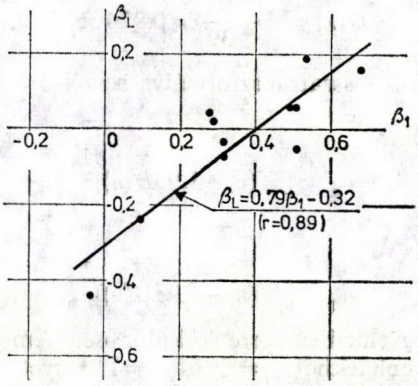
A csapadéksorokra kapott α_{kr} és m értékeket az 1. táblázat 5. és 6. oszlopa tünteti fel. Amint ebből láthatjuk, α_{kr} hat állomásra negatív, ötre pozitív, az $f(\alpha)$ görbe négy esetben konvex, hét esetben konkáv.

Az m, β_1 és β_L értékek korrelációs mátrixa a következő:

	m	β_1	β_L
m	1	0,768	0,627
β_1	0,768	1	0,893
β_L	0,627	0,893	1

Az 1. táblázat szerint $|m|$ a vizsgált tartományban közel zérus, összhangban a β_1 és β_L értékek közötti szoros lineáris korrelációval (3. ábra).

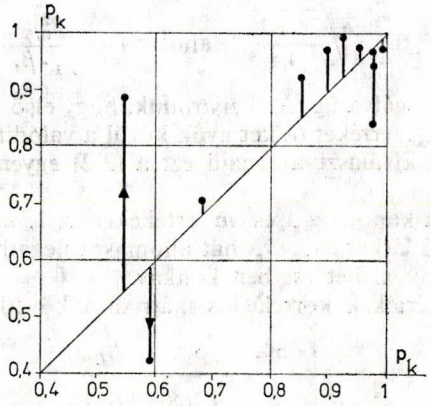
A (3.2) és (3.3) egyenletek gyakorlati használhatóságát bizonyítja, hogy
a) a számított (1. táblázat, 5. oszlop) és szerkesztett α_{kr} értékek igen jól egyeznek,
b) az $\alpha = \alpha_{kr}$ kitevővel számolt sorok ferdesége átlagosan $|\beta| = 0{,}007$ (1. táblázat, 7. oszlop).



3. ábra
A csapadék-idősorokra számított β_1, β_L értékpárok és regressziós egyenesük

3.2. Az eredeti és javított sorok normalitásvizsgálata

Az eredeti idősorok *Kolmogorov*-féle valószínűségi értékei $p_K=0{,}544$ és $p_K=0{,}990$ között változtak, az átlag 0,851 volt (1. táblázat, 8. oszlop). Az aszimmetriát megszüntetve ez az átlag $\bar{p}_K=0{,}877$-re módosult, a változás tehát nem jelentős. Ha az $\alpha=1$ és az $\alpha=\alpha_{kr}$ kitevőkhöz tartozó p_K értékeket állomásonként összehasonlítjuk, azt figyelhetjük meg, hogy az utóbbiak mindössze hat esetben nagyobbak az előbbieknél (4. ábra).



4. ábra
A csapadék-idősorok eredeti p_K-értéke (abszcissza) és a $\beta\sim 0$-hoz tartozó p_K-értéke (ordináta)

Mivel az idősorok elemei éves csapadékösszegek — tehát nem extremumok —, jogosnak látszik a $p=1/39$ valószínűséggel várható értéknek a normális eloszlás alapján való számítása (különösen a transzformált sorokra). Mivel

$$\Phi(\lambda) = 1-\frac{1}{39} = 0{,}974\,36,$$

így $\lambda=1{,}95$. Az η-sorokat visszatranszformálva a szélső értékek a következő kifejezésekből számíthatók:

$$F_\eta = [M(\eta)+1{,}95D(\eta)]^{\frac{1}{\alpha}} \tag{3.4}$$

és

$$A_\eta = [M(\eta)-1{,}95D(\eta)]^{\frac{1}{\alpha}}, \tag{3.5}$$

feltéve, hogy az utóbbi egyenletben szereplő különbség nem negatív.

A (3.4) és (3.5) egyenletekből $\eta=\xi$ és $\alpha=1$ helyettesítéssel számíthatók az eredeti sornak a normális eloszláshoz tartozó szélső értékei is (F_ξ és A_ξ), amelyek általában nem azonosak a ténylegesen észlelt szélső értékekkel (F_t és A_t). A szá-

mított és tényleges extrémumok relatív eltérését az

$$E^F_{r,\xi} = \frac{F_\xi - F_t}{D(\xi)} \quad \text{és} \quad E^A_{r,\xi} = \frac{A_\xi - A_t}{D(\xi)} \tag{3.6}$$

hányadosokkal definiálhatjuk.

Az η-sorokat visszatranszformálva a (3.4) és (3.5) egyenletekből számítani tudjuk a várható szélső értékeket, majd az

$$E^F_{r,\eta} = \frac{F_\eta - F_t}{D(\zeta)} \quad \text{és} \quad E^A_{r,\eta} = \frac{A_\eta - A_t}{D(\xi)} \tag{3.7}$$

kifejezésekből ezeknek a tényleges extrémumokhoz viszonyított relatív eltéréseit.

Az 1. táblázat 10—13. oszlopai szerint a relatív eltérések abszolút értékeinek átlaga

$$\bar{E}^F_{r,\xi} = 0{,}45, \quad \text{illetve} \quad \bar{E}^A_{r,\xi} = 0{,}28$$

az eredeti sorokra, és

$$\bar{E}^F_{r,\eta} = 0{,}29, \quad \text{illetve} \quad \bar{E}^A_{r,\eta} = 0{,}12$$

a transzformált ($\alpha = \alpha_{kr}$, $\beta \sim 0$) sorokra. Mivel az éves csapadékösszegek szórásának átlaga a 11 mérőállomásra 116 mm, az idősor felső és alsó elemének becslését egyaránt

$$(0{,}45 - 0{,}29)116 = (0{,}28 - 0{,}12)116 \cong 19 \text{ mm-rel}$$

sikerült javítanunk, jóllehet az eredeti sorok ferdeségének ($|\beta|$) átlaga mindössze 0,368 volt.

Arra vonatkozóan, hogy a becslés pontosságát sikerül-e és ha igen, milyen mértékben fokoznunk, sem az eredeti, sem a transzformált sorral végzett *Kolmogorov*-próba nem ad egyértelmű választ. A két p_K-érték különbsége valamivel több információt nyújt, de ez a kapcsolat sem szoros (a felső szélső értékre pl. $r = 0{,}39$).

4. Következtetések

A ferde eloszlású adat-, ill. idősorok aszimmetriája — függetlenül a sor statisztikai szerkezetétől — az $\eta = \xi^\alpha$ transzformációval megszüntethető.

A második és harmadik centrális momentummal definiált β ferdeségi együttható és az α hatványkitevő közötti kapcsolatot sima lefutású, enyhe görbülettel jellemzett, a kitevő állandójától függően konvex vagy konkáv exponenciális függvény írja le. A függvénynek az $\alpha = 0$ helyen szakadása van: ezen a helyen $|\beta|$ az $\eta = \ln \xi$ transzformációval előállított sor β_L ferdeségi együtthatójával azonos.

A közelítő függvény tulajdonságainak ismeretében felírható ennek a gyakorlati feladatok megoldására alkalmas alakja. Az $\alpha = \alpha_{kr}$ kitevő (amelyre $\beta = 0$) számításához az $\eta = \sqrt{\xi}$ és $\eta = \ln \xi$ transzformációkat célszerű elvégeznünk.

A transzformálás segítségével kapott sorból a tetszőleges valószínűséghez rendelt ξ-értékek pontosabban közelíthetők, a javítás mértékére azonban az illeszkedésvizsgálat (pl. a *Kolmogorov*-próba) csak tájékoztató információt nyújt.

Végezetül meg kívánom köszönni PRÉKOPA ANDRÁS akadémikusnak a kézirattal kapcsolatban adott értékes tanácsait.

(Beérkezett: 1979. május 15.)

RÉTHÁTI LÁSZLÓ
1092 BUDAPEST, RÁDAY U. 43.

SYMMETRIESIERUNG VON SCHRÄG VERTEILTEN DATENREIHEN MIT POTENZIERUNG

L. RÉTHÁTI

Mit der Transformation $\eta=\xi^{\alpha}$ ist erreichbar, dass der Schrägefaktor (skewness) der so hergestellten Reihe Null sei. Den dazu gehörende Exponent α_{kr} kann man aus einer exponentiellen Gleichung rechnen; die mit diesem Wert transformierte Reihe ist für praktische Zwecke als eine normal verteilte Reihe annehmbar.

Alkalmazott Matematikai Lapok **5** *(1979)* 303—374

AUTOREGRESSZIÓS TÍPUSÚ GAUSS-FOLYAMATOK NÉHÁNY JELLEMZÉSI PROBLÉMÁJÁRÓL

PHAM NGOC PHUC
Hanoi

Bevezetés

Az utóbbi években jelentős szerepet játszanak statisztikai kutatásokban azok a feladatok, amelyek eloszlásoknak a statisztikák tulajdonságaival történő jellemzését tűzik ki célul. KAGAN—LINNIK—RAO [21] 1972-ben megjelent könyve az utóbbi évtizedben ebben a kérdéskörben elért jelentős eredmények összefoglalása.

Az eloszlások jellemzésének feladatai mind elméleti, mind gyakorlati alkalmazási szempontból fontos szerepet játszanak a matematikai statisztika olyan ágaiban, mint a becsléselmélet, a hipotézis vizsgálat, a szekvenciális analízis. Érdekes nemparaméteres kritériumok megszerkesztésében játszanak szerepet (lásd pl. SARKADI [49]), továbbá hasznosak a lineáris modellek elméletében a biometriában (lásd RAO [46]).

Az eloszlások jellemzésében központi helyet foglalnak el a normális eloszlás jellemzésével foglalkozó feladatok.

Független megfigyeléssorozatokra a hatvanas években sok olyan probléma érdekes megoldását adták meg a matematikusok, amelyben a normális eloszlás jellemezhető a statisztikában gyakran használt becslések megengedhetőségével és optimalitásával (a definíciók megtalálhatók I. fejezet 1. pontjában).

Az első jelentős eredmény, amely a problémakör további vizsgálatának széles körű irodalmát indította meg, KAGAN—LINNIK—RAO [20] nevéhez fűződik. Az eredmény a következő:

Legyenek a megfigyelések

$$x(j) = \varepsilon(j) + \theta, \quad j = 1, \ldots, n$$

alakúak, ahol θ ismeretlen paraméter, $\varepsilon(j)$ a megfigyelés hibája. $n \geqq 3$ esetén az $\bar{x} = \frac{x_1 + \ldots + x_n}{n}$ empirikus közép akkor és csak akkor megengedhető a négyzetes veszteségfüggvénnyel θ torzítatlan becslései osztályában, ha az $\varepsilon(j)$ valószínűségi változók normális eloszlásúak.

További eredmények, amelyek különböző megfigyelési sémákhoz és különböző veszteségfüggvényekhez kapcsolódnak, megtalálhatók az irodalomban (lásd pl. KAGAN [17], STEIN [50], FINTUSAL [10], FARREL [7], JOSHI [15], KLEBANOV [27]).

Később KAGAN és KARPOV [22] egy másik irányban — a *Bayes-féle* megfogalmazásban — vizsgálták a feladatot, és az invariáns becsléselméletben elért eredmények *Bayes-féle* analogonjait adták meg.

A független megfigyelési sémára vonatkozó statisztikai vizsgálatok fejlődésével párhuzamosan, az ötvenes években megindult a sztochasztikus folyamatok statisztikájának kidolgozása is. Ismeretes, hogy a matematikai statisztika klasszikus problémáinak a független megfigyeléssorozat esetére vonatkozó sok megoldása átvehető sztochasztikus folyamatokra is. Így, természetes módon merül fel a sztochasztikus folyamatok statisztikájában az a kérdéskör, amely a folyamatok — és elsősorban *Gauss-folyamatok* — paraméter becslésének valamilyen jó tulajdonságaival való jellemzésére vonatkozik. Jelen dolgozatban megvizsgáljuk a *Gauss-stacionárius* (speciálisan autoregressziós típusú) folyamatok jellemzését a paraméter becslésének megengedhetőségével és a *Bayes-féle becslés* linearitásával.

Az első fejezet bevezető jellegű. Ebben a fejezetben a téma fejlődését is figyelembe véve röviden ismertetjük az alapfogalmakat és a legfontosabb tételeket, amelyekre a későbbi tárgyalásban szükség van.

Az új eredmények ismertetésére a II., III. és IV. fejezetekben kerül sor. A II. fejezet első két pontjában vizsgáljuk az $x(j)=\xi(j)+\theta$ sémát, ahol $\xi(j)$ p-edrendű ($p\geqq 1$) autoregressziós folyamat, és θ ismeretlen paraméter. *Kagan—Linnik—Rao-módszerét* felhasználva bebizonyítjuk a 2.1., 2.2., 2.3. és 2.4. tételeket, amelyek az $x(j)$ folyamat *Gauss* volta, a legjobb lineáris becslés megengedhetősége és a legjobb lineáris becslés polinomjainak optimalitása közötti kapcsolattal foglalkoznak. Ezek az eredmények KAGAN—LINNIK—RAO eredményeinek továbbfejlesztései. A 3. pontban a problémát az $x(j)=\xi(j)+\theta$ séma esetén vizsgáljuk, ahol $\xi(j)$ k-dimenziós reguláris stacionárius *Gauss—Markov-folyamat.*

A dolgozat III. fejezete az $x(j)=\xi(j)+\theta$($\xi(j)$ p-edrendű autoregressziós folyamat) alakú *Gauss-folyamatok* θ paraméterének becslésével foglalkozik *Bayes-féle* megfogalmazásban. Ebben a fejezetben bizonyítjuk a 3.1., 3.1.′, 3.2. és 3.4. tételeket, amelyek az $x(j)$ *Gauss-folyamat* jellemzését vizsgálják a θ paraméter *Bayes-féle becslésének* és a *Bayes-féle polinomiális becslésének* linearitásával (a 3.3. tétel KAGAN—KARPOV eredményének megismétlése).

Az itt ismertetésre kerülő eredmények KAGAN—KARPOV [22] független megfigyelési sémára vonatkozó eredményeinek kiterjesztései a függő esetre. A bizonyítás KAGAN—KARPOV gondolatmenetén alapszik, de már bonyolultabb, mint a független megfigyelési séma esetén.

A IV. — utolsó — fejezet stacionárius folyamat szórásnégyzete becslésének problémáját tárgyalja. ARATÓ [2] és KAGAN—RUHIN [17], [23] gondolatmenetén alapulva a skála paraméter KAGAN által definiált szabályos becsléseinek és *Pitman-féle becsléseinek* fogalmát alkalmazzuk a sztochasztikus folyamatok szórásnégyzetének becslésére. A fejezet 2. pontjában egy konkrét esetet vizsgáltunk, a p-edrendű autoregressziós stacionárius *Gauss-folyamat* esetét. Bizonyítjuk, hogy ebben az esetben a folyamat σ^2-paraméterének *Pitman-féle becslése* és *maximum likelihood becslése* megegyezik egymással, és egyben a legjobb torzítatlan becslést adja σ^2-re (4.2. tétel). (A folyamat szórásnégyzete $\beta\sigma^2$, ahol β egy ismert pozitív állandó.)

Befejezésül szeretném hálámat kifejezni ARATÓ MÁTYÁSnak, aki támogatott munkámban, a dolgozat témájának kiválasztásában és annak elkészítésében.

Köszönettel tartozom BÁRTFAI PÁLnak, PERGEL JÓZSEFnek és KRÁMLI ANDRÁSnak, a matematikai tudományok kandidátusainak, a dolgozat gondos átnézéséért és értékes megjegyzéseikért.

I. FEJEZET

1. Paraméter becslések standard sémája. Megengedhető becslések. Az eltolási paraméter. Pitman-féle becslések.

Egy θ paraméter becslésének feladatában az $(\mathcal{X}, \mathcal{A}, P_\theta)$ valószínűségi mezőt vizsgáljuk, ahol $\mathcal{X}$ az x valószínűségi elem megfigyeléseinek tere, $\mathcal{A}$ az $\mathcal{X}$ részhalmazaiból alkotott σ algebra és $\{P_\theta, \theta \in \Theta\}$ valószínűségi mértékek egy, a θ ismeretlen paramétertől függő összessége.

Ezt tekintjük egy ismeretlen θ paraméterre vonatkozó becslés standard sémájának. A θ paraméter értéke ismeretlen, és a feladat az x megfigyelések alapján a $t(\theta)$ paraméterfüggvény becslése.

Legyen a $\tau(x)$ statisztika $t(\theta)$-nak valamilyen becslése, és $r(\tau, t)$ egy nemnegatív veszteségfüggvény.

A $t(\theta)$ paraméterfüggvény $\tau(x)$ becslése és $r(\tau, t)$ veszteségfüggvénye esetén a megfelelő rizikó, amikor a paraméter igazi értéke θ, definíció szerint a következő:

$$R(\tau; \theta) = E_\theta r[\tau(x); t(\theta)],$$

ahol itt és a továbbiakban E_θ jelöli a P_θ eloszlásnak megfelelő várható értéket.

Legyen $\tau_1(x)$ és $\tau_2(x)$ a $t(\theta)$ paraméterfüggvény két becslése. Ha fennáll az

$$R(\tau_1(x); \theta) \leqq R(\tau_2(x); \theta), \quad \text{minden } \theta \in \Theta\text{-ra}$$

egyenlőtlenség, akkor azt mondjuk, hogy a $\tau_1(x)$ becslés nem rosszabb mint $\tau_2(x)$.

Legyen $\mathcal{U}$ a $t(\theta)$ becsléseinek egy osztálya.

1.1. Definíció. Az $\mathcal{U}$-hoz tartozó $\tau^*(x)$ becslést akkor nevezzük *megengedhetőnek*, ha nincs olyan $\tau(x) \in \mathcal{U}$ becslés, amelyre

$$R(\tau; \theta) \leqq R(\tau^*; \theta), \quad \theta \in \Theta\text{-ra}, \tag{1.1}$$

és (1.1)-ben legalább egy θ-ra egyenlőtlenség áll fenn.

Azt a $\tau^*(x)$ becslést, amely a $t(\theta)$ összes becslései osztályában megengedhető, *abszolút megengedhetőnek* nevezzük.

1.2. Definíció. Az $\mathcal{U}$-hoz tartozó $\tau^0(x)$ becslést akkor nevezzük *optimálisnak* az $\mathcal{U}$-osztályban az $r(\tau, t)$ veszteségfüggvénnyel, ha minden $\mathcal{U}$-hoz tartozó $\tau(x)$ becslésre

$$R(\tau^0; \theta) \leqq R(\tau; \theta), \quad \text{minden } \theta \in \Theta\text{-ra}. \tag{1.2}$$

1.3. Definíció. A $\tau_0(x) \in \mathcal{U}$ becslést az $\mathcal{U}$ osztályban akkor nevezzük a θ_0-*pontban optimálisnak* (néha *lokálisan optimálisnak*), ha minden $\tau(x) \in \mathcal{U}$-ra

$$R(\tau_0; \theta_0) \leqq R(\tau; \theta_0).$$

Most feltesszük, hogy az $\mathcal{X}$ mintatér az $\mathcal{R}^n$ n-dimenziós euklideszi tér, $\mathcal{A}$ a *Borel mérhető halmazok* összessége, a Θ paramétertér pedig a valós egyenes ($\Theta = \mathcal{R}^1$) és a becsülendő paraméterfüggvény $t(\theta) \equiv \theta$.

Ha a P_θ mértékek a θ paramétertől a következőképpen függenek:

$$P_\theta(A) = \int_A \cdots \int dF(x_1-\theta, \ldots, x_n-\theta), \quad A \in \mathscr{A}, \tag{1.3}$$

akkor azt mondjuk, hogy θ eltolási paraméter.

Tekintsük példaként a közvetlen mérések sémáját. A megfigyelések a következő alakúak:

$$x_j = \theta + \varepsilon_j, \quad j = 1, \ldots, n,$$

ahol ε_j, $j=1, \ldots, n$, a véletlen eltéréseket jelentik. Ha $F(x_1, \ldots, x_n)$ az $\varepsilon_1, \ldots, \varepsilon_n$ valószínűségi változók együttes eloszlásfüggvénye, akkor a P_θ mértékek (1.3) alakúak.

Az eltolás paramétere becslésének feladatában természetes megvizsgálni az úgynevezett korrekt becslések osztályát, amelyen a következőt értjük:

1.4. Definíció. A θ eltolási paraméter $\tilde{\theta}(x_1, \ldots, x_n)$ becslését akkor nevezzük *korrekt* becslésnek, ha tetszőleges $-\infty < c < +\infty$ esetén teljesül

$$\tilde{\theta}(x_1+c, \ldots, x_n+c) = \tilde{\theta}(x_1, \ldots, x_n)+c, \tag{1.4}$$

Könnyű látni, hogy pl. az $\bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j$ empirikus közép, és az

$$l = \sum_{j=1}^{n} c_j x_j \left(\sum_{j=1}^{n} c_j = 1 \right)$$

lineáris becslés korrekt becslések.

Az irodalomban többen a korrekt becslést a koordináta-rendszer megválasztásával szemben invariáns becslésnek nevezik.

Független megfigyeléssorozat esetén a korrekt becslés fogalmával PITMAN [42] foglalkozott először. Ő részletes tárgyalását adta az eltolási és később ismertetésre kerülő skála paraméter becslésének olyan eloszlások esetén, amelyek ezen két paraméter egyikétől vagy mindkettőtől függenek.

A definícióból nyilvánvaló, hogy ha $\hat{\theta}_1$ és $\hat{\theta}_2$ korrekt becslések, akkor

$$\hat{\theta}_1 - \hat{\theta}_2 = h(x_2-x_1, \ldots, x_n-x_1), \tag{1.5}$$

ahol $h(y)$ y-nak valamilyen függvényét jelöli.

Jelöljük a továbbiakban $\mathscr{K}$-val a korrekt becslések osztályát. Ha a veszteségfüggvény $r(\tilde{\theta}; \theta) = r(\tilde{\theta}-\theta)$ csak a változók különbségétől függ, akkor könnyű belátni, hogy minden $\hat{\theta} \in \mathscr{K}$-ra a rizikó

$$R(\hat{\theta}; \theta) = E_\theta r(\hat{\theta}-\theta) = \text{konstans},$$

független θ-tól. Ezért minden korrekt becslés vagy optimális, vagy nem megengedhető $\mathscr{K}$-ban, más szóval a korrekt becslések osztályában az optimalitás ekvivalens a megengedhetőséggel.

1.5. Definíció. A θ eltolási paraméter *Pitman-féle becslésének* (az $r(\hat{\theta}, \theta)$ veszteségfüggvénnyel) azt a $\hat{\theta}_0 = \hat{\theta}_0(x_1, \ldots, x_n)$ $\mathcal{K}$-ban optimális becslést nevezzük, amelyre

$$R(\hat{\theta}_0; \theta) = \min_{\hat{\theta} \in \mathcal{K}} R(\hat{\theta}; \theta).$$

Ha az $R(\hat{\theta}; \theta)$ minimumát egyetlen $\hat{\theta}_0$ becslésre éri el, (azonosnak tekintjük azokat a statisztikákat, amelyek minden P_θ-mértékre nézve m. m. megegyeznek), akkor világos, hogy a $\hat{\theta} \in \mathcal{K}$ korrekt becslés megengedhetőségének szükséges feltétele az, hogy

$$\hat{\theta} = \hat{\theta}_0 \quad \text{m. m. } P_\theta, \quad \theta \in \mathcal{R}^1,$$

vagy ezzel ekvivalens

$$\hat{\theta} = \hat{\theta}_0 \quad P_0 \text{ m.m.}$$

A továbbiakban — hacsak nincs külön megjegyzés, a tárgyalás az

$$r(\hat{\theta}, \theta) = |\hat{\theta} - \theta|^2$$

négyzetes (*Gauss-féle*) veszteségfüggvényhez kapcsolódik.

Tegyük most fel, hogy

$$\int_{R^n} x_j^2 \, dF(x_1, \ldots, x_n) = \sigma_j^2 < \infty, \quad j = 1, \ldots, n. \tag{1.6}$$

Legyen $l = \sum_{j=1}^{n} c_j x_j \left(\sum_{j=1}^{n} c_j = 1 \right)$ egy lineáris statisztika. Legyen továbbá

$$\mathbf{y} = (x_2 - x_1, \ldots, x_n - x_1).$$

Vizsgáljuk a

$$\hat{\theta}_0 = l - E_0(l|\mathbf{y}) \tag{1.7}$$

statisztikát. Nyilvánvaló, hogy $\hat{\theta}_0 \in \mathcal{K}$. Érvényes a következő tétel, amely szerepel pl. [21]-ben.[1]

1.1. Tétel. 1°. Az (1.6) feltétel teljesülése esetén a

$$\hat{\theta}_0 = l - E_0(l|\mathbf{y})$$

becslés *Pitman-féle becslés,* miközben a $\hat{\theta} \in \mathcal{K}$ becslésre

$$E_\theta(\hat{\theta} - \theta)^2 > E_\theta(\hat{\theta}_0 - \theta)^2, \quad \theta \in \mathcal{R}^1 \tag{1.8}$$

azon eset kivételével, amikor $\hat{\theta} = \hat{\theta}_0$ P_0 majdnem mindenütt.

2°. Ha az $F(x_1, \ldots, x_n)$ eloszlás abszolút folytonos (a *Lebesgue-mértékre* nézve),

$$F(x_1, \ldots, x_n) = \int_{-\infty}^{x_1} \ldots \int_{-\infty}^{x_n} f(u_1, \ldots, u_n) \, du_1 \ldots du_n,$$

akkor

$$\hat{\theta}_0 = \frac{\int_{-\infty}^{+\infty} \xi f(x_1 - \xi, \ldots, x_n - \xi) \, d\xi}{\int_{-\infty}^{+\infty} f(x_1 - \xi, \ldots, x_n - \xi) \, d\xi}. \tag{1.9}$$

Bizonyítás. (A tétel bizonyítása megtalálható pl. [17]-ben és [21]-ben.)

1°. Legyen $\hat{\theta}\in\mathcal{K}$. (1.5) szerint $\hat{\theta}-\hat{\theta}_0=h(\mathbf{y})$. $E_0\hat{\theta}^2=\infty$ esetén nyilvánvaló, hogy $E_\theta(\hat{\theta}_0-\theta)^2<E_\theta(\hat{\theta}-\theta)^2$, $\theta\in R^1$. Feltesszük most, hogy $E_0\hat{\theta}^2<\infty$. Ekkor

$$E_\theta(\hat{\theta}-\theta)^2 = E_\theta(\hat{\theta}-\hat{\theta}_0+\hat{\theta}_0-\theta)^2 = E_\theta(\hat{\theta}-\hat{\theta}_0)^2+E_\theta(\hat{\theta}_0-\theta)^2+2E_\theta\{(\hat{\theta}-\hat{\theta}_0)(\hat{\theta}_0-\theta)\}.$$

Figyelembe vesszük, hogy

$$E_\theta\{(\hat{\theta}-\hat{\theta}_0)(\hat{\theta}_0-\theta)\} = E_0\{\hat{\theta}_0(\hat{\theta}-\hat{\theta}_0)\} = E_0\{E_0[\hat{\theta}_0(\hat{\theta}-\hat{\theta}_0)|\mathbf{y}]\} = E_0\{(\hat{\theta}-\hat{\theta}_0)E_0(\hat{\theta}_0|\mathbf{y})\} = 0,$$

ahonnan adódik, hogy

$$E_\theta(\hat{\theta}-\theta)^2 \geqq E_\theta(\hat{\theta}_0-\theta)^2, \quad \text{minden } \hat{\theta}\in\mathcal{K}\text{-ra.}$$

Tehát igazoltuk az 1.1. tétel 1°. állítását.

2°. Legyen speciálisan $l=\frac{1}{n}\sum_{j=1}^{n}x_j=\bar{x}$, akkor $\hat{\theta}_0=\bar{x}-E_0(\bar{x}|\mathbf{y})$. Minthogy

$$E_0(\bar{x}|\mathbf{y}) = E_0\left(x_1+\frac{(x_2-x_1)+\ldots+(x_n-x_1)}{n}\middle|\mathbf{y}\right) = \bar{x}-x_1+E_0(x_1|\mathbf{y}),$$

ennélfogva

$$\hat{\theta}_0 = x_1-E_0(x_1|\mathbf{y}).$$

Bevezetjük az

$$u_1 = x_1$$
$$u_2 = x_2-x_1$$
$$\ldots\ldots\ldots$$
$$u_n = x_n-x_1$$

új változókat. Könnyű belátni, hogy ennek a leképezésnek *Jacobi-determinánsa* 1-gyel egyenlő, ezért az $u_1, \ldots, u_n$ változók együttes sűrűségfüggvénye (feltéve, hogy $\theta=0$) a következő:

$$p(u_1, u_2, \ldots, u_n) = f(u_1, u_1+u_2, \ldots, u_1+u_n).$$

Ebből adódik, hogy

$$E_0(u_1|u_2, \ldots, u_n) = \frac{\int_{-\infty}^{+\infty} u_1 f(u_1, u_1+u_2, \ldots, u_1+u_n)\,du_1}{\int_{-\infty}^{+\infty} f(u_1, u_1+u_2, \ldots, u_1+u_n)\,du_1}.$$

Visszatérve az $x_1, \ldots, x_n$ változókra, kapjuk

$$E_0(x_1|x_2-x_1, \ldots, x_n-x_1) = \frac{\int_{-\infty}^{+\infty} u_1 f(u_1, u_1+x_2-x_1, \ldots, u_1+x_n-x_1)\,du_1}{\int_{-\infty}^{+\infty} f(u_1, u_1+x_2-x_1, \ldots, u_1+x_n-x_1)\,du_1} =$$

$$= \frac{\int_{-\infty}^{+\infty} (x_1-\xi) f(x_1-\xi, x_2-\xi, \ldots, x_n-\xi)\, d\xi}{\int_{-\infty}^{+\infty} f(x_1-\xi, x_2-\xi, \ldots, x_n-\xi)\, d\xi} = x_1 - \frac{\int_{-\infty}^{+\infty} \xi f(x_1-\xi, x_2-\xi, \ldots, x_n-\xi)\, d\xi}{\int_{-\infty}^{+\infty} f(x_1-\xi, x_2-\xi, \ldots, x_n-\xi)\, d\xi}$$

összefüggést, amiből

$$\hat{\theta}_0 = \frac{\int_{-\infty}^{+\infty} \xi f(x_1-\xi, \ldots, x_n-\xi)\, d\xi}{\int_{-\infty}^{+\infty} f(x_1-\xi, \ldots, x_n-\xi)\, d\xi}$$

következik. Ezzel kész az 1.1. tétel bizonyítása.

Abban az esetben, amikor $x_1, \ldots, x_n$ független azonos eloszlású megfigyelések, amelyeknek sűrűségfüggvénye (a *Lebesgue-mértékre* nézve) $f(x)$, akkor a *Pitman-féle becslést* a következő alakban kapjuk:

$$\hat{\theta}_0 = \frac{\int_{-\infty}^{+\infty} \xi \prod_{j=1}^{n} f(x_j-\xi)\, d\xi}{\int_{-\infty}^{+\infty} \prod_{j=1}^{n} f(x_j-\xi)\, d\xi}. \tag{1.10}$$

PITMAN [42] már 1938-ban bebizonyította, hogy $\hat{\theta}_0$ legjobb eltolásinvariáns becslése θ-nak, azaz $\hat{\theta}_0$ minimális szórású becslése θ-nak a korrekt becslések osztályában.

GIRSHICK és SAVAGE [13] megmutatta, hogy $\hat{\theta}_0$ minimax θ összes becsléseinek osztályában. Ez az eredmény következik KUDÓ [30] és KIEFER [25] későbbi általánosabb eredményeiből is.

STEIN [50] (1959) szerint, ha az $f(x)$ sűrűségfüggvény eleget tesz az

$$\int_{-\infty}^{+\infty} \prod_{j=1}^{n} f(x_j) \left\{ \frac{\int_{-\infty}^{+\infty} \xi^2 \prod_{j=1}^{n} f(x_j-\xi)\, d\xi}{\int_{-\infty}^{+\infty} \prod_{j=1}^{n} f(x_j-\xi)\, d\xi} - \left(\frac{\int_{-\infty}^{+\infty} \xi \prod_{j=1}^{n} f(x_j-\xi)\, d\xi}{\int_{-\infty}^{+\infty} \prod_{j=1}^{n} f(x_j-\xi)\, d\xi} \right)^2 \right\}^{3/2} dx_1 \ldots dx_n < \infty \tag{1.11}$$

feltételnek, akkor $\hat{\theta}_0$ abszolút megengedhető.

Legyen

$$p(\xi) = \frac{\prod_{j=1}^{n} f(x_j-\xi)}{\int_{-\infty}^{+\infty} \prod_{j=1}^{n} f(x_j-\xi)\, d\xi},$$

akkor az (1.10) *Pitman-féle becslés*

$$\hat{\theta}_0 = \int_{-\infty}^{+\infty} \xi p(\xi)\, d\xi \tag{1.12}$$

alakú.

Az (1.11) *Stein-féle feltétel* $p(\xi)$ segítségével az alábbi alakban írható fel:

$$E_0\Big\{\int_{-\infty}^{+\infty} \xi^2 p(\xi)\, d\xi - \Big(\int_{-\infty}^{+\infty} \xi p(\xi)\, d\xi\Big)^2\Big\}^{3/2} < \infty. \tag{1.13}$$

(1.11)-ből (vagy ekvivalens módon (1.13)-ból) speciálisan következik, hogy ha $\int_{-\infty}^{+\infty} |x|^3 f(x)dx < \infty$, akkor a $\hat{\theta}_0$ *Pitman-féle becslés* abszolút megengedhető.

IBRAHIMOV és HASZMINSZKIJ [15] (1.13)-ból levezették a megengedhetőség olyan feltételét, amely közvetlenül az $f(x)$ függvény segítségével van megfogalmazva.

A stacionárius sztochasztikus folyamatok esetén az $Ex(t)=\theta$ paraméter korrekt becslésével foglalkozik ARATÓ [2] dolgozata.

Legyen az $x(t)$ $(0 \leqq t \leqq T_0)$ valós, szigorú értelemben stacionárius folyamat 1 valószínűséggel folytonos és $\theta = Ex(t)$ a folyamat várható értéke ismeretlen, míg a $B(t)=E[x(s+t)-\theta][x(s)-\theta]$ kovariancia függvénye legyen ismert.

ARATÓ a korrekt becslés fogalmát a következőképpen definiálja:

A $\hat{\theta}(x(t))$ funkcionált a θ paraméter korrekt becslésének nevezzük, ha tetszőleges $-\infty < c < +\infty$-re

$$\hat{\theta}(x(t)+c) = \hat{\theta}(x(t))+c.$$

Bebizonyította, hogy az így definiált $\hat{\theta}_0 = x(0) - E_0(x(0)|x(t)-x(0),\ 0 \leqq t \leqq T_0)$ *Pitman-féle becslés* (a szórás létezését feltételezve) minimális szórású a korrekt becslések osztályában.

Az egydimenziós *Doob-féle elemi Gauss-stacionárius folyamatok* (lásd [6], [3]) esetére a *Pitman-féle becslés* megengedhetősége *Stein-féle* bizonyításának kiterjesztését szintén ARATÓ végezte el (lásd [2]).

Legyen az $x(t)$ stacionárius *Gauss-folyamat* $(n-1)$-szer differenciálható és elégítse ki a

$$dx^{(n-1)}(t) + \left[a_{n-1}x^{(n-1)}(t) + \ldots + a_0(x(t)-\theta)\right] dt = dw(t) \tag{1.14}$$

differenciálegyenletet, ahol $w(t)$ az ismert *Wiener-folyamat* $Edw(t)=0$, $E(dw)^2=dt$ paraméterekkel. Ismert, hogy $x(t)$ spektrál sűrűsége

$$\frac{1}{2\pi} \frac{1}{|(i\lambda)^n + a_{n-1}(i\lambda)^{n-1} + \ldots + a_0|^2}$$

alakú.

A sztochasztikus differenciálegyenletek tulajdonságain alapuló módszerek felhasználásával ARATÓ megmutatta, hogy az

$$m^* = \frac{\sum_{k=1}^{n-1} a_{k+1}[x^{(k)}(T_0) + (-1)^k x^{(k)}(0)] + a_0 \int_0^{T_0} x(t)\, dt}{2a_1 + a_0 T_0} \tag{1.15}$$

maximum likelihood becslés egyben *Pitman-féle becslés* is és az $E_\theta x(t)=\theta$ $(-\infty<\theta<+\infty)$ paraméternek megengedhető becslése.

Vizsgáljuk most azt az esetet, amikor a veszteségfüggvény

$$r(\tilde{\theta};\theta)=|\tilde{\theta}-\theta|$$

alakú, azaz *Laplace-féle veszteségfüggvény*. Az

$$\int_{R^n}|x_j|\,dF(x_1,\ldots,x_n)<\infty,\quad j=1,\ldots,n \tag{1.16}$$

feltétel mellett a *Pitman-féle becslés* a következő:

$$\bar{\theta}_0=l-\mathrm{med}_0(l|\mathbf{y}), \tag{1.17}$$

ahol $l=\sum_{j=1}^{n}c_jx_j\left(\sum_{j=1}^{n}c_j=1\right)$ lineáris statisztika, $\mathbf{y}=(x_2-x_1,\ldots,x_n-x_1)$ és $\mathrm{med}_0(l|\mathbf{y})$-on a feltételes eloszlás (adott $\mathbf{y}$ mellett az l statisztika P_0 eloszlásának megfelelő feltételes eloszlás) mediánjai közül azt a tetszőleges variánst értjük, amelyre $\mathrm{med}_0(l|\mathbf{y})$ statisztika.

Valóban, legyen $\hat{\theta}\in\mathcal{K}$ — azaz $\hat{\theta}$ tetszőleges korrekt becslés. Ekkor $\hat{\theta}=l+h(\mathbf{y})$ ($h(\mathbf{y})$ $\mathbf{y}$-nak valamely függvénye).

Figyelembe véve, hogy minden $\mathbf{y}$-ra

$$\min_{h(\mathbf{y})}E_0\{|l+h(\mathbf{y})||\mathbf{y}\}=E_0\{|l-\mathrm{med}_0(l|\mathbf{y})||\mathbf{y}\}$$

kapjuk az

$$E_\theta|\hat{\theta}-\theta|=E_\theta|l-\theta+h(\mathbf{y})|=E_0|l+h(\mathbf{y})|=E_0\{E_0(|l+h(\mathbf{y})||\mathbf{y})\}\geqq$$
$$\geqq E_0\{E_0(|l-\mathrm{med}_0(l|\mathbf{y})||\mathbf{y}\}=E_0|\bar{\theta}_0|=E_\theta|\bar{\theta}_0-\theta|$$

összefüggést. Tehát $\bar{\theta}_0$ optimális becslés a $\mathcal{K}$ osztályban a *Laplace-féle veszteségfüggvénnyel.*

Speciálisan, amikor $l=\bar{x}=\frac{1}{n}\sum_{j=1}^{n}x_j$, az (1.17) *Pitman-féle becslés*

$$\bar{\theta}_0=\bar{x}-\mathrm{med}_0(\bar{x}|\mathbf{y}) \tag{1.18}$$

alakú.

Már láttuk, hogy a *Pitman-féle becslés,* a négyzetes veszteségfüggvény, illetve *Laplace-féle veszteségfüggvény* esetén, a

$$\hat{\theta}_0=\bar{x}-E_0(\bar{x}|\mathbf{y}),$$

illetve a

$$\bar{\theta}_0=\bar{x}-\mathrm{med}_0(\bar{x}|\mathbf{y})$$

statisztika. Tehát, általában ugyanarra az $F(x)$ eloszlásfüggvényre a *Pitman-féle becslés* függ a veszteségfüggvény választásától. Normális eloszlás esetén azonban megmutatható (lásd 2. pont), hogy a négyzetes és a *Laplace-féle* (és sok más) *veszteség függvényre* a θ paraméter *Pitman-féle becslése* ugyanaz, mármint $\hat{l}$, a θ legjobb lineáris torzítatlan becslése. (Független azonos eloszlású megfigyelések esetén $\hat{l}=\bar{x}$, az empirikus közép.)

Tekintsük most a többdimenziós eltolási paraméter becslésének feladatát. Legyen $\mathbf{x}=\begin{pmatrix} x^{(1)} \\ \vdots \\ x^{(k)} \end{pmatrix}$ k-dimenziós valószínűségi vektorváltozó, amelynek eloszlásfüggvénye

$$F(\mathbf{x}-\boldsymbol{\theta}) = F(x^{(1)}-\theta^{(1)}, \ldots, x^{(k)}-\theta^{(k)})$$

függ a $\boldsymbol{\theta}=\begin{pmatrix} \theta^{(1)} \\ \vdots \\ \theta^{(k)} \end{pmatrix}$ k-dimenziós eltolási paramétertől. Ekkor a mintatér $\mathscr{H}=\mathscr{R}^{kn}=$ $=\mathscr{R}^k\times\ldots\times\mathscr{R}^k$, és a paramétertér $\Theta=\mathscr{R}^k$. Feltesszük, hogy a becsülendő paraméterfüggvény $t(\boldsymbol{\theta})=\boldsymbol{\theta}$, és $\tilde{\boldsymbol{\theta}}(x_1, \ldots, x_n)$ az $(x_1, \ldots, x_n)$ ismétléses mintán alapuló becslése $\boldsymbol{\theta}$-nak.

A $\tilde{\boldsymbol{\theta}}(x_1, \ldots, x_n)$ becslés jóságának mértékét a következő nemnegatív definit mátrix méri:

$$\mathbf{R}(\tilde{\boldsymbol{\theta}}; \boldsymbol{\theta}) = E_{\boldsymbol{\theta}}(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta})(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta})^T, \tag{1.19}$$

itt és a továbbiakban ^{T}-vel a transzponáltat jelöljük. Ha csak $\boldsymbol{\theta}$ torzítatlan becsléseinek osztályát vizsgáljuk, akkor $\mathbf{R}(\tilde{\boldsymbol{\theta}}; \boldsymbol{\theta})$ — a rizikó — nem más, mint $\boldsymbol{\theta}$ kovarianciamátrixa.

Az $\mathbf{R}(\tilde{\boldsymbol{\theta}}_1; \boldsymbol{\theta}) \leqq \mathbf{R}(\tilde{\boldsymbol{\theta}}_2; \boldsymbol{\theta}), \boldsymbol{\theta}\in\mathscr{R}^k$ reláció teljesülése, ahol $\tilde{\boldsymbol{\theta}}_1$, $\tilde{\boldsymbol{\theta}}_2$ becslései $\boldsymbol{\theta}$-nak, azt jelenti, hogy az $\mathbf{R}(\tilde{\boldsymbol{\theta}}_2; \boldsymbol{\theta})-\mathbf{R}(\tilde{\boldsymbol{\theta}}_1; \boldsymbol{\theta})$ mátrix pozitív szemidefinit.

E fogalmak bevezetése után a becslések megengedhetősége és optimalitása természetes módon értendő.

A korrekt becslések 1.4. definícióját triviális módon értelmezzük, feltételezve, hogy az (1.4) formula igaz minden $\mathbf{c}\in\mathscr{R}^k$-ra.

A *A Pitman-féle becslés* 1.5. definíciója változtatás nélkül érvényben marad.

Tegyük fel, hogy

$$\int_{R^k} x^{(i)}\, dF(x^{(1)}, \ldots, x^{(k)}) = 0, \quad i = 1, \ldots, k, \tag{1.20}$$

$$\int_{R^k} x^{(i)^2}\, dF(x^{(1)}, \ldots, x^{(k)}) < \infty, \quad i = 1, \ldots, k. \tag{1.21}$$

Legyen

$$\mathbf{y} = (x_2-x_1, \ldots, x_n-x_1) = (x_2^{(1)}-x_1^{(1)}, \ldots, x_n^{(k)}-x_1^{(k)})$$

$k(n-1)$ komponensű vektor. Legyen továbbá

$$\hat{\theta}_0 = \bar{\mathbf{x}} - E_0(\bar{\mathbf{x}}|\mathbf{y}), \tag{1.22}$$

ahol

$$\hat{\boldsymbol{\theta}}_0 = (\hat{\theta}_0^{(1)}, \ldots, \hat{\theta}_0^{(k)})^T, \quad \bar{\mathbf{x}} = \frac{1}{n}\sum_{j=1}^{n}\mathbf{x}_j = \left(\frac{1}{n}\sum_{j=1}^{n}x_j^{(1)}, \ldots, \frac{1}{n}\sum_{j=1}^{n}x_j^{(k)}\right)^T,$$

és az $E_0(\bar{\mathbf{x}}|\mathbf{y})$ feltételes várható értéket komponensenként értjük.

Érvényes a következő tétel, amely nem más, mint az 1.1. tétel többdimenziós analogonja, és megtalálható pl. [17]-ben.

1.2. Tétel. 1°. Az (1.20), (1.21) feltételek teljesülése esetén a

$$\hat{\boldsymbol{\theta}}_0 = \bar{\mathbf{x}} - E_0(\bar{\mathbf{x}}|\mathbf{y})$$

becslés *Pitman-féle becslés.*

2°. Ha

$$F(\mathbf{x}) = F(x^{(1)}, \ldots, x^{(k)}) = \int_{-\infty}^{x^{(1)}} \ldots \int_{-\infty}^{x^{(k)}} f(u_1, \ldots, u_k)\, du_1 \ldots du_k,$$

akkor

$$\hat{\theta}_0^{(i)} = \frac{\int \cdots \int \xi^{(i)} \prod_{j=1}^{n} f(x_j^{(1)} - \xi^{(1)}, \ldots, x_j^{(k)} - \xi^{(k)})\, d\xi^{(1)} \ldots d\xi^{(k)}}{\int \cdots \int \prod_{j=1}^{n} f(x_j^{(1)} - \xi^{(1)}, \ldots, x_j^{(k)} - \xi^{(k)})\, d\xi^{(1)} \ldots d\xi^{(k)}}, \quad i = 1, \ldots, k. \tag{1.23}$$

FINTUSAL [10] (1971) bebizonyította az (1.22)-ben szereplő *Pitman-féle becslés* abszolút megengedhetőségét. STEIN [50] gondolatmenetének kiterjesztésével megmutatta, hogy ha az $\mathbf{x}$ valószínűségi vektorváltozó $F(\mathbf{x})$ eloszlásfüggvénye abszolút folytonos (az $\mathscr{R}^k$-ban levő *Lebesgue-mértékre* nézve), és minden $i=1, \ldots, k$-ra teljesül a következő feltétel:

$$E_0\{E_0[[x_1^{(i)} - E_0(x_1^{(i)} - |x_2 - x_1, \ldots, x_n - x_1)]^{2k} | x_2 - x_1, \ldots, x_n - x_1]\}^{(k+2)/2k} < \infty, \tag{1.24}$$

akkor a *Pitman-féle becslés* abszolút megengedhető.

Érdekes megjegyezni, hogy ha a veszteségfüggvény az eltérések négyzetének összege, azaz

$$r(\tilde{\boldsymbol{\theta}}, \boldsymbol{\theta}) = (\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta})^T (\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \|\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2 = \sum_{j=1}^{k} [\tilde{\theta}^{(j)} - \theta^{(j)}]^2, \tag{1.25}$$

akkor a *Pitman-féle becslés* egyáltalában nem lesz megengedhető $k \geqq 3$ esetén.

Tekintsük STEIN [51], [52] következő példáját, amely megmutatja, hogy $k \geqq 3$ esetén az (1.25) veszteségfüggvénnyel a *Pitman-féle becslés* nem megengedhető.

Legyen $\mathbf{x} = (x^{(1)}, \ldots, x^{(k)})^T$ k-dimenziós normális eloszlású vektorváltozó, amelynek $\boldsymbol{\theta} = E\mathbf{x} = (\theta^{(1)}, \ldots, \theta^{(k)})^T$ várható értékvektora teljesen ismeretlen, a kovarianciamátrixa pedig egységmátrix: $E(\mathbf{x} - \boldsymbol{\theta})(\mathbf{x} - \boldsymbol{\theta})^T = \mathbf{I}$.

Tegyük fel, hogy a veszteségfüggvény (1.25) alakú.

Legyen $\hat{\boldsymbol{\theta}}_0$ az úgynevezett szokásos becslés, azaz a következő becslés:

$$\hat{\boldsymbol{\theta}}_0(\mathbf{x}) = \mathbf{x}, \tag{1.26}$$

amelynek rizikója

$$\mathbf{R}(\hat{\boldsymbol{\theta}}_0; \boldsymbol{\theta}) = E_{\boldsymbol{\theta}} r(\hat{\boldsymbol{\theta}}_0, \boldsymbol{\theta}) = E_{\boldsymbol{\theta}} (\hat{\boldsymbol{\theta}}_0 - \boldsymbol{\theta})^T (\hat{\boldsymbol{\theta}}_0 - \boldsymbol{\theta}) = k.$$

Ismeretes, hogy a torzítatlan becslések közül, vagy a korrekt becslések közül, a $\hat{\boldsymbol{\theta}}_0$ szokásos becslés minimális veszteségű minden $\boldsymbol{\theta}$-ra, tehát az 1.5. definíció szerint $\hat{\boldsymbol{\theta}}_0$ *Pitman-féle becslés.* Legyen

$$\hat{\boldsymbol{\theta}}_1(\mathbf{x}) = \left(1 - \frac{k-2}{\|\mathbf{x}\|^2}\right)\mathbf{x}, \tag{1.27}$$

ahol $\|\mathbf{x}\|^2=\mathbf{x}^T\mathbf{x}=\sum_{j=1}^{k} x^{(j)2}$. Megmutatható, hogy $k\geqq 3$ esetén

$$E_{\boldsymbol{\theta}}\left\|\left(1-\frac{k-2}{\|\mathbf{x}\|^2}\right)\mathbf{x}-\boldsymbol{\theta}\right\|^2 = k-E\frac{(k-2)^2}{k-2+2\zeta} < k, \quad \text{minden } \boldsymbol{\theta}\text{-ra,}$$

ahol a ζ változó $\frac{\|\boldsymbol{\theta}\|^2}{4}$ várható értékű *Poisson-eloszlású* valószínűségi változó, azaz a $\hat{\boldsymbol{\theta}}_0(\mathbf{x})$ szokásos becslés nem megengedhető.

$k=1$ esetén a $\hat{\boldsymbol{\theta}}_0$ szokásos becslés megengedhetősége jól ismert (lásd pl. LEHMANN—HODGES [14], GIRSHIK és SAVAGE [13], BLYTH [5]).

$k=2$ esetén $\hat{\boldsymbol{\theta}}_0$ megengedhetőségét STEIN [51] bebizonyította.

2. A normális eloszlás jellemzése az eltolás paramétere legjobb lineáris becslésének megengedhetőségével

Legyenek $x_1, \ldots, x_n$ független megfigyelések az $F_1(x-\theta), \ldots, F_n(x-\theta)$ eloszlásfüggvényekkel, amelyek a θ valós eltolási paramétertől függenek. Tegyük fel, hogy

$$\int x\,dF_j(x)=0, \quad j=1, \ldots, n, \tag{1.28}$$

$$0<\int x^2\,dF_j(x)=\sigma_j^2<\infty, \quad j=1, \ldots, n. \tag{1.29}$$

Az (1.28) feltételből látható, hogy $Ex_j=\theta$, $j=1, \ldots, n$. A további vizsgálatban a veszteségfüggvény legyen a paraméter igazi értékétől való eltérés négyzete, azaz $r(\tilde{\theta}, \theta)=|\tilde{\theta}-\theta|^2$. Ebben az esetben θ-nak az $(x_1, \ldots, x_n)$ mintán alapuló legjobb lineáris torzítatlan becslése az $\hat{l}=\sum_{j=1}^{n} c_j^0 x_j$ statisztika, ahol $c_j^0=\frac{1}{\sigma_j^2}\left(\sum_{j=1}^{n}\frac{1}{\sigma_j^2}\right)^{-1}$. Megmutatjuk, hogy $n\geqq 3$ esetén $\hat{l}$ megengedhetősége, θ torzítatlan becsléseinek osztályában, jellemző tulajdonsága a normális eloszlásoknak.

A következő tétel, amely nem más, mint közvetlen általánosítása KAGAN—LINNIK—RAO egy tételének [20], megtalálható [21]-ben.

A tétel szükséges feltételének bizonyítását, RAO egy általános tételének (lásd [43] 4. tétel) alapgondolatát alkalmazva, egy kissé módosítva adjuk.

1.3. Tétel. Tegyük fel, hogy $x_1, \ldots, x_n$ $(n\geqq 3)$ független megfigyelések, amelyek $F_1(x-\theta), \ldots, F_n(x-\theta)$ eloszlásfüggvényei eleget tesznek az (1.28) és (1.29) feltételeknek, $\theta\in R^1$.

Az $\hat{l}=\sum_{j=1}^{n} c_j^0 x_j$ legjobb lineáris torzítatlan becslés akkor és csak akkor megengedhető — a négyzetes veszteségfüggvénnyel — θ összes torzítatlan becslései osztályában, ha az $F_j(x)$ eloszlásfüggvények normálisak.

Bizonyítás 1°. *Szükségesség.*

Vegyük figyelembe, hogy az (1.7) *Pitman-féle becslés*

$$\hat{\theta}_0=\hat{l}-E_0(\hat{l}\,|\,x_2-x_1, \ldots, x_n-x_1)$$

alakban írható fel.

Az 1.1. tétel alapján az $\hat{l}$ becslés megengedhetőségéből adódik, hogy

$$E_0(\hat{l}|x_2-x_1, \ldots, x_n-x_1) = 0, \tag{1.30}$$

ahonnan

$$E_0\left(\hat{l}\exp i\sum_{j=2}^{n} t_j(x_j-x_1)\right) = E_0\left\{\exp\left(i\sum_{j=2}^{n} t_j(x_j-x_1)\right) E_0(\hat{l}|x_2-x_1, \ldots, x_n-x_1)\right\} = 0 \tag{1.31}$$

következik.

(1.31)-ből, az $f_j(t)=E_0e^{itx_j}$, $j=1, \ldots, n$, jelölésekkel a következő összefüggést kapjuk:

$$c_1^0 f_1'[-(t_2+\ldots+t_n)]\prod_{j=2}^{n} f_j(t_j)+c_2^0 f_2'(t_2)f_1[-(t_2+\ldots+t_n)]\prod_{j=3}^{n-1} f_j(t_j)+\ldots+ \tag{1.32}$$

$$+\ldots+c_k^0 f_k'(t_k)f_1[-(t_2+\ldots+t_n)]\prod_{\substack{j=2\\ j\neq k}}^{n} f_j(t_j)+\ldots+c_n^0 f_n'(t_n)f_1[-(t_2+\ldots+t_n)]\prod_{j=2}^{n} f_j(t_j)=0.$$

Mivel $f_j(t)$ karakterisztikus függvény, $j=1, \ldots, n$, létezik olyan $\delta>0$, hogy ha $|t_j|<\delta$, $j=2, \ldots, n$, akkor $f_1[-(t_2+\ldots+t_n)]\prod_{j=2}^{n} f_j(t_j)\neq 0$.

Legyen

$$g_j(t) = \frac{f_j'(t)}{f_j(t)}, \quad j = 1, \ldots, n,$$

ahol a $g_j(t)$ függvények a 0-pont egy környezetében vannak értelmezve. (1.32) mindkét oldalát $f_1[-(t_2+\ldots+t_n)]\prod_{j=2}^{n} f_j(t_j)$-val osztva, kapjuk a

$$c_1^0 g_1(-(t_2+\ldots+t_n))+c_2^0 g_2(t_2)+\ldots+c_n^0 g_n(t_n) = 0 \tag{1.33}$$

egyenletet. A t_r, t_s kivételével a $t_k=0$, $k=2, \ldots, n$ helyettesítéssel (1.32)-ből a

$$c_1^0 g_1(-(t_r+t_s))+c_r^0 g_r(t_r)+c_s^0 g_s(t_s) = 0 \tag{1.34}$$

egyenlethez jutunk.

Linnik—Rao tételét (lásd [33], [44]) alkalmazva és figyelembe véve, hogy $E_0(x_i)=0$, $i=1, \ldots, n$ miatt $g_i(0)=0$ minden i-re, kapjuk, hogy

$$g_i(t) = \alpha_i t, \quad i = 1, \ldots, n,$$

ahol $|t|<\varepsilon$, $\varepsilon>0$, eléggé kicsiny és α_i komplex szám. Ha $\alpha_i=0$ a probléma triviális. Ha $\alpha_i\neq 0$, akkor $f_i(t)=\exp\frac{\alpha_i}{2}t^2$ ($|t|<\varepsilon$). Az analitikus folytatás elvével belátható, hogy

$$f_i(t) = \exp\frac{\alpha_i}{2}t^2, \quad \text{minden } t\text{-re},$$

és mivel $f_i(t)$ karakterisztikus függvény, $\alpha_i<0$, valós. Tehát az x_i, $i=1, \ldots, n$, valószínűségi változók normális eloszlásúak.

1. Megjegyzés. Marcinkiewicz tételének (lásd pl. [36], [33] 64—65. o.) alkalmazásával a fenti bizonyítás végén az indoklást a következőképpen rövidíthetjük:

Már láttuk, hogy $g_j(t)=\frac{f_j'(t)}{f_j(t)}$ polinom, ahonnan következik, hogy $\log f_j(t)$ is polinom. *Marcinkiewicz tétele* szerint ekkor $f_j(t)$ normális eloszlás karakterisztikus függvénye.

2. Megjegyzés. $n=2$ esetén az állítás nem igaz. Valóban, vizsgáljuk azt az esetet, amikor x_1, x_2 független azonos eloszlású valószínűségi változók $F(x-\theta)$ eloszlásfüggvénnyel.

Legyen

$$f(t) = E\exp itx, \quad g(t) = \frac{f'(t)}{f(t)}.$$

Ebben az esetben az $\hat{l}=\bar{x}=\frac{x_1+x_2}{2}$ becslés megengedhetőségéből kapjuk az

$$\frac{f'(t_1)}{f(t_1)}+\frac{f'(-t_1)}{f(-t_1)}=0$$

egyenletet a 0-pont egy környezetében.

Ennek teljesüléséhez elegendő az $F(x-\theta)$ eloszlás szimmetrikus volta. Könnyen belátható, hogy tetszőleges szimmetrikus eloszlás és $n=2$ esetén az

$$E(\hat{l}|x_2-x_1, \ldots, x_n-x_1)=0$$

feltétel automatikusan teljesül, azaz $E(x_1+x_2|x_2-x_1)=0$.

2°. *Elégségesség.*

Tegyük fel, hogy $F_j(x-\theta)$ az $N(\theta, \sigma_j^2)$ normális eloszlásfüggvény. Így, az $x_1, \ldots, x_n$ valószínűségi változók együttes sűrűségfüggvénye a következő:

$$(1.35) \qquad p(x_1, \ldots, x_n; \theta) = \prod_{j=1}^{n}(2\pi\sigma_j^2)^{-1/2}\exp\left\{-\frac{1}{2}\sum_{j=1}^{n}\left(\frac{x_j-\theta}{\sigma_j}\right)^2\right\} =$$

$$= \prod_{j=1}^{n}(2\pi\sigma_j^2)^{-1/2}\exp\left\{-\frac{1}{2}\sum_{j=1}^{n}\frac{x_j^2}{\sigma_j^2}+\theta\sum_{j=1}^{n}\frac{1}{\sigma_j^2}x_j-\frac{\theta^2}{2}\sum_{j=1}^{n}\frac{1}{\sigma_j^2}\right\}.$$

Ebből belátható, hogy a $\sum_{j=1}^{n}\frac{1}{\sigma_j^2}x_j$, vagy az ezzel ekvivalens $\hat{l}=\sum_{j=1}^{n}c_j^0x_j$ statisztika elégséges a $p(x_1, \ldots, x_n; \theta)$ sűrűségfüggvények összességére nézve. Mivel $x_1, \ldots, x_n$ független normális eloszlású valószínűségi változók $Ex_j=\theta$, $D^2x_j=\sigma_j^2$, az $\hat{l}=\sum_{j=1}^{n}c_j^0x_j$ statisztika szintén normális eloszlású, $\theta=E\hat{l}$, $\sigma^2=D^2\hat{l}=\sum_{j=1}^{n}c_j^{02}\sigma_j^2$ paraméterekkel. Ily módon $\hat{l}$ sűrűségfüggvénye a következő:

$$\pi(\hat{l}; \theta) = \frac{1}{\sqrt{2\pi}\sigma}\exp-\frac{1}{2\sigma^2}(\hat{l}-\theta)^2.$$

A *Lehmann-tétel* (lásd [31]) alapján belátható, hogy $\hat{l}$ elégséges és egyben teljes is. A *Blackwell—Kolmogorov—Rao-tétel* (lásd BLACKWELL [4], KOLMOGOROV [29], RAO [45]) szerint $\hat{l}$ az egyetlen (P_0 m. m.) legjobb torzítatlan becslése θ-nak, amit bizonyítani akartunk.

Vegyük észre, hogy egyrészt definíció szerint a becslés optimalitásából következik a megengedhetősége, másrészt az 1.3. tétel elégségességének bizonyításában már láttuk, hogy normális eloszlású sokaság esetén az $\hat{l}$ legjobb lineáris becslés nemcsak megengedhető, hanem optimális is θ torzítatlan becsléseinek osztályában. Így az 1.3. tétel a következő ekvivalens módon fogalmazható meg:

1.3'. Tétel. Az 1.3. tételben mondott feltevések teljesülése esetén az $\hat{l}$ legjobb lineáris torzítatlan becslés optimalitása, θ torzítatlan becsléseinek osztályában, jellemző tulajdonsága a normális eloszlásnak.

Speciálisan, amikor $F_j(x)=F(x)$, $j=1, \ldots, n$, a most bizonyított tételből közvetlenül adódik az alábbi következmény, amely nem más mint az említett *Kagan—Linnik—Rao-tétel* egy másik megfogalmazása.

Következmény. Ha a független és azonos eloszlású $x_1, \ldots, x_n$ megfigyelések $F(x)$ eloszlásfüggvénye eleget tesz az (1.28), (1.29) feltételeknek, akkor az $\bar{x}=\frac{x_1+\ldots+x_n}{n}$ empirikus közép megengedhetősége, a θ eltolási paraméter torzítatlan becsléseinek osztályában, jellemző tulajdonsága a normális eloszlásnak.

Megemlítjük, hogy a normális eloszlásoknak van a fentinél erősebb jellemző tulajdonsága — az optimális lineáris becslés abszolút megengedhetősége. Ehhez kapcsolódóan bizonyítás nélkül kimondjuk a következő tételt, amelyet KAGAN (lásd [21]) bizonyított HODGES és LEHMANN [14] dolgozata módszerének alkalmazásával (azonos eloszlású megfigyelések esetén, lásd [17]).

1.4. Tétel. Az 1.3. tétel feltevései teljesülése esetén az $\hat{l}=\sum_{j=1}^{n} c_j^0 x_j$ legjobb lineáris becslés akkor és csak akkor abszolút megengedhető, ha $F_j(x)$ normális eloszlás, $j=1, \ldots, n$.

Ily módon, normális eloszlás esetén nyilvánvaló, hogy az $\hat{l}$ legjobb lineáris becslés megengedhető a korrekt becslések osztályában. Ismeretes, hogy abban az esetben, amikor a veszteségfüggvény $r(\tilde{\theta}, \theta)=r(\tilde{\theta}-\theta)$ csak $(\tilde{\theta}-\theta)$-tól függ, az optimalitás és a megengedhetőség ekvivalens egymással a korrekt becslések osztályában, ezért $\hat{l}$ legjobb korrekt becslés.

Normális eloszlás esetén tehát az $\hat{l}$ legjobb lineáris becslés és a *Pitman-féle becslés* ugyanaz. (Független azonos eloszlású megfigyelések esetén $\hat{l}=\bar{x}=\frac{1}{n}\sum_{j=1}^{n} x_j$ az empirikus közép.)

A négyzetes és a *Laplace-féle veszteségfüggvény* esetén a fenti állítás megfordítása is igaz. Ha ugyanis $\hat{l}$ *Pitman-féle becslés,* akkor (1.7)-ből, illetve (1.17)-ből következik, hogy $E_0(\hat{l}|\mathbf{y})=0$, illetve $\mathrm{med}_0(\hat{l}|\mathbf{y})=0$, amiből — mint az 1.3. tétel szükségességének bizonyításánál, illetve KAGAN—ZINGER [24] cikkében szereplő 1. tétel bizonyításánál látjuk — adódik, hogy az $x_1, \ldots, x_n$ megfigyelések normális eloszlásúak.

KAGAN [19] (1970) bevezette a becslések ε-megengedhetőségének fogalmát. Bebizonyította, hogy független azonos $F(x-\theta)$-eloszlású $x_1, \ldots, x_n$ megfigyelések

esetén, ha $F(-x)=1-F(x+0)$ és $\int x^2\,dF=\sigma^2<\infty$, akkor az $\bar{x}=\frac{1}{n}\sum_{j=1}^{n}x_j$ empirikus közép $\frac{\sigma^2\varepsilon^2}{n}$-megengedhető θ torzítatlan becsléseinek osztályában, azaz nem létezik olyan $\tilde{\theta}=\tilde{\theta}(x_1, ..., x_n)$ torzítatlan becslés, amelyre

$$E_\theta(\tilde{\theta}-\theta)^2 < E_\theta(\bar{x}-\theta)^2-\frac{\sigma^2\varepsilon^2}{n}, \quad \text{minden} \quad \theta\in\mathscr{R}^1\text{-re.}$$

Ez az eredmény nem más mint *Kagan—Linnik—Rao tételének* [20] egy kiterjesztése.

Vizsgáljuk a feladatot a többdimenziós eltolási paraméter esetén, mégpedig feltesszük, hogy $\boldsymbol{\theta}=(\theta^{(1)}, ..., \theta^{(k)})^T\in\mathscr{R}^k$ k-dimenziós vektorparaméter.

Legyen $(\mathbf{x}_1, ..., \mathbf{x}_n)$, ahol $\mathbf{x}_j=(x_j^{(1)}, ..., x_j^{(k)})^T$, $j=1, ..., n$, az $F(\mathbf{x}-\boldsymbol{\theta})=F(x^{(1)}-\theta^{(1)}, ..., x^{(k)}-\theta^{(k)})$ eloszlásfüggvényű k-dimenziós sokaságból vett ismétléses minta. Legyen továbbá $\tilde{\boldsymbol{\theta}}(x_1, ..., x_n)=(\tilde{\theta}^{(1)}, ..., \tilde{\theta}^{(k)})^T$ egy becslése $\boldsymbol{\theta}$-nak. A veszteségfüggvény és a rizikó az (1.19) formula segítségével definiálható, azaz $\mathbf{r}(\tilde{\boldsymbol{\theta}}, \boldsymbol{\theta})=(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta})(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta})^T$ és $\mathbf{R}(\tilde{\boldsymbol{\theta}}; \boldsymbol{\theta})=E_\theta(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta})(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta})^T$.

Többdimenziós esetben igaz az 1.3. tétel következő analogonja, amely megtalálható pl. [17]-ben és [18]-ban.

1.5. Tétel. $n\geqq 3$ esetén az empirikus közép vektor $\bar{\mathbf{x}}=(\bar{x}^{(1)}, ..., \bar{x}^{(k)})^T$ akkor és csak akkor megengedhető a $\boldsymbol{\theta}$ k-dimenziós eltolási parméter torzítatlan osztályában, ha $F(\mathbf{x})$ normális eloszlás.

Bizonyítás. 1°. *Szükségesség.*

Ismeretes, hogy az 1.2. tétel szerint a $\hat{\boldsymbol{\theta}}_0$ *Pitman-féle becslés* a következő alakú:

$$\hat{\boldsymbol{\theta}}_0 = \bar{\mathbf{x}}-E_0(\bar{\mathbf{x}}|\mathbf{y}),$$

ahol $\mathbf{y}=(x_2-x_1, ..., x_n-x_1)=(x_2^{(1)}-x_1^{(1)}, ..., x_n^{(k)}-x_1^{(k)})$. Belátható, hogy

$$\begin{aligned}(1.36)\qquad E_\theta(\bar{\mathbf{x}}-\boldsymbol{\theta})(\bar{\mathbf{x}}-\boldsymbol{\theta})^T &= E_\theta\big(\hat{\boldsymbol{\theta}}_0-\boldsymbol{\theta}+E_0(\bar{\mathbf{x}}|\mathbf{y})\big)\big(\hat{\boldsymbol{\theta}}_0-\boldsymbol{\theta}+E_0(\bar{\mathbf{x}}|\mathbf{y})\big)^T = \\ &= E_\theta(\hat{\boldsymbol{\theta}}_0-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}_0-\boldsymbol{\theta})^T+E_\theta\big\{(\hat{\boldsymbol{\theta}}_0-\boldsymbol{\theta})\big(E_0(\bar{\mathbf{x}}|\mathbf{y})\big)^T\big\}+E_\theta\{E_0(\bar{\mathbf{x}}|\mathbf{y})(\hat{\boldsymbol{\theta}}_0-\boldsymbol{\theta})^T\}+ \\ &+E_\theta\big\{E_0(\bar{\mathbf{x}}|\mathbf{y})\big(E_0(\bar{\mathbf{x}}|\mathbf{y})\big)^T\big\} = E_\theta(\hat{\boldsymbol{\theta}}_0-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}_0-\boldsymbol{\theta})^T+E_0\{\hat{\boldsymbol{\theta}}_0E_0(\bar{\mathbf{x}}|\mathbf{y})^T\}+ \\ &+E_0\{E_0(\bar{\mathbf{x}}|\mathbf{y})\hat{\boldsymbol{\theta}}_0^T\}+E_0\{E_0(\bar{\mathbf{x}}|\mathbf{y})E_0(\bar{\mathbf{x}}|\mathbf{y})^T\}.\end{aligned}$$

Minthogy $\hat{\boldsymbol{\theta}}_0$ nem más, mint $\bar{\mathbf{x}}$ vetülete a $h(\mathbf{x}_2-\mathbf{x}_1, ..., \mathbf{x}_n-\mathbf{x}_1)$ mérhető függvények alterére, ennélfogva

$$E_0(\hat{\boldsymbol{\theta}}_0E_0(\bar{\mathbf{x}}|\mathbf{y})^T = E_0(E_0(\bar{\mathbf{x}}|\mathbf{y})\hat{\boldsymbol{\theta}}_0^T) = 0.$$

Ily módon (1.36)-ból kapjuk az

$$(1.37)\qquad E_\theta(\bar{\mathbf{x}}-\boldsymbol{\theta})(\bar{\mathbf{x}}-\boldsymbol{\theta})^T = E_\theta(\hat{\boldsymbol{\theta}}_0-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}_0-\boldsymbol{\theta})^T+E_0\{E_0(\bar{\mathbf{x}}|\mathbf{y})E_0(\bar{\mathbf{x}}|\mathbf{y})^T\}$$

összefüggést.

Mivel az $E_0\{(\bar{\mathbf{x}}|\mathbf{y})E_0(\bar{\mathbf{x}}|\mathbf{y})^T\}$ mátrix pozitív szemidefinit, (1.37)-ből belátható, hogy

$$E_\theta(\bar{\mathbf{x}}-\boldsymbol{\theta})(\bar{\mathbf{x}}-\boldsymbol{\theta})^T \geqq E_\theta(\hat{\boldsymbol{\theta}}_0-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}_0-\boldsymbol{\theta})^T, \tag{1.38}$$

azonosság minden $\boldsymbol{\theta}$-ra akkor és csak akkor áll fenn, ha

$$E_0(\bar{\mathbf{x}}|\mathbf{x}_2-\mathbf{x}_1, \ldots, \mathbf{x}_n-\mathbf{x}_1) = \mathbf{0}. \tag{1.39}$$

Tehát $\hat{\boldsymbol{\theta}}_0$ mindig jobb, mint $\bar{\mathbf{x}}$ az empirikus középvektor, azon eset kivételével, amikor (1.39) teljesül.

Legyenek $\beta_1, \ldots, \beta_k$ tetszőleges valós számok. Tekintsük a

$$z_j = \beta_1 x_j^{(1)} + \ldots + \beta_k x_j^{(k)}, \quad j = 1, \ldots, n,$$

valószínűségi változókat.

(1.39)-ből következik, hogy

$$E_0(\bar{z}|\mathbf{x}_2-\mathbf{x}_1, \ldots, \mathbf{x}_n-\mathbf{x}_1) = 0,$$

ahol $\bar{z}=\dfrac{z_1+\ldots+z_n}{n}$, és még inkább

$$E_0(\bar{z}|z_2-z_1, \ldots, z_n-z_1) = 0. \tag{1.40}$$

Az 1.3. tétel bizonyításából belátható, hogy $n\geqq 3$ esetén az (1.40) reláció teljesülése ekvivalens a z_i valószínűségi változók normalitásával.

Ismeretes, hogy ebben az esetben az $\mathbf{x}_j=(x_j^{(1)}, \ldots, x_j^{(k)})^T$ $j=1, \ldots, n$, valószínűségi vektorváltozók szintén normálisak. Így, bebizonyítottuk, hogy ha $\bar{\mathbf{x}}$ megengedhető (a torzítatlan becslések osztályában), akkor $F(\mathbf{x})$ normális eloszlás.

2°. *Elégségesség.*

Tegyük fel, hogy $F(\mathbf{x})$ normális eloszlás. Az 1.3. tétel elégségességének bizonyításához hasonlóan könnyű megmutatni, hogy az $\bar{\mathbf{x}}$ elégséges statisztika egyben teljes is az

$$F(\mathbf{x}_1-\boldsymbol{\theta})\ldots F(\mathbf{x}_n-\boldsymbol{\theta})$$

eloszláscsaládra nézve. A *Blackwell—Kolmogorov—Rao-tétel* többdimenziós variánsából (lásd pl. [35]) adódik, hogy $\bar{\mathbf{x}}$ optimális a $\boldsymbol{\theta}$ összes torzítatlan becsléseinek osztályában. Ezzel a tétel bizonyítását befejeztük.

Többdimenziós normális eloszlások esetén szintén igaz az $\bar{\mathbf{x}}$ empirikus középvektor abszolút megengedhetősége, és ez jellemző tulajdonsága a normális eloszlásoknak.

KAGAN [17] bebizonyította a következő tételt, amelyet bizonyítás nélkül mondunk ki:

1.6. Tétel. Ha $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ egy k-dimenziós sokaságból vett minta $\Phi(x^{(1)}-\theta^{(1)}, \ldots, x^{(k)}-\theta^{(k)})$ normális eloszlásfüggvénnyel, akkor az $\bar{\mathbf{x}}=(\bar{x}^{(1)}, \ldots, \bar{x}^{(k)})^T$ empirikus középvektor abszolút megengedhető becslése a $\boldsymbol{\theta}=(\theta^{(1)}, \ldots, \theta^{(k)})^T$ paraméternek.

A nem kvadratikus, hanem valamilyen általános alakú veszteségfüggvény esetét az utóbbi években sok szerző vizsgálta.

Az $r(\tilde{\theta}, \theta)=|\tilde{\theta}-\theta|$ *Laplace-féle veszteségfüggvény* esetére KAGAN és ZINGER [24] megmutatta, hogy ha $x_1, \ldots, x_n$ $(n \geqq 6)$ az $F(x-\theta)$-eloszlású sokaságból vett független minta, és $F(x)$-re teljesülnek a következő feltételek:

(i) $F(x)$ sűrűségfüggvénye létezik és $f(x)=F'(x)$ folytonosan differenciálható,
(ii) $F(x)$ egycsúcsos, azaz létezik olyan x_0, hogy $f'(x) \geqq 0$, ha $x \leqq x_0, f'(x) \leqq 0$, ha $x \geqq x_0$,

akkor az $\bar{x}$ empirikus közép abszolút megengedhetőségéből következik, hogy $F(x)$ normális eloszlás.

Az állítás megfordítása is igaz, (sőt az $n \geqq 6$ feltétel nélkül). Ez következik FOX és RUBIN [11] általános tételeiből, (amelyek a *Pitman-féle becslés* — a *Laplace-féle veszteségfüggvénnyel* — megengedhetőségéről szólnak), vagy bebizonyítható FARREL [7] dolgozata módszerének felhasználásával.

Az alább felsorolt veszteségfüggvények esetén, hasonló eredmények találhatók KAGAN—ZINGER [24], FOX—RUBIN [11], ZINGER—KAGAN—KLEBANOV [54], KLEBANOV [26], KAGAN [17] és JOSHI [16] dolgozatában:

$$r(\tilde{\theta}; \theta) = \begin{cases} -\alpha(\tilde{\theta}-\theta), & \tilde{\theta} \leqq \theta, \\ \beta(\tilde{\theta}-\theta), & \tilde{\theta} \geqq \theta, \end{cases} \tag{1.41}$$

ahol $\alpha>0, \beta>0$ állandók. (Speciálisan, ha $\alpha=\beta=1$, $r(\tilde{\theta}, \theta)=|\tilde{\theta}-\tilde{\theta}|$, azaz a *Laplace-féle veszteségfüggvény* adódik);

$$r(\tilde{\theta}, \theta) = |\tilde{\theta}-\theta|^{2m+1}, \tag{1.42}$$

ahol m természetes szám;

$$r(\tilde{\theta}, \theta) = Q((\tilde{\theta}-\theta)^2), \tag{1.43}$$

ahol $Q(u)$ m-edfokú pozitív együtthatójú polinom;

$$r(\tilde{\theta}, \theta) = \begin{cases} 0, & |\tilde{\theta}-\theta| \leqq b, \\ 1, & |\tilde{\theta}-\theta| > b, \end{cases} \tag{1.44}$$

ahol b állandó (ez a konfidenciai becslésnek megfelelő veszteségfüggvény).

Nemrég FIEGER—WERNER [9] (1971) és KLEBANOV [27] (1973) a feladatot elég általános veszteségfüggvényekkel vizsgálta.

Legyen $x_1, \ldots, x_n$ $(n \geqq 3)$ az $F(x-\theta)$-eloszlású sokaságból vett független minta, legyen továbbá $f(x)$ sűrűségfüggvénye $F(x)$-nek (a *Lebesgue-mértékre* nézve). Tegyük fel, hogy a veszteségfüggvény a következő alakú:

$$r(\tilde{\theta}, \theta) = w(\tilde{\theta}-\theta), \tag{1.45}$$

ahol $w(x)$ konvex, páros, folytonosan differenciálható függvény, $(w(x) \geqq 0, w(0)=0, w'(x)=0 \Leftrightarrow x=0)$.

KLEBANOV bebizonyította, hogy az $f(x)$ sűrűség függvényre vonatkozó néhány további feltétel teljesülése esetén az $\bar{x}$ empirikus közép megengedhetőségéből — az (1.45) veszteségfüggvény esetén, a korrekt becslések osztályában — következik, hogy $f(x)$ *Gauss-féle sűrűségfüggvény.*

Legyen most $\hat{\theta}(x_1, ..., x_n)$ egy korrekt becslés és $\mathscr{D}_w$ a következő alakú veszteségfüggvények egy sokasága:

$$r(\hat{\theta}, \theta) = \max [0, w(|\hat{\theta}-\theta|-\gamma)], \quad \gamma \in (0, \infty). \tag{1.46}$$

FIEGER—WERNER megmutatta, hogy ha $w(x)$ konvex, differenciálható függvény és $w(x)>w(0)=0$, minden $x>0$-re, továbbá

$$E_0 w'(|x_i|) < \infty,$$

$$E_\theta w(|x_i|) < \infty, \quad \text{ahol } |\theta| \text{ elég kicsiny},$$

akkor az $\bar{x}$ empirikus közép — mint θ becslése — optimalitása, minden $\mathscr{D}_w$-hez tartozó $r(o, o)$ veszteségfüggvény esetén, jellemző tulajdonsága az $F(x)$ normális eloszlásnak.

Ez az eredmény visszavezethető a négyzetes veszteségfüggvény esetére.

Az utóbbi időben az eltolási paraméter vizsgálatában jelentősek azok a valószínűségi feladatok, amelyek csoportokhoz és topologikus struktúrákhoz kapcsolódnak.

Legyen $(\mathscr{X}, \mathscr{A}, P)$ valószínűségi mező. Tegyük fel, hogy $\mathscr{X}$ lokálisan kompakt *Abel-csoport,* amely eleget tesz a második megszámlálhatósági axiómának, $\mathscr{A}$ pedig $\mathscr{X}$ részhalmazaiból alkotott σ-algebra.

Építsük fel a P mértékből kiindulva a $\{P_\theta, \theta \in \mathscr{X}, P_\theta(A)=P(A-\theta), A \in \mathscr{A}\}$ θ paramétertől függő mértékek családját. Az ilyen θ paramétert a csoport paraméterének (vagy eltolás paraméterének) nevezzük.

A matematikai statisztika olyan klasszikus problémáiból, mint a valós eltolási paraméter és skálaparaméter becsléselméletéből természetes módon adódnak azok a feladatok, amelyek a θ csoportparaméter becslésével foglalkoznak egy $\mathbf{x}=(x_1, ..., x_n)$ minta alapján.

A következőkben megmutatjuk RUHIN [48] eredményeit, amelyek θ becsléseinek megengedhetőségéhez és a normális eloszlások jellemzéséhez kapcsolódnak.

Először megemlítjük a csoportokon adott mértékek harmonikus analízisének néhány fogalmát. Jelöljük $\mathscr{Y}$-nal az $\mathscr{X}$ csoport karaktereinek csoportját, azaz a valós számok moduló 2π vett additív csoportjába való folytonos homomorfizmusok csoportját.

Ily módon, ha $\eta \in \mathscr{Y}$, $x, z \in \mathscr{X}$, akkor $e^{i\eta(x+z)}=e^{i\eta(x)}e^{i\eta(z)}$. A dualitáselmélet alapján az $\mathscr{X}$ és $\mathscr{Y}$ közötti kapcsolat teljesen szimmetrikus, tehát $\mathscr{X}$ az $\mathscr{Y}$ csoport karaktereinek csoportja.

Legyen μ mérték az $\mathscr{X}$ csoporton. μ karakterisztikus függvényének, vagy *Fourier-transzformáltjának* nevezzük azt a $\hat{\mu}(\eta)$ függvényt, amelyet a következő reláció határoz meg:

$$\hat{\mu}(\eta) = \int_{\mathscr{X}} e^{i\eta(t)}\, d\mu(t). \tag{1.47}$$

A normális eloszlások definícióját csoportokon a következő módon szokás megadni (lásd [37]).

Az $\mathscr{X}$ lokálisan kompakt Abel-csoporton értelmezett P mértéket akkor nevezzük normálisnak, ha ennek *Fourier-transzformáltja*

$$\hat{p}(\eta) = \exp \{i\eta(x_0)-\varphi(\eta)\}$$

alakú, ahol x_0 valamely fix eleme $\mathscr{X}$-nek, $\varphi(\eta)$ folytonos nemnegatív függvény $\mathscr{Y}$-on, amelyre teljesül a

$$\varphi(\xi+\eta)+\varphi(\xi-\eta) = 2[\varphi(\xi)+\varphi(\eta)]$$

egyenlőség minden $\xi, \eta \in \mathscr{Y}$-ra.

Tegyük fel, hogy az $r(\tilde{\theta}, \theta)$ veszteségfüggvény eleget tesz az invariancia

$$r(\tilde{\theta}+c, \theta+c) = r(\tilde{\theta}, \theta) \tag{1.48}$$

feltételének. Belátható, hogy $r(\tilde{\theta}, \theta)=r(\tilde{\theta}-\theta, 0)=r(\tilde{\theta}-\theta)$, azaz a veszteségfüggvény csak a paraméter igazi értéke és ennek becslése közötti eltéréstől függ.

Az (1.48) invariáns veszteségfüggvény esetén természetes megvizsgálni az olyan $\hat{\theta}(x_1, \ldots, x_n)$ becslések osztályát, amelyekre fennáll

$$\hat{\theta}(x_1+c, \ldots, x_n+c) = \hat{\theta}(x_1, \ldots, x_n)+c$$

minden $\mathbf{x}=(x_1, \ldots, x_n) \in \mathscr{X}^n$-re és $c \in \mathscr{X}$-re (a *Haar-mértékre* m. m.). Az Ilyen becsléseket, RUHIN [48] nyomán, homogén additív becsléseknek nevezzük. Nyilvánvaló, hogy az (1.48) veszteségfüggvény esetén a homogén becslések osztályában az optimalitás és a megengedhetőség ekvivalensek egymással.

A következő eredmények az

$$r(\hat{\theta}, \theta) = E_\theta |e^{i\eta(\theta(x))} - e^{i\eta(\theta)}|^2 \tag{1.49}$$

veszteségfüggvényre vonatkoznak. Akkor mondjuk, hogy a $\hat{\theta}_0$ becslés nem rosszabb, mint $\hat{\theta}_1$ becslés, ha az $\mathscr{Y}$ csoport 0 elemének valamely az egész $\mathscr{Y}$ csoportot generáló $\mathscr{V}$ környezetének minden η karakterére és minden $\theta \in \mathscr{X}$-re teljesül

$$E_\theta |e^{i\eta(\theta_0(x))} - e^{i\eta(\theta)}|^2 \leqq E_\theta |e^{i\eta(\theta_1(x))} - e^{i\eta(\theta)}|^2. \tag{1.50}$$

Az előző pontokban megmutattuk, hogy abban az esetben, amikor a mintatér $\mathscr{X}=\mathscr{R}^1$, a paramétertér $\Theta=\mathscr{X}=\mathscr{R}^1$, azaz $\mathscr{X}$ az összes valós számok additív csoportja, akkor az $\bar{x}=\frac{1}{n}\sum_{j=1}^{n} x_j$ empirikus közép megengedhetősége jellemző tulajdonsága a normális eloszlásoknak.

Ennek analogonjai a csoportparaméter becslése esetén megtalálhatók RUHIN [48] említett dolgozatában.

RUHIN megmutatta, hogy ha $\mathscr{X}$ *torziómentes teljes Abel-csoport,* és P abszolút folytonos szimmetrikus *Gauss-féle mérték* $\mathscr{X}$-en, akkor $\bar{x}=\frac{1}{n}\sum_{j=1}^{n} x_j$ — mint a csoportparaméter becslése — optimális (1.50) értelemben a homogén becslések $\mathscr{H}$ osztályában.

Megfordítva, $n \geqq 3$ esetén, ha $\bar{x}$ optimális a $\mathscr{H}$ osztályban, akkor a P mérték normális.

Bebizonyította továbbá, hogy a P mérték abszolút folytonossága esetén a $\sum_{j=1}^{n} x_j$ ($n \geqq 3$) becslés akkor és csak akkor abszolút megengedhető becslése az $n\theta$ paraméternek, ha P normális mérték.

II. FEJEZET

1. Gauss-folyamatok jellemzése az eltolási paraméter legjobb lineáris becslésének megengedhetőségével

Vizsgáljuk az

$$x(j) = \xi(j)+\theta, \quad j = 1, 2, \ldots, n \tag{2.1}$$

folyamatot, ahol a $\xi(1), \ldots, \xi(n)$ valószínűségi változók (általában nem függetlenek) eleget tesznek az $E\xi(j)=0$, $E\xi^2(j)<\infty$ feltételeknek. Legyen a θ paraméter ($\theta\in\mathscr{R}^1$) legjobb lineáris torzítatlan becslése az $\hat{l}=\sum_{j=1}^{n} c_j^0 x(j)\left(\sum_{j=1}^{n} c_j^0=1\right)$ statisztika, ahol c_j^0 a $\xi(j)$ folyamat kovariancia függvénye segítségével meghatározható.

Felmerül a kérdés, hogy az $\hat{l}$ becslés megengedhetősége a θ torzítatlan becslései osztályában jellemző-e *Gauss-folyamatokra?* A. M. KAGAN, JU. V. LINNIK, C. R. RAO a [21] könyvben vizsgálják ezt a problémát abban az esetben, amikor $\xi(j)$ elsőrendű autoregressziós folyamat. Az eredmény a következő (lásd [21], 341. o.):

Legyenek az $x(1), \ldots, x(n)$, $n\geqq 3$ *megfigyelések* (2.1) *alakúak, és* $\xi(j)$ *a következő elsőrendű autoregressziós folyamat:*

$$\xi(1) = \varepsilon(1) \tag{2.2}$$

$$\xi(j) = a\xi(j-1)+\varepsilon(j), \quad j = 2, \ldots, n,$$

ahol az $\varepsilon(1), \ldots, \varepsilon(n)$ *valószínűségi változók függetlenek* $F_j(x)$ *eloszlással (az* $F_j(x)$ *eloszlások nem szükségképpen azonosak). Ha* $a\neq 1$, *akkor az* $\hat{l}=\sum_{j=1}^{n} c_j^0 x(j)$ *becslés megengedhetősége a* θ *torzítatlan becslései osztályában jellemző tulajdonsága az* $\varepsilon(1), \ldots, \varepsilon(n)$ *Gauss-féle valószínűségi változóknak, egyben az* $x(j)$ *Gauss-folyamatnak.*

A következőkben a $\xi(j)$ p-edrendű autoregressziós folyamat esetén vizsgáljuk a problémát. KAGAN—LINNIK—RAO gondolatmenetének kiterjesztésével megmutatjuk, hogy bizonyos egyszerű feltételek mellett az elsőrendű autoregressziós sémára vonatkozó eredmény általánosítható.

Megemlítjük, hogy a továbbiakban, ha nincs külön megjegyzés, a tárgyalás csak a négyzetes veszteségfüggvényhez kapcsolódik.

Legyen $\xi(t)$ stacionárius folyamat és elégítse ki a

$$\xi(t)+a_1\xi(t-1)+\ldots+a_p\xi(t-p) = \varepsilon(t) \tag{2.3}$$

sztochasztikus differenciaegyenletet, ahol $\varepsilon(t)$ $(E\varepsilon(t)=0, E\varepsilon^2(t)=\sigma_\varepsilon^2)$ azonos eloszlású független sorozat, melyre $\varepsilon(t)$ független $\xi(t-1)$, $\xi(t-2)$, ...,-től is. Legyen továbbá

$$x(j) = \xi(j)+\theta, \quad j = 1, \ldots, n.$$

Tegyük fel, hogy $x(j)$ *Gauss-folyamat.* $n\geqq 2p+1$ esetén $x(1), \ldots, x(n)$ együttes sűrűségfüggvényét a következőképpen határozhatjuk meg.

Figyelembe véve, hogy a $\xi(1), \ldots, \xi(p)$ változók függetlenek az $\varepsilon(p+1), \ldots, \varepsilon(n)$ változóktól, kapjuk, hogy

$$p_{\xi(1),\ldots,\xi(p),\varepsilon(p+1),\ldots,\varepsilon(n)}(u_1, \ldots, u_p, z_{p+1}, \ldots, z_n) = \tag{2.4}$$

$$= (2\Pi)^{-n/2}|\mathbf{R}_p|^{-1/2}\sigma_\varepsilon^{-(n-p)} \exp\left\{-\frac{1}{2}\left(\mathbf{U}_p^T\mathbf{R}_p^{-1}\mathbf{U}_p + \frac{1}{\sigma_\varepsilon^2}\sum_{i=p+1}^{n} z_i^2\right)\right\},$$

ahol $\mathbf{R}_p$ jelöli a $\xi(1), \ldots, \xi(p)$ változók kovarianciamátrixát, $\mathbf{R}_p^{-1}$ ennek inverze, és $\mathbf{U}_p = \begin{pmatrix} u_1 \\ \vdots \\ u_p \end{pmatrix}$, $\mathbf{U}_p^T = (u_1, \ldots, u_p)$ $\mathbf{U}_p$-nek transzponáltja.

Felhasználva a

$$\begin{aligned} &\xi(1) = \xi(1) \\ &\cdots\cdots\cdots \\ &\xi(p) = \xi(p) \\ &\xi(p+1) + a_1\xi(p) + \ldots + a_p\xi(1) = \varepsilon(p+1) \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ &\xi(n) + a_1\xi(n-1) + \ldots + a_p\xi(n-p) = \varepsilon(n) \end{aligned} \tag{2.5}$$

leképezést, melynek determinánsa 1, (2.4)-ből belátható, hogy

$$p_{\xi(1),\ldots,\xi(n)}(u_1, \ldots, u_n) = \tag{2.6}$$

$$= (2\Pi)^{-n/2}|\mathbf{R}_p|^{-1/2}\sigma_\varepsilon^{-(n-p)} \exp\left\{-\frac{1}{2}\left[\mathbf{U}_p^T\mathbf{R}_p^{-1}\mathbf{U}_p + \frac{1}{\sigma_\varepsilon^2}\sum_{i=p+1}^{n}(u_i + a_1u_{i-1} + \ldots + a_pu_{i-p})^2\right]\right\}.$$

Innen $x(1), \ldots, x(n)$ együttes sűrűségfüggvénye a következő:

$$p(x_1, \ldots, x_n; \theta) = (2\Pi)^{-n/2}|\mathbf{R}_p|^{-1/2}\sigma_\varepsilon^{-(n-p)} \exp - \tag{2.7}$$

$$-\frac{1}{2}\left\{(\mathbf{X}_p - \boldsymbol{\theta})^T\mathbf{R}_p^{-1}(\mathbf{X}_p - \boldsymbol{\theta}) + \frac{1}{\sigma_\varepsilon^2}\sum_{i=p+1}^{n}[(x_i - \theta) + a_1(x_{i-1} - \theta) + \ldots + a_p(x_{i-p} - \theta)]^2\right\},$$

ahol $\mathbf{X}_p - \boldsymbol{\theta} = (x_1 - \theta, \ldots, x_p - \theta)^T$.

Ismeretes (lásd Arató [1]), hogy ($a_0 = 1$)

$$\mathbf{R}_p^{-1} = \frac{1}{\sigma_\varepsilon^2}\begin{pmatrix} a_0^2 & a_0a_1 & a_0a_2 & \ldots & a_0a_{p-1} \\ a_0a_1 & a_0^2 + a_1^2 & a_0a_1 + a_1a_2 & \ldots & a_0a_{p-2} + a_1a_{p-1} \\ a_0a_2 & a_0a_1 + a_1a_2 & a_0^2 + a_1^2 + a_2^2 & \ldots & \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_0a_{p-1} & a_0a_{p-2} + a_1a_{p-1} & & \ldots & a_0^2 + a_1^2 + \ldots + a_{p-1}^2 \end{pmatrix}. \tag{2.8}$$

Innen

$$(\mathbf{X}_p - \boldsymbol{\theta})^T\mathbf{R}_p^{-1}(\mathbf{X}_p - \boldsymbol{\theta}) = \frac{1}{\sigma_\varepsilon^2}[a_0^2(x_1 - \theta)^2 + a_0a_1(x_1 - \theta)(x_2 - \theta) + \ldots +$$

$$+ \ldots + a_0a_{p-1}(x_1 - \theta)(x_p - \theta) + a_0a_1(x_2 - \theta)(x_1 - \theta) + (a_0^2 + a_1^2)(x_2 - \theta)^2 + \ldots +$$

$$+\ldots+(a_0a_{p-2}+a_1a_{p-1})(x_2-\theta)(x_p-\theta)+\ldots+(a_0^2+a_1^2+\ldots+a_{p-1}^2)(x_p-\theta)^2] =$$

$$= -\frac{2\theta}{\sigma_\varepsilon^2}\left[x_1(a_0^2+a_0a_1+\ldots+a_0a_{p-1})+x_2(a_0a_1+(a_0^2+a_1^2)+\ldots+(a_0a_{p-2}+a_1a_{p-1}))+\ldots\right.$$

$$\left.\ldots+x_p(a_0a_{p-1}+(a_0a_{p-2}+a_1a_{p-1})+\ldots+(a_0^2+a_1^2+\ldots+a_{p-1}^2))\right]+f_1(x_1,\ldots,x_n)+g_1(\theta),$$

ahol $f_1(x_1,\ldots,x_n)$ és $g_1(\theta)$ értelmezése leolvasható.

Ily módon,

$$(2.9)\quad (\mathbf{X}_p-\boldsymbol{\theta})^T\mathbf{R}_p^{-1}(\mathbf{X}_p-\boldsymbol{\theta}) = -\frac{2\theta}{\sigma_\varepsilon^2}(\alpha_1x_1+\alpha_2x_2+\ldots+\alpha_px_p)+f_1(x_1,\ldots,x_n)+g_1(\theta),$$

ahol

$$(2.10)\qquad \begin{aligned}
\alpha_1 &= a_0^2+a_0a_1+\ldots+a_0a_{p-1}\\
\alpha_2 &= a_0a_1+(a_0^2+a_1^2)+\ldots+(a_0a_{p-2}+a_1a_{p-1})\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
\alpha_p &= a_0a_{p-1}+(a_0a_{p-2}+a_1a_{p-1})+\ldots+(a_0^2+a_1^2+\ldots+a_{p-1}^2).
\end{aligned}$$

Másrészt

$$(2.11)\qquad \sum_{i=p+1}^{n}[(x_i-\theta)+a_1(x_{i-1}-\theta)+\ldots+a_p(x_{i-p}-\theta)]^2 =$$

$$= \sum_{i=p+1}^{n}[(x_i+a_1x_{i-1}+\ldots+a_px_{i-p})-\theta(1+a_1+\ldots+a_p)]^2 =$$

$$= \sum_{i=p+1}^{n}(x_i+a_1x_{i-1}+\ldots+a_px_{i-p})^2-2\theta(1+a_1+\ldots+a_p)\sum_{i=p+1}^{n}(x_i+a_1x_{i-1}+\ldots+a_px_{i-p})+$$

$$+\theta^2(n-p)(1+a_1+\ldots+a_p)^2 =$$

$$= -2\theta(1+a_1+\ldots+a_p)\sum_{i=p+1}^{n}(x_i+a_1x_{i-1}+\ldots+a_px_{i-p})+f_2(x_1,\ldots,x_n)+g_2(\theta).$$

(2.9) és (2.11) behelyettesítésével $x(1),\ldots,x(n)$ (2.7 alatti együttes sűrűségfüggvényét a következő alakba írhatjuk:

$$(2.12)\qquad p(x_1,\ldots,x_n;\theta) = H(x_1,\ldots,x_n)G(\theta)\exp\frac{\theta}{\sigma_\varepsilon^2}\left\{\alpha_1x_1+\ldots+\alpha_px_p+\right.$$

$$\left.+(1+a_1+\ldots+a_p)\sum_{i=p+1}^{n}(x_i+a_1x_{i-1}+\ldots+a_px_{i-p})\right\},$$

ahol a $H(x_1,\ldots,x_n)$ és $G(\theta)$ függvények konkrét alakukra nem lesz szükségünk.

Tekintsük az

(2.13)

$$l = \alpha_1 x(1) + \ldots + \alpha_p x(p) + (1 + a_1 + \ldots + a_p) \sum_{i=p+1}^{n} [x(i) + a_1 x(i-1) + \ldots + a_p x(i-p)] =$$

$$= \alpha_1 x(1) + \ldots + \alpha_p x(p) + (1 + a_1 + \ldots + a_p) \left[\sum_{j=1}^{p} (a_p + \ldots + a_{p-j+1}) x(j) + \right.$$

$$\left. + \sum_{j=p+1}^{n-p} (1 + a_1 + \ldots + a_p) x(j) + \sum_{j=n-p+1}^{n} (1 + a_1 + \ldots + a_{n-j}) x(j) \right]$$

lineáris statisztikát.

Látható, hogy

$$l = \lambda_1 x(1) + \ldots + \lambda_n x(n)$$

alakú, ahol

(2.14)
$$\lambda_j = \alpha_j + (1 + a_1 + \ldots + a_p)(a_{p-j+1} + \ldots + a_p), \quad \text{ha} \quad j = 1, \ldots, p,$$
$$\lambda_j = (1 + a_1 + \ldots + a_p)^2, \quad \text{ha} \quad p+1 \leqq j \leqq n-p,$$
$$\lambda_j = (1 + a_1 + \ldots + a_p)(1 + a_1 + \ldots + a_{n-j}), \quad \text{ha} \quad n-p+1 \leqq j \leqq n.$$

(2.12)-ből nyilvánvaló, hogy az l statisztika elégséges az $x(1), \ldots, x(n)$ együttes sűrűségfüggvényeinek összességére nézve. Mivel $x(1), \ldots, x(n)$ normális eloszlású valószínűségi változók, könnyű megmutatni, hogy az l lineáris statisztika egyben teljes is. Vizsgáljuk most az $\hat{l} = \sum_{j=1}^{n} c_j^0 x(j)$ statisztikát. Mivel $\hat{l}$ legjobb lineáris torzítatlan becslése θ-nak, és a vizsgálandó esetben létezik az l elégséges statisztika, akkor $\hat{l}$ szükségképpen csak l-től függ, más szóval $\hat{l}$ függvénye l-nek.

Valójában, ellenkező esetben $\hat{l}$ nem volna függvénye l-nek. Tekintsük az $L(l) = E(\hat{l}|l)$ függvényt. Minthogy $x(1), \ldots, x(n)$ *Gauss-eloszlásúak,* ennélfogva $\hat{l}$ és l szintén *Gauss-eloszlású,* így

$$L(l) = E(\hat{l}|l) = \alpha l + \beta$$

alakú, azaz $L(l)$ lineáris függvénye l-nek, ebből következik, hogy $E(\hat{l}|l)$ lineáris függvénye $x(1), \ldots, x(n)$-nek.

A *Blackwell—Kolmogorov—Rao-tétel* szerint $E(\hat{l}|l)$ torzítatlan becslése θ-nak (így $E(\hat{l}|l)$ lineáris torzítatlan becslése θ-nak) és

(2.15)
$$E_\theta(\hat{l} - \theta)^2 \geqq E_\theta[E(\hat{l}|l) - \theta]^2, \quad \text{minden } \theta \in \mathscr{R}^1\text{-re.}$$

A feltevés szerint $\hat{l}$ legjobb torzítatlan lineáris becslése θ-nak, ezért a (2.15) relációban szükségképpen teljesül az egyenlőség:

(2.16)
$$E_\theta(\hat{l} - \theta)^2 = E_\theta[E(\hat{l}|l) - \theta]^2 \quad \text{minden } \theta \in \mathscr{R}^1\text{-re.}$$

Ha visszaemlékszünk a *Blackwell—Kolmogorov—Rao-tétel* bizonyítására (lásd pl. [46]), (2.16) akkor és csak akkor teljesül, ha

$$\hat{l} = E(\hat{l}|l) = L(l), \quad \text{m. m. } P_\theta, \quad \theta \in \mathscr{R}^1,$$

ami ellentmond a feltevésünknek, hogy $\hat{l}$ nem függ l-től. Ezzel igazoltuk állításunkat.

Az l statisztika teljessége miat $\hat{l}$ θ-nak egyetlen olyan torzítatlan becslése, amely függvénye l-nek, így $\hat{l}$ legjobb torzítatlan becslés θ-ra.

A fenti eredményeket összefoglalva a következő állítást kapjuk:

2.1. Tétel. Ha $x(j)$ *Gauss-folyamat* és $x(j)=\xi(j)+\theta$ alakú, ahol $\xi(j)$ stacionárius folyamat és kielégíti a (2.1.3) sztochasztikus differenciaegyenletet, akkor az $\hat{l}=\sum_{j=1}^{n} c_j^0 x(j)$ $(n\geqq 2p+1)$ legjobb lineáris torzítatlan becslés megengedhető — sőt optimális — a θ torzítatlan becsléseinek osztályában.

Megjegyzés. Ha speciálisan, $\xi(t)$ a következő differenciaegyenletnek eleget tevő stacionárius elsőrendű autoregressziós Gauss-folyamat:

$$\xi(t) = a\xi(t-1)+\varepsilon(t),$$

ahol $\varepsilon(t)$ $(E\varepsilon(t)=0,\ E\varepsilon^2(t)=\sigma_\varepsilon^2)$ független azonos eloszlású *Gauss-sorozat,* akkor a fentiek alapján belátható, hogy az $x(1), \ldots, x(n)$ valószínűségi változók együttes sűrűségfüggvénye felírható a következő alakban:

$$p(x_1, \ldots, x_n; \theta) = H(x_1, \ldots, x_n)G(\theta)\exp\frac{\theta}{\sigma_\varepsilon^2}(\lambda_1 x_1+\ldots+\lambda_n x_n),$$

ahol

$$\lambda_1 = 1+a(1-a),$$
$$\lambda_j = (1-a)^2, \qquad 2\leqq j\leqq n-1$$
$$\lambda_n = 1-a,$$

vagy ami ezzel ekvivalens,

$$p(x_1, \ldots, x_n; \theta) = H(x_1, \ldots, x_n)G(\theta)\exp\frac{\theta}{c}(c_1^0 x_1+\ldots+c_n^0 x_n),$$

ahol $\hat{l}=c_1^0 x(1)+\ldots+c_n^0 x(n)$ legjobb lineáris torzítatlan becslése θ-nak, és c egy meghatározott állandó.

Legyen most $x(j)=\xi(j)+\theta$, $j=1, \ldots, n$ és $\xi(j)$ a következő p-edrendű autoregressziós folyamat:

$$\begin{aligned}
\xi(1) &= \varepsilon(1),\\
\xi(2)+a_1\xi(1) &= \varepsilon(2),\\
&\cdots\cdots\cdots\cdots\\
\xi(p)+a_1\xi(p-1)+\ldots+a_{p-1}\xi(1) &= \varepsilon(p),\\
\xi(k)+a_1\xi(k-1)+\ldots+a_p\xi(k-p) &= \varepsilon(k), \quad k=p+1, \ldots, n,
\end{aligned} \tag{2.17}$$

ahol $\varepsilon(k)$ $(E\varepsilon(k)=0,\ E\varepsilon^2(k)=\sigma_k^2,\ 0<\sigma_k^2<\infty)$ független *Gauss-sorozat.*

Tegyük fel, hogy $n\geqq 2p+1$. Könnyen belátható, hogy $\varepsilon(1), \ldots, \varepsilon(n)$ együttes sűrűségfüggvénye a következő:

$$p_{\xi(1),\ldots,\varepsilon(n)}(z_1, \ldots, z_n) = (2\Pi)^{-n/2}(\sigma_1 \ldots \sigma_n)^{-1}\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\frac{z_i^2}{\sigma_i^2}\right\}. \tag{2.18}$$

Felhasználva a (2.17) egyenletek által meghatározott leképezést, amelynek determinánsa 1, kapjuk $\xi(1), \ldots, \xi(n)$ együttes sűrűségfüggvényét a $(a_0=1)$

(2.19)

$$p_{\xi(1),\ldots,\xi(n)}(u_1, \ldots, u_n)=(2\Pi)^{-n/2}(\sigma_1 \ldots \sigma_n)^{-1}\exp\left\{-\right.$$

$$-\frac{1}{2}\left[\sum_{i=1}^{p}\frac{(a_0u_i+a_1u_{i-1}+\ldots+a_{i-1}u_1)^2}{\sigma_i^2}+\right.$$

$$\left.\left.+\sum_{i=p+1}^{n}\frac{(a_0u_i+a_1u_{i-1}+\ldots+a_pu_{i-p})^2}{\sigma_i^2}\right]\right\}$$

alakban.

Vegyük figyelembe, hogy

$$\sum_{i=1}^{p}\frac{(a_0u_i+a_1u_{i-1}+\ldots+a_{i-1}u_1)^2}{\sigma_i^2}=\frac{a_0^2}{\sigma_1^2}u_1^2+\frac{(a_0u_2+a_1u_1)^2}{\sigma_2^2}+\ldots+$$

$$+\frac{(a_0u_p+a_1u_{p-1}+\ldots+a_{p-1}u_1)^2}{\sigma_p^2}=$$

$$=u_1^2\left(\frac{a_0^2}{\sigma_1^2}+\frac{a_1^2}{\sigma_2^2}+\ldots+\frac{a_{p-1}^2}{\sigma_p^2}\right)+u_2^2\left(\frac{a_0^2}{\sigma_2^2}+\ldots+\frac{a_{p-2}^2}{\sigma_p^2}\right)+\ldots+u_p^2\frac{a_0^2}{\sigma_p^2}+$$

$$+2u_1u_2\left(\frac{a_0a_1}{\sigma_2^2}+\frac{a_1a_2}{\sigma_3^2}+\ldots+\frac{a_{p-2}a_{p-1}}{\sigma_p^2}\right)+\ldots+2u_1u_p\frac{a_0a_{p-1}}{\sigma_p^2}+$$

$$+\ldots+2u_{p-1}u_p\frac{a_0a_1}{\sigma_p^2}=\mathbf{U}_p^T\mathbf{D}_p\mathbf{U}_p,$$

ahol $\mathbf{U}_p=(u_1, \ldots, u_p)^T$ és $(a_0=1)$

(2.20)

$$\mathbf{D}_p=\begin{Bmatrix}\frac{a_0^2}{\sigma_1^2}+\frac{a_1^2}{\sigma_2^2}+\ldots+\frac{a_{p-1}^2}{\sigma_p^2} & \frac{a_0a_1}{\sigma_2^2}+\frac{a_1a_2}{\sigma_3^2}+\ldots+\frac{a_{p-2}a_{p-1}}{\sigma_p^2} & \ldots & \frac{a_0a_{p-1}}{\sigma_p^2}\\ \frac{a_0a_1}{\sigma_p^2}+\frac{a_1a_2}{\sigma_3^2}+\ldots+\frac{a_{p-2}a_{p-1}}{\sigma_p^2} & \frac{a_0^2}{\sigma_2^2}+\frac{a_1^2}{\sigma_3^2}+\ldots+\frac{a_{p-2}^2}{\sigma_p^2} & \ldots & \frac{a_0a_{p-2}}{\sigma_p^2}\\ \vdots & & \ddots & \\ \frac{a_0a_{p-1}}{\sigma_p^2} & & & \frac{a_0^2}{\sigma_p^2}\end{Bmatrix}=$$

$$=(d_{ij})\quad i=1, \ldots, p,\quad j=1, \ldots, p$$

(a $\mathbf{D}_p$ mátrix szimmetrikus a főátlóra nézve). Ily módon

(2.21)

$$p_{\xi(1),\ldots,\xi(n)}(u_1, \ldots, u_n)=$$

$$=2\Pi^{-n/2}(\sigma_1 \ldots \sigma_n)^{-1}\exp\left\{-\frac{1}{2}\left[\mathbf{U}_p^T\mathbf{D}_p\mathbf{U}_p+\sum_{i=p+1}^{n}\left(\frac{a_0u_i+a_1u_{i-1}+\ldots+a_pu_{i-p}}{\sigma_i}\right)^2\right]\right\}.$$

Ebből már adódik $x(1), \ldots, x(n)$ együttes sűrűségfüggvénye:

$$(2.22)\quad p(x_1, \ldots, x_n; \theta) = (2\Pi)^{-n/2}(\sigma_1 \ldots \sigma_n)^{-1} \exp\left[-\frac{1}{2}\left[(\mathbf{X}_p-\boldsymbol{\theta})^T \mathbf{D}_p(\mathbf{X}_p-\boldsymbol{\theta})+\right.\right.$$

$$\left.\left.+\sum_{i=p+1}^{n}\left(\frac{a_0(x_i-\theta)+a_1(x_{i-1}-\theta)+\ldots+a_p(x_{i-p}-\theta)}{\sigma_i}\right)^2\right]\right],$$

ahol

$$\mathbf{X}_p-\boldsymbol{\theta} = (x_1-\theta, \ldots, x_n-\theta)^T.$$

Egyrészt,

$$(2.23)\quad (\mathbf{X}_p-\boldsymbol{\theta})^T \mathbf{D}_p(\mathbf{X}_p-\boldsymbol{\theta}) = -2\theta\{d_{11}x_1+d_{12}(x_1+x_2)+\ldots+d_{1p}(x_1+x_p)+$$

$$+d_{22}x_2+\ldots+d_{2p}(x_2+x_p)+\ldots+d_{pp}x_p\}+f_1^*(x_1, \ldots, x_n)+g_1^*(\theta) =$$

$$= -2\theta\{\alpha_1^* x_1+\ldots+\alpha_p^* x_p\}+f_1^*(x_1, \ldots, x_n)+g_1^*(\theta),$$

ahol

$$(2.24)\quad \alpha_1^* = d_{11}+d_{12}+\ldots+d_{1p}$$

$$\alpha_2^* = d_{21}+d_{22}+\ldots+d_{2p}$$

$$\alpha_p^* = d_{p1}+d_{p2}+\ldots+d_{pp}$$

(az $f_1^*(x_1, \ldots, x_n)$, $g_1^*(\theta)$ függvények konkrét alakjára nem lesz szükségünk). Másrészt,

$$(2.25)\quad \sum_{i=p+1}^{n}\left[\frac{(x_i-\theta)+a_1(x_{i-1}-\theta)+\ldots+a_p(x_{i-p}-\theta)}{\sigma_i}\right]^2 =$$

$$= \sum_{i=p+1}^{n}\left[\frac{(x_i+a_1x_{i-1}+\ldots+a_px_{i-p})}{\sigma_i} - \frac{\theta(1+a_1+\ldots+a_p)}{\sigma_i}\right]^2 =$$

$$= \sum_{i=p+1}^{n}\frac{(x_i+a_1x_{i-1}+\ldots+a_px_{i-p})^2}{\sigma_i^2} - 2\theta(1+a_1+\ldots+a_p)\times$$

$$\times\sum_{i=p+1}^{n}\frac{(x_i+a_1x_{i-1}+\ldots+a_px_{i-p})}{\sigma_i^2}+\theta^2(1+a_1+\ldots+a_p)^2\sum_{i=p+1}^{n}\frac{1}{\sigma_i^2} =$$

$$= -2\theta(1+a_1+\ldots+a_p)\sum_{i=p+1}^{n}\frac{x_i+a_1x_{i-1}+\ldots+a_px_{i-p}}{\sigma_i^2}+f_2^*(x_1, \ldots, x_n)+g_2^*(\theta).$$

(2.23) és (2.25) behelyettesítésével kapjuk $x(1), \ldots, x(n)$ együttes sűrűségfüggvényét a következő alakban:

$$(2.26)\quad p(x_1, \ldots, x_n; \theta) = H^*(x_1, \ldots, x_n)G^*(\theta)\exp\theta\times$$

$$\times\left\{\alpha_1^* x_1+\ldots+\alpha_p^* x_p+(1+a_1+\ldots+a_p)\sum_{i=p+1}^{n}\frac{x_i+a_1x_{i-1}+\ldots+a_px_{i-p}}{\sigma_i^2}\right\}.$$

Legyen

(2.27)

$$l^* = \alpha_1^* x(1)+\ldots+\alpha_p^* x(p)+(1+a_1+\ldots+a_p)\sum_{i=p+1}^{n}\frac{x(i)+a_1x(i-1)+\ldots+a_px(i-p)}{\sigma_i^2} =$$

$$= \alpha_1^* x(1)+\ldots+\alpha_p^* x(p)+(1+a_1+\ldots+a_p)\left[\sum_{j=1}^{p}\left(\frac{a_{p-j+1}}{\sigma_{p+1}^2}+\ldots+\frac{a_p}{\sigma_{p+j}^2}\right)x(j)+\right.$$

$$\left.+\sum_{j=p+1}^{n-p}\left(\frac{1}{\sigma_j^2}+\frac{a_1}{\sigma_{j+1}^2}+\ldots+\frac{a_p}{\sigma_{j+p}^2}\right)x(j)+\sum_{j=n-p+1}^{n}\left(\frac{1}{\sigma_j^2}+\frac{a_1}{\sigma_{j+1}^2}+\ldots+\frac{a_{n-j}}{\sigma_n^2}\right)x(j)\right] =$$

$$= \lambda_1^* x(1)+\ldots+\lambda_n^* x(n),$$

ahol

(2.28)
$$\lambda_j^* = \alpha_j^*+(1+a_1+\ldots+a_p)\left(\frac{a_{p-j+1}}{\sigma_{p+1}^2}+\ldots+\frac{a_p}{\sigma_{p+j}^2}\right), \quad 1 \leqq j \leqq p,$$

$$\lambda_j = (1+a_1+\ldots+a_p)\left(\frac{1}{\sigma_j^2}+\frac{a_1}{\sigma_{j+1}^2}+\ldots+\frac{a_p}{\sigma_{j+p}^2}\right), \quad p+1 \leqq j \leqq n-p,$$

$$\lambda_j^* = (1+a_1+\ldots+a_p)\left(\frac{1}{\sigma_j^2}+\frac{a_1}{\sigma_{j+1}^2}+\ldots+\frac{a_{n-j}}{\sigma_n^2}\right), \quad n-p+1 \leqq j \leqq n.$$

(2.26)-ból belátható, hogy az l^* statisztika elégséges (és egyben teljes is) a $p(x_1, \ldots, x_n; \theta)$ sűrűségfüggvények összességére nézve.

A 2.1. tétel bizonyításához hasonlóan megmutatható, hogy az $\hat{l}^* = \sum_{j=1}^{n} c_j^{0*} x(j)$ legjobb lineáris torzítatlan becslés szükségképpen függvénye l^*-nak. Az l^* teljessége miatt $\hat{l}^*$ egyetlen olyan torzítatlan becslése θ-nak, amely függvénye l^*-nak. A *Blackwell—Kolmogorov—Rao-tétel* szerint $\hat{l}^*$ a legjobb torzítatlan becslés θ-ra.

Tehát igazoltuk a következő állítást:

2.2. Tétel. Legyen $x(j)$ *Gauss-folyamat* és $x(j)=\xi(j)+\theta$, ahol $\xi(j)$ a (2.17) alakú p-edrendű autoregressziós folyamat. Legyen továbbá $n \geqq 2p+1$. Az $\hat{l}^* = \sum_{j=1}^{n} c_j^{0*} x(j)$ legjobb lineáris torzítatlan becslés ekkor megengedhető — sőt optimális — a θ torzítatlan becsléseinek osztályában.

1. Megjegyzés. A 2.2. tétel bizonyításából belátható, hogy abban az esetben, amikor a (2.17) egyenletben $\varepsilon(k)$ $(E\varepsilon(k)=0, E\varepsilon^2(k)=\sigma_\varepsilon^2)$ független azonos eloszlású *Gauss-sorozat,* akkor az

(2.29)
$$\bar{l}^* = \bar{\lambda}_1^* x(1)+\ldots+\bar{\lambda}_n^* x(n)$$

lineáris statisztika, ahol

(2.30)
$$\bar{\lambda}_j^* = \alpha_{p-j+1}+(1+a_1+\ldots+a_p)(a_{p-j+1}+\ldots+a_p), \quad \text{ha} \quad 1 \leqq j \leqq p,$$

$$\bar{\lambda}_j^* = (1+a_1+\ldots+a_p)^2, \quad \text{ha} \quad p+1 \leqq j \leqq n-p,$$

$$\bar{\lambda}_j^* = (1+a_1+\ldots+a_p)(1+a_1+\ldots+a_{n-j}), \quad \text{ha} \quad n-p+1 \leqq j \leqq n$$

elégséges, egyben teljes is az $x(1), \ldots, x(n)$ együttes sűrűségfüggvényeinek összességére nézve.

(A (2.30)-ban szereplő α_j, $j=1, \ldots, p$ állandók (2.10)-ből meghatározhatók.)

2. Megjegyzés. A

$$\xi(1) = \xi(1)$$

$$\xi(k) = a\xi(k-1)+\varepsilon(k), \quad k = 2, \ldots, n, \tag{2.31}$$

elsőrendű autoregressziós folyamat esetén, ahol $\varepsilon(k)$, $k=1, \ldots, n$, $(E\varepsilon(k)=0$, $E\varepsilon^2(k)=\sigma_2^k$, $0<\sigma_k^2<\infty)$ független *Gauss-sorozat*, könnyű belátni, hogy az l^* lineáris statisztika a következő alakú:

$$l^* = \left[\frac{1}{\sigma_1^2} - \frac{a(1-a)}{\sigma_2^2}\right] x(1)+1(-a)\sum_{j=2}^{n-1}\left(\frac{1}{\sigma_j^2}-\frac{a}{\sigma_{j+1}^2}\right)x(j)+\frac{1-a}{\sigma_n^2}x(n). \tag{2.32}$$

Vizsgáljuk az

$$x(j) = \xi(j)+\theta, \quad j = 1, \ldots, n$$

sztochasztikus folyamatot. Legyen $\xi(j)$ (2.17) alakú p-edrendű autoregressziós folyamat azzal a feltevéssel, hogy $\varepsilon(k)$ $(E\varepsilon(k)=0$, $E\varepsilon^2(k)=\sigma_\varepsilon^2)$ független, azonos eloszlású sorozat.

Tegyük fel, hogy az $\hat{l}^0 = \sum_{j=1}^{n} \bar{c}_j^0 x(j)\left(\sum_{j=1}^{n}\bar{c}_j^0 = 1\right)$ legjobb lineáris torzítatlan becslés megengedhető θ torzítatlan becsléseinek osztályában.

Megmutatjuk, hogy az $x(j)$ folyamat *Gauss-folyamat*. Legyen

$$\mathbf{y} = (x(2)-x(1), \ldots, x(n)-x(1))$$

és vezessük be a

$$\hat{\theta}_0 = \hat{l}^0 - E_0(\hat{l}^0|\mathbf{y}) \tag{2.33}$$

Pitman-féle becslést. Belátható, hogy

$$E_\theta(\hat{\theta}_0) = E_\theta(\hat{l}^0) - E_\theta(E_0(\hat{l}^0|\mathbf{y})) = \theta - E_0(E_0(\hat{l}^0|\mathbf{y})) = \theta,$$

azaz $\hat{\theta}_0$ torzítatlan becslése θ-nak. Mivel $\hat{l}^0$ korrekt becslés, az 1.1. tétel első állítása szerint

$$E_\theta(\hat{\theta}_0-\theta)^2 \leqq E_\theta(\hat{l}^0-\theta)^2,$$

az egyenlőség minden $\theta\in\mathscr{R}^1$-ra akkor és csak akkor áll fenn, ha $\hat{\theta}_0=\hat{l}^0$ (m. m. P_θ, $\theta\in\mathscr{R}^1$) vagy, ami ugyanaz,

$$E_0(\hat{l}^0|\mathbf{y}) = 0.$$

Ebből már következik, hogy az $\hat{l}^0$ becslés akkor és csak akkor megengedhető θ torzítatlan becsléseinek osztályában, ha

$$E_0(\hat{l}^0|\mathbf{y}) = 0. \tag{2.34}$$

Legyenek

$$f(t) = E \exp it\varepsilon_j, \quad g(t) = \frac{f'(t)}{f(t)},$$

$$\Phi(t_1, \ldots, t_n) = E \exp i \sum_{j=1}^{n} t_j \xi_j = E_0 \exp i \sum_{j=1}^{n} t_j x(j),$$

$$\varphi(t_1, \ldots, t_n) = \log \Phi(t_1, \ldots, t_n),$$

ahol $t_1, \ldots, t_n$ valós számok, a $g(t)$ függvény a $t=0$ pont valamilyen környezetében, $\varphi(t_1, \ldots, t_n)$ pedig a $(0, \ldots, 0)$ pont egy környezetében van értelmezve.

A $\xi(t)$ elsőrendű autoregressziós folyamat esetén teljesül az alábbi lemma (lásd [21], 7.8.2. lemma), amely érvényben marad a $\xi(t)$ p-edrendű autoregressziós folyamatra is.

2.1. Lemma. Tegyük fel, hogy $E_0(\hat{l}^0|\mathbf{y})=0$, akkor a $(0, \ldots, 0)$ pont valamilyen környezetében

$$\sum_{j=1}^{n} \bar{c}_j^0 \frac{\partial \varphi(\tau_1, \ldots, \tau_n)}{\partial \tau_j} = 0, \quad \text{ha} \quad \sum_{j=1}^{n} \tau_j = 0. \tag{2.35}$$

Bizonyítás. Mivel feltevésünk szerint $E_0(\hat{l}^0|\mathbf{y})=0$, ezért

$$E_0\left\{\hat{l}^0 \exp i \sum_{j=2}^{n} t_j(x_j - x_1)\right\} = E_0\left\{E_0\left(\hat{l}^0 \exp i \sum_{j=2}^{n} t_j(x_j - x_1)|\mathbf{y}\right)\right\} = \tag{2.36}$$

$$= E_0\left\{\exp i \sum_{j=2}^{n} t_j(x_j - x_1) E_0(\hat{l}^0|\mathbf{y})\right\} = 0.$$

Figyelembe véve, hogy

$$\frac{\partial \Phi\left(-\sum_{j=2}^{n} t_j, t_2, \ldots, t_n\right)}{\partial t_1} = E_0\left\{i x_1 \exp i \sum_{j=2}^{n} t_j(x_j - x_1)\right\},$$

$$\frac{\partial \Phi\left(-\sum_{j=1}^{n} t_j, t_2, \ldots, t_n\right)}{\partial t_j} = E_0\left\{i x_j \exp i \sum_{j=1}^{n} t_j(x_j - x_1)\right\},$$

(2.36)-ból kapjuk a

$$0 = E_0\left\{\hat{l}^0 \exp i \sum_{j=2}^{n} t_j(x_j - x_1)\right\} = E_0\left\{\sum_{j=1}^{n} \bar{c}_j^0 x_j \exp i \sum_{j=2}^{n} t_j(x_j - x_1)\right\} = \tag{2.37}$$

$$= \bar{c}_1^0 \frac{\partial \Phi\left\{-\sum_{j=2}^{n} t_j, t_2, \ldots, t_n\right\}}{\partial t_1} + \sum_{j=2}^{n} \bar{c}_j^0 \frac{\partial \Phi\left(-\sum_{j=2}^{n} t_j, t_2, \ldots, t_n\right)}{\partial t_j}$$

összefüggést.

Minthogy $\Phi(t_1, ..., t_n)$ karakterisztikus függvénye $(\xi_1, ..., \xi_n)$-nek, ennélfogva $\Phi(0, ..., 0)=1$, $\Phi(t_1, ..., t_n)$ t_j-ben egyenletesen folytonos, amiből következik, hogy a $(0, ..., 0)$ pont egy környezetében $\Phi\left(-\sum_{j=2}^{n} t_j, t_2, ..., t_n\right)\neq 0$. (2.37) mindkét oldalát $\Phi\left(-\sum_{j=2}^{n} t_j, t_2, ..., t_n\right)$-nel osztva kapjuk, hogy

$$\sum_{j=1}^{n} \bar{c}_j^0 \frac{\partial \varphi\left(-\sum_{j=2}^{n} t_j, t_2, ..., t_n\right)}{\partial t_j} = 0,$$

amivel ekvivalens

$$\sum_{j=1}^{n} \bar{c}_j^0 \frac{\partial \varphi(\tau_1, ..., \tau_n)}{\partial \tau_j} = 0, \quad \text{ha} \quad \sum_{j=1}^{n} \tau_j = 0.$$

Ezzel kész a 2.1. lemma bizonyítása.

A (2.17) alakú p-edrendű autoregressziós folyamat esetén $\xi(n)$ előállítható az $\varepsilon(k)$ sorozat segítségével.

Valóban, írjuk fel a (2.17) egyenleteket $k=n, n-1, ..., 2, 1$-re

$$\begin{aligned}
\xi(n)+a_1\xi(n-1)+...+a_p\xi(n-p) &= \varepsilon(n)\\
\xi(n-1)+a_1\xi(n-2)+...+a_p\xi(n-p-1) &= \varepsilon(n-1)\\
&\cdots\cdots\cdots\\
\xi(2)+a_1\xi(1) &= \varepsilon(2)\\
\xi(1) &= \varepsilon(1),
\end{aligned}$$

azután szorozzuk meg ezeket rendre a $b_0(b_0=1), b_1, ..., b_{n-2}, b_{n-1}$ számokkal majd így adjuk össze. A kapott összefüggésben $\xi(n-1), ..., \xi(2), \xi(1)$ akkor és csak akkor nem szerepel, ha

$$\begin{aligned}
a_0b_1+a_1b_0 &= 0\\
a_0b_2+a_1b_1+a_2b_0 &= 0\\
&\cdots\cdots\cdots\\
a_0b_p+a_1b_{p-1}+...+a_pb_0 &= 0\\
&\cdots\cdots\cdots\\
a_0b_{n-1}+a_1b_{n-2}+...+a_pb_{n-(p+1)}+a_{p+1}b_{n-(p+2)}+...+a_{n-1}b_0 &= 0,
\end{aligned} \tag{2.38}$$

ahol $a_0=b_0=1$ és $a_k=0$, ha $k\geqq p+1$. (2.38) teljesülése esetén kapjuk, hogy

$$\xi(n) = \varepsilon(n)+b_1\varepsilon(n-1)+...+b_{n-1}\varepsilon(1) = \sum_{k=0}^{n-1} b_k\varepsilon(n-k). \tag{2.39}$$

Ily módon a $\xi(n)$ sorozat előállítható a (2.39) alakban, ahol a b_k együtthatók

(2.38)-ból határozhatók meg. (Könnyen belátható, hogy pl.

$$b_1 = -a_1,$$
$$b_2 = -a_2 + a_1^2,$$
$$b_3 = -a_3 + 2a_1 a_2 - a_1^3$$
$$b_4 = -a_4 + 2a_1 a_3 + a_2^2 - 3a_1^2 a_2 + a_1^4, \text{ stb.})$$

2.2. Lemma. A (2.17) $\xi(t)$ p-edrendű autoregressziós folyamat esetén

$$\Phi(t_1, \ldots, t_n) = f(t_1 + b_1 t_2 + \ldots + b_{n-1} t_n) \ldots f(t_{n-1} + b_1 t_n) \cdot f(t_n). \tag{2.40}$$

Bizonyítás. Valóban (2.39) szerint $\xi(k) = b_{k-1}\varepsilon(1) + b_{k-2}\varepsilon(2) + \ldots + \varepsilon(k)$, $k = 1, 2, \ldots, n$, ekkor

$$\begin{aligned}
\Phi(t_1, \ldots, t_n) &= E \exp i \sum_{j=1}^{n} t_j \xi(j) = \\
&= E \exp i\{t_1\varepsilon_1 + t_2(b_1\varepsilon_1 + \varepsilon_2) + \ldots + t_n(b_{n-1}\varepsilon_1 + \ldots + b_1\varepsilon_{n-1} + \varepsilon_n)\} = \\
&= E \exp i\{(t_1 + b_1 t_2 + \ldots + b_{n-1} t_n)\varepsilon_1 + \ldots + (t_{n-1} + b_1 t_n)\varepsilon_{n-1} + t_n\varepsilon_n\} = \\
&= E e^{i(t_1 + b_1 t_2 + \ldots + b_{n-1} t_n)\varepsilon_1} \ldots E e^{i(t_{n-1} + b_1 t_n)\varepsilon_{n-1}} E e^{i t_n \varepsilon_n} = \\
&= f(t_1 + b_1 t_2 + \ldots + b_{n-1} t_n) \ldots f(t_{n-1} + b_1 t_n) f(t_n).
\end{aligned}$$

A 2.2. lemma alapján

$$\varphi(t_1, \ldots, t_n) = \log \Phi(t_1, \ldots, t_n) =$$
$$= \log f(t_1 + b_1 t_2 + \ldots + b_{n-1} t_n) + \ldots + \log f(t_{n-1} + b_1 t_n) + \log f(t_n).$$

Innen könnyen belátható, hogy a (2.17) p-edrendű autoregressziós folyamat esetén a (2.35) összefüggés felírható a következő alakban:

$$\bar{c}_n^0 \left\{ \frac{f'(t_n)}{f(t_n)} + b_1 \frac{f'(t_{n-1} + b_1 t_n)}{f(t_{n-1} + b_1 t_n)} + \ldots + b_{n-1} \frac{f'(t_1 + b_1 t_2 + \ldots + b_{n-1} t_n)}{f(t_1 + b_1 t_2 + \ldots + b_{n-1} t_n)} \right\} +$$

$$+ \bar{c}_{n-1}^0 \left\{ \frac{f'(t_{n-1} + b_1 t_n)}{f(t_{n-1} + b_1 t_n)} + b_1 \frac{f'(t_{n-2} + b_1 t_{n-1} + b_2 t_n)}{f(t_{n-2} + b_1 t_{n-1} + b_2 t_n)} + \ldots + b_{n-2} \frac{f'(t_1 + b_1 t_2 + \ldots + b_{n-1} t_n)}{f(t_1 + b_1 t_2 + \ldots + b_{n-1} t_n)} \right\} +$$

$$+ \ldots + \bar{c}_1^0 \frac{f'(t_1 + b_1 t_2 + \ldots + b_{n-1} t_n)}{f(t_1 + b_1 t_2 + \ldots + b_{n-1} t_n)} = 0,$$

vagy

$$\bar{c}_n^0 g(t_n) + (b_1 \bar{c}_n^0 + \bar{c}_{n-1}^0) g(t_{n-1} + b_1 t_n) + \ldots + \tag{2.41}$$
$$+ (b_{n-1}\bar{c}_n^0 + b_{n-2}\bar{c}_{n-1}^0 + \ldots + \bar{c}_1^0) g(t_1 + b_1 t_2 + \ldots + b_{n-1} t_n) = 0,$$

ha

$$\sum_{j=1}^{n} t_j = 0.$$

Legyenek

(2.42)
$$\begin{aligned}\gamma_n &= \bar{c}_n^0\\ \gamma_{n-1} &= b_1\bar{c}_n^0+\bar{c}_{n-1}^0\\ \gamma_{n-2} &= b_2\bar{c}_n^0+b_1\bar{c}_{n-1}^0+\bar{c}_{n-2}^0\\ &\cdots\cdots\cdots\cdots\cdots\\ \gamma_1 &= b_{n-1}\bar{c}_n^0+b_{n-2}\bar{c}_{n-1}^0+\ldots+\bar{c}_1^0.\end{aligned}$$

Vezessük be a

(2.43)
$$\begin{aligned}s_n &= t_n\\ s_{n-1} &= b_1t_n+t_{n-1}\\ s_{n-2} &= b_2t_n+b_1t_{n-1}+t_{n-2}\\ &\cdots\cdots\cdots\cdots\cdots\\ s_1 &= b_{n-1}t_n+b_{n-2}t_{n-1}+\ldots+t_1\end{aligned}$$

új változókat, akkor (2.41) a

(2.44)
$$\sum_{j=1}^{n}\gamma_j g(s_j)=0$$

alakra hozható. Legyen $b_i^*=-b_i$, $i=1, 2, \ldots, n-1$, (2.43)-ból belátható, hogy

(2.45)
$$\begin{aligned}t_n &= v_0s_n\\ t_{n-1} &= v_1s_n+v_0s_{n-1}\\ &\cdots\cdots\cdots\cdots\\ t_{n-k} &= v_ks_n+v_{k-1}s_{n-1}+\ldots+v_0s_{n-k}\\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\ t_1 &= v_{n-1}s_n+v_{n-2}s_{n-1}+\ldots+v_0s_1\end{aligned}$$

ahol

$$\begin{aligned}v_0 &= a_0=1\\ v_1 &= b_1^*=a_1\\ v_2 &= b_1^{*2}+b_2^*=a_2\\ v_3 &= b_1^{*3}+2b_1^*b_2^*+b_3^*=a_3\\ v_4 &= b_1^{*4}+3b_1^{*2}b_2^*+2b_1^*b_3^*+b_2^{*2}+b_4^*=a_4\end{aligned}$$

és általában

(2.46)
$$v_{k+1}=b_1^*v_k+b_2^*v_{k-1}+\ldots+b_{k+1}^*v_0.$$

Ily módon a $\sum_{j=1}^{n}t_j=0$ feltételt átírhatjuk s_i segítségével az

$$(1+v_1+\ldots+v_{n-1})s_n+(1+v_1+\ldots+v_{n-2})s_{n-1}+\ldots+(1+v_1)s_2+s_1=0$$

alakba.

A fentieket a következőképpen foglalhatjuk össze: Ha az $\hat{l}^0 = \sum_{j=1}^{n} \bar{c}_j^0 x(j)$ legjobb lineáris torzítatlan becslés megengedhető a θ paraméter torzítatlan becslései osztályában, akkor

$$\sum_{j=1}^{n} \gamma_j g(s_j) = 0, \tag{2.47}$$

ahol $(1+v_1+\ldots+v_{n-1})s_n+\ldots+(1+v_1)s_2+s_1=0$.

Tegyük fel, hogy a $j=2, \ldots, n-1$ indexek közül létezik legalább két index, mondjuk i és k, hogy

$$(1+v_1+\ldots+v_{i-1})(1+v_i+\ldots+v_{k-1}) \neq 0.$$

(Egyszerűség kedvéért tegyük fel, hogy létezik pontosan két index, i és k, hogy

$$(1+v_1+\ldots+v_{i-1})(1+v_1+\ldots+v_{k-1}) \neq 0).$$

E feltételek teljesülése esetén a (2.47) egyenlet a következő egyenletre vezethető vissza:

$$g[(1+v_1+\ldots+v_{i-1})s_i+(1+v_1+\ldots+v_{k-1})s_k] = -\frac{\gamma_i}{\gamma_1} g(s_i) - \frac{\gamma_k}{\gamma_i} g(s_k). \tag{2.48}$$

Ismerjük, hogy a (2.48) egyenlet a *Cauchy-féle egyenlet* általánosítása (ACZÉL [0], és annak megoldása lineáris függvény. Ily módon $g(t)$ lineáris függvénye t-nek, amiből következik, hogy

$$f(t) = \exp P(t),$$

ahol $P(t)$ másodfokú polinom. *Marcinkiewicz tétele* szerint $f(t)$ normális eloszlás karakterisztikus függvénye.

Ily módon igazoltuk a következő állítást:

2.3. Tétel. Legyen $x(j)=\xi(j)+\theta$, ahol $\xi(j)$ (2.1.17) alakú p-edrendű autoregressziós folyamat ($\varepsilon(j)$ független azonos eloszlású sorozat).

Tegyük fel, hogy $f''(t)$ folytonos a $t=0$ pont környezetében. Akkor

$$\frac{\gamma_1}{\gamma_2}(1+a_1) = 1,$$

és ha az $\hat{l}^0 = \sum_{j=1}^{n} \bar{c}_j^0 x(j)$ legjobb lineáris becslés megengedhető a θ torzítatlan becsléseinek osztályában, akkor az $x(j)$ folyamat *Gauss-folyamat.*

Tekintsük speciálisan az

$$x(j) = \xi(j)+\theta, \quad j = 1, \ldots, n \tag{2.49}$$

sztochasztikus folyamatot (θ valós paraméter), ahol $\xi(j)$ a következő elsőrendű autoregressziós stacionárius folyamat:

$$\xi(1) = \frac{1}{\sqrt{1-a^2}} \varepsilon(1) \tag{2.50}$$

$$\xi(j) = a\xi(j-1)+\varepsilon(j), \quad j = 2, \ldots, n,$$

ahol $|a|<1$ és $\varepsilon(j)$ egy független, azonos eloszlású sorozat

$$(E\varepsilon(j)=0,\ E\varepsilon(j)^2=\sigma_\varepsilon^2,\ 0<\sigma_\varepsilon^2<\infty).$$

Tegyük fel, hogy $x(j)$ *Gauss-folyamat*. A (2.49), (2.50) séma esetén látható, hogy az $x(1), \ldots, x(n)$ változók együttes sűrűségfüggvénye a következő:

$$(2.51)\qquad p(x_1-\theta, \ldots, x_n-\theta) = (2\Pi\sigma_\varepsilon^2)^{-n/2}\sqrt{1-a^2}\times$$

$$\times\exp -\frac{1}{2\sigma_\varepsilon^2}\left\{(1-a^2)(x_1+\theta)^2+\sum_{j=2}^{n}[(x_j-\theta)-a(x_{j-1}-\theta)]^2\right\} =$$

$$= (2\Pi\sigma_\varepsilon^2)^{-n/2}\sqrt{1-a^2}\exp -\frac{1}{2\sigma_\varepsilon^2}\left\{(1-a^2)(x_1-\theta)^2+\sum_{j=2}^{n}[(x_j-ax_{j-1})-\theta(1-a)]^2\right\} =$$

$$= G(\theta; a)H(x_1, \ldots, x_n)\exp\frac{\theta}{\sigma_\varepsilon^2}\left\{(1-a^2)x_1+(1-a)\sum_{j=2}^{n}(x_j-ax_{j-1})\right\},$$

ahol $G(\theta; a)$ és $H(x_1, \ldots, x_n)$ függvények konkrét alakjára nem lesz szükségünk.

Tekintsük az

$$(2.52)\quad l = (1-a^2)x_1+(1-a)\sum_{j=2}^{n}(x_j-ax_{j-1}) = (1-a)x_1+(1-a)^2\sum_{j=2}^{n-1}x_j+(1-a)x_n$$

lineáris statisztikát. Nyilvánvaló, hogy az l lineáris statisztika elégséges (egyben teljes is) az $x_1, \ldots, x_n$ együttes sűrűségfüggvényeinek összességére nézve.

Ily módon az $\hat{l}^0=\sum_{j=1}^{n}c_j^0x(j)$ $\left(\sum_{j=1}^{n}c_j^0=1\right)$ legjobb lineáris becslés egyetlen olyan torzítatlan becslése θ-nak, amely függvénye l-nek.

Tegyük most fel, hogy a (2.49), (2.50) séma esetén az $\hat{l}^0=\sum_{j=1}^{n}c_j^0x(j)$ $\left(\sum_{j=1}^{n}c_j^0=1\right)$ legjobb lineáris torzítatlan becslés megengedhető θ torzítatlan becsléseinek osztályában. Megmutatjuk, hogy az $x(j)$ folyamat *Gauss-folyamat*. Ismerjük, hogy az $\hat{l}^0$ becslés akkor és csak akkor megengedhető θ torzítatlan becsléseinek osztályában, ha

$$(2.53)\qquad E_0(\hat{l}^0|x_2-x_1, \ldots, x_n-x_1) = 0,$$

Legyenek

$$f(t) = E\exp it\varepsilon_j,\quad g(t) = \frac{f'(t)}{f(t)},$$

$$\Phi(t_1, \ldots, t_n) = E\exp i\sum_{j=1}^{n}t_j\xi_j$$

$$\varphi(t_1, \ldots, t_n) = \log\Phi(t_1, \ldots, t_n).$$

Láthatjuk, hogy a (2.50) elsőrendű autoregressziós stacionárius folyamat esetén a 2.1. lemma érvényben marad. Ily módon (2.53)-ból következik, hogy a $(0, \ldots, 0)$

pont valamilyen környezetében

$$\sum_{j=1}^{n} c_j^0 \frac{\partial \varphi(\tau_1, \ldots, \tau_n)}{\partial \tau_j} = 0, \quad \text{ha} \quad \sum_{j=1}^{n} \tau = 0. \tag{2.54}$$

(2.50)-ből nyilvánvaló, hogy

$$\xi(1) = \frac{1}{\sqrt{1-a^2}} \varepsilon(1) \tag{2.55}$$

$$\xi(2) = \frac{a}{\sqrt{1-a^2}} \varepsilon(1) + \varepsilon(2)$$

$$\ldots\ldots\ldots\ldots\ldots\ldots$$

$$\xi(n) = \frac{1}{\sqrt{1-a^2}} a^{n-1}\varepsilon(1) + a^{n-2}\varepsilon(2) + \ldots + a\varepsilon(n-1) + \varepsilon(n).$$

Így, a (2.50) folyamat esetén

$$\Phi(t_1, \ldots, t_n) = E \exp i \sum_{j=1}^{n} t_j \xi_j = f\left[\frac{1}{\sqrt{1-a^2}} (t_1 + a_1 t_2 + \ldots + a^{n-1} t_n)\right] \cdot$$

$$\cdot f(t_2 + at_3 + \ldots + a^{n-2} t_n) \ldots f(t_n),$$

és a (2.54) egyenlet a következő alakú:

$$c_n^0 g(t_n) + (ac_n^0 + c_{n-1}^0) g(at_n + t_{n-1}) + \ldots + \tag{2.56}$$

$$+ \frac{1}{\sqrt{1-a^2}} (a^{n-1} c_n^0 + \ldots + ac_2^0 + c_1^0) g\left[\frac{1}{\sqrt{1-a^2}} (a^{n-1} t_n + \ldots + t_1)\right] = 0, \quad \text{ha} \quad \sum_{j=1}^{n} t_j = 0,$$

ahol

$$c_1^0 = 1 - a,$$

$$c_j^0 = (1-a)^2, \quad j = 2, \ldots, n-1,$$

$$c_n^0 = 1 - a.$$

Legyenek

$$\gamma_n = c_n^0$$

$$\gamma_{n-1} = ac_n^0 + c_{n-1}^0$$

$$\ldots\ldots\ldots\ldots$$

$$\gamma_1 = u^{n-1} c_n^0 + \ldots + ac_2^0 + c_1^0.$$

Belátható, hogy

$$\gamma_1 = 1 - a^2$$

$$\gamma_j = 1 - a, \quad j = 2, \ldots, n.$$

Vezessük be az

$$s_n = t_n$$
$$s_{n-1} = at_n + t_{n-1}$$
$$\cdots\cdots\cdots\cdots$$
$$s_1 = a^{n-1}t_n + \ldots + t_1.$$

új változókat.

Akkor

$$t_n = s_n$$
$$t_{n-1} = s_{n-1} - as_n$$
$$\cdots\cdots\cdots\cdots$$
$$t_1 = s_1 - as_2$$

és a (2.56) egyenlet a következő egyenletre vezethető vissza:

$$(2.57)\quad g\left(-\sqrt{\frac{1-a}{1+a}}\,s_2 - \ldots - \sqrt{\frac{1-a}{1+a}}\,s_n\right) = -\sqrt{\frac{1-a}{1+a}}\,g(s_2) - \ldots - \sqrt{\frac{1-a}{1+a}}\,g(s_n).$$

Ismerjük, hogy a (2.57) egyenlet nem más, mint lineáris függvényegyenlet — a *Cauchy-féle egyenlet* általánosítása (lásd pl. J. ACZÉL [0]), melynek megoldása lineáris függvény. Ily módon $g(t)$ lineáris függvénye t-nek, amiből következik, hogy $f(t)$ normális eloszlás karakterisztikus függvénye — azaz $\varepsilon(j)$ *Gauss-sorozat.*

A fenti eredményeket összefoglalva a következő állítást kapjuk:

2.4. Tétel. A (2.49)—(2.50) séma esetén az $\hat{l}^0 = \sum_{j=1}^{n} c_j^0 x(j)$ legjobb lineáris becslés megengedhetősége a θ torzítatlan becsléseinek osztályában jellemző tulajdonsága az $x(j)$ *Gauss-folyamatnak.*

2. Gauss-folyamatok jellemzése $\hat{l}^* = \sum_{j=1}^{n} c_j^{0*} x_j$ polinomjának optimalitásával

Legyen L_θ^2 azoknak az $(\mathcal{X}, \mathcal{A})$ téren mérhető függvényeknek *Hilbert-tere,* amelyek a P-mérték szerint négyzetes középben integrálhatók, azaz ha $h(x) \in L_\theta^2$, akkor $\int_{\mathcal{X}} h^2(x)\,dP_\theta(x) < \infty$. A skalárszorzat a szokásos módon definiálható. Legyen továbbá

$$L^2 = \bigcap_{\theta \in \Theta} L_\theta^2.$$

2.2. Definíció. A $h = h(x)$ statisztikát akkor nevezzük nulla torzítatlan becslésének, ha $E_\theta(h) \equiv 0$ minden $\theta \in \Theta$-ra. $\mathcal{H}_\theta$-val jelöljük az összes nulla torzítatlan becslései összességét, amelyek L_θ^2-hez tartoznak. Legyen

$$\mathcal{H} = \bigcap_{\theta \in \Theta} \mathcal{H}_\theta.$$

Az alábbi tárgyaláshoz szükségünk lesz a következő lemmára, amely RAO [47] nevéhez fűződik. A lemma általánosításai és analogonjai megtalálhatók STEIN [53], LEHMANN—SCHEFFÉ [32], LINNIK—RUHIN [34] és KLEBANOV—LINNIK—RUHIN [28] cikkeiben is.

2.3. Lemma. A $\tau(x)\in L^2$ (ill. $\tau(x)\in L^2_{\theta_0}$) statisztika akkor és csak akkor optimális (ill. a θ_0-pontban lokálisan optimális) torzítatlan becslése L^2-ben a $t(\theta)=E_\theta\tau$ paraméterfüggvénynek, ha

$$E_\theta(\tau h)\equiv 0, \quad \text{minden } h\in\overline{\mathscr{H}}\text{-ra}, \ \theta\in\Theta\text{-ra}$$

(ill. ha az adott θ_0-ra, $E_{\theta_0}(\tau h)\equiv 0$, minden $h\in\mathscr{H}_{\theta_0}$-ra).

Bizonyítás. (A lemma bizonyítása szerepel pl. [21]-ben.)

1°. Szükségesség. Tegyük fel, hogy $\tau\in L^2_{\theta_0}$ a θ_0-pontban lokálisan optimális torzítatlan becslése $t(\theta)=E_\theta\tau$-nak, és $h\in\mathscr{H}_{\theta_0}$.

Belátható, hogy tetszőleges állandó λ-ra $\tau+\lambda h$ $L^2_{\theta_0}$-hez tartozik és $\tau+\lambda h$ torzítatlan becslés $t(\theta)$-ra. Továbbá,

$$E_{\theta_0}(\tau+\lambda h-t)^2=E_{\theta_0}(\tau-t)^2+2\lambda E_{\theta_0}(\tau h)+\lambda^2 E_{\theta_0}(h^2).$$

Ha $E_{\theta_0}(\tau h)\neq 0$, akkor λ megfelelő választásával

$$2\lambda E_{\theta_0}(\tau h)+\lambda^2 E_{\theta_0}(h^2) < 0.$$

Ebből következik, hogy

$$E_{\theta_0}(\tau+\lambda h-t)^2 < E_{\theta_0}(\tau-t)^2,$$

ami ellentmond annak a feltevésünknek, hogy τ optimális becslés θ_0-ban.

Tehát $E_{\theta_0}(\tau h)=0$, minden $h\in\mathscr{H}_{\theta_0}$-ra.

2°. Elégségesség. Legyen $\tau(x)\in L^2_{\theta_0}$ olyan torzítatlan becslése $t(\theta)=E_\theta\tau$-nak, amelyre

$$E_{\theta_0}(\tau h)=0, \quad \text{minden } h\in\mathscr{H}_{\theta_0}\text{-ra}.$$

Legyen továbbá $\tau^*(x)\in L^2_{\theta_0}$ egy másik torzítatlan becslése $t(\theta)$-nak. Nyilván $\tau^*-\tau=h\in\mathscr{H}_{\theta_0}$ és mivel a feltevés szerint $E_{\theta_0}(\tau h)=0$,

$$E_{\theta_0}(\tau^*-t)^2=E_{\theta_0}(\tau-t)^2+2E_{\theta_0}(\tau h)+E_{\theta_0}(h^2)\geqq E_{\theta_0}(\tau-t)^2.$$

Hasonló módon bizonyíthatjuk a lemma másik állítását, amely az L^2-re vonatkozó optimalitásra vonatkozik.

Tekintsük az

$$x(j)=\xi(j)+\theta, \quad j=1,\ldots,n, \tag{2.58}$$

folyamatot. Legyen $\xi(j)$ a következő elsőrendű autoregressziós folyamat:

$$\begin{aligned}\xi(1)&=\varepsilon(1)\\ \xi(k)&=a\xi(k-1)+\varepsilon(k), \quad k=2,\ldots,n,\end{aligned} \tag{2.59}$$

ahol $\varepsilon(j)$, $j=1,\ldots,n$, független valószínűségi változók, $F_j(x)$ eloszlásfüggvénnyel,

és $a \neq 1$. Feltesszük, hogy $F_j(x)$ eleget tesz a következő feltételeknek:

$$\int x\, dF_j = 0, \quad j = 1, \ldots, n, \tag{2.60}$$

$$0 < \int x^2\, dF_j = \sigma_j^2 < \infty, \quad j = 1, \ldots, n. \tag{2.61}$$

Legyen $Q(\hat{l}^*) = q_0 \hat{l}^{*m} + \ldots + q_m$ $\hat{l}^*$-nak egy m-edfokú polinomja, ahol

$$\hat{l}^* = \sum_{j=1}^{n} c_j^{0*} x(j) \left(\sum_{j=1}^{n} c_j^{0*} = 1 \right) \text{ legjobb lineáris becslése } \theta\text{-nak.}$$

Figyelembe vesszük, hogy a (2.59) elsőrendű autoregressziós folyamat esetén

$$\xi(k) = a^{k-1}\varepsilon(1) + \ldots + a\varepsilon(k-1) + \varepsilon(k),$$

így

$$\hat{l}^* = \sum_{j=1}^{n} c_j^{0*} x(j) = \theta + \sum_{j=1}^{n} c_j^{0*} (a^{j-1}\varepsilon_1 + \ldots + \varepsilon_j).$$

Az

$$\int x^{2m}\, dF_j = E(\varepsilon_j^{2m}) < \infty, \quad j = 1, \ldots, n \tag{2.62}$$

feltétel mellett, a $Q(\hat{l}^*) = q_0 \hat{l}^{*m} + \ldots + q_m$ polinom torzítatlan becslése a $q(\theta) = E_\theta Q = c_0\theta^m + \ldots + c_m$ paraméterpolinomnak véges szórásnégyzettel minden $\theta \in \mathscr{R}^1$-re.

Megmutatjuk, hogy $Q(\hat{l}^*)$ optimalitása — $q(\theta)$ torzítatlan becsléseinek osztályában, a négyzetes veszteségfüggvénnyel — jellemző tulajdonsága az $x(j)$ *Gauss-folyamatnak*. Pontosabban mondva, igaz a következő

2.5. Tétel. Tegyük fel, hogy a (2.58)—(2.59) sémában $n \geqq 3$, $\theta \in \Theta \subset \mathscr{R}^1$, ahol Θ nem elfajult intervallum, és $F_j(x)$-re teljesülnek a (2.60)—(2.61)—(2.62) feltételek.

A $Q(\hat{l}^*) = q_0 \hat{l}^{*m} + \ldots + q_m$, $q_0 \neq 0$, $m \geqq 1$ polinom akkor és csak akkor optimális a $q(\theta) = E_\theta Q$ torzítatlan (L^2-ben levő) becsléseinek osztályában, a négyzetes veszteségfüggvénnyel, ha $x(j)$ *Gauss-folyamat.*

Bizonyítás. 1°. *Szükségesség.* A 2.3. lemma alapján $Q(\hat{l}^*)$ optimalitásából következik, hogy tetszőleges $h(x_2 - x_1, \ldots, x_n - x_1) = h \in \overline{\mathscr{H}}$-ra

$$E_\theta[Q(\hat{l}^*)h] = 0, \quad \text{minden } \theta \in \Theta\text{-ra.} \tag{2.63}$$

Belátható, hogy

$$E_\theta(Qh) = E_\theta[Q(\hat{l}^* + \theta)h] = E_0\{[q_0(\hat{l}^* + \theta)^m + \ldots + q_m]h\}. \tag{2.64}$$

A (2.64) jobb oldala nem más mint θ-nak polinomja, amelyben (2.63) miatt minden együttható nulla. A polinom $(m-1)$-edfokú tagjának együtthatója az

$$E_0[(q_0 m\hat{l}^* + q_1)h] = q_0 m E_0(\hat{l}^* h),$$

amiből a $q_0 \neq 0$ feltevés miatt

$$E_0(\hat{l}^* h) = 0, \quad \text{minden } h \in \overline{\mathscr{H}}\text{-ra.} \tag{2.65}$$

Legyen

$$h^* = E_0(\hat{l}^* | \mathbf{y}), \quad \mathbf{y} = (x_2 - x_1, \ldots, x_n - x_1).$$

Könnyű belátni, hogy $h^* \in \overline{\mathscr{H}}$. (2.65) alapján kapjuk a

$$0 = E_0(\hat{l}^* h^*) = E_0[E_0(\hat{l}^* h^* | \mathbf{y})] = E_0[h^* E_0(\hat{l}^* | \mathbf{y})] = E_0(h^*)^2 \tag{2.66}$$

összefüggést, ahonnan

$$E_0(\hat{l}^* | \mathbf{y}) = 0, \quad \text{m. m. } P_0 \tag{2.67}$$

adódik.

Ismeretes (lásd [21], § 7.8), hogy (2.67)-ből következik az $\varepsilon(1), \ldots, \varepsilon(n)$ valószínűségi változók eloszlásának normalitása.

2°. Elégségesség. Tegyük fel, hogy $\varepsilon(j)$ $N(0, \sigma_j^2)$ normális eloszlású valószínűségi változó. A 2.2. tétel bizonyításában láttuk, hogy az $\hat{l}^* = \sum_{j=1}^{n} c_j^{0*} x_j$ statisztika elégséges, és egyben teljes az $x(1), \ldots, x(n)$ valószínűségi változók együttes sűrűségfüggvényeinek összességére nézve.

A *Blackwell—Kolmogorov—Rao-tételből* következik, hogy $\hat{l}^*$-nak minden véges szórásnégyzetű függvénye L^2-ben optimális torzítatlan becslése a saját várható értéknek.

Ezzel a 2.5. tétel bizonyítását befejeztük.

3. Többdimenziós Gauss—Markov-folyamatok paraméterének optimális becslései

Legyen $\mathbf{x}(t) = \{x^{(i)}(t)\}_{i=1,\ldots,k}$ a következő k-dimenziós sztochasztikus folyamat:

$$\mathbf{x}(t) = \boldsymbol{\zeta}(t) + \boldsymbol{\theta}, \tag{2.68}$$

ahol $\boldsymbol{\theta} = (\theta^{(1)}, \ldots, \theta^{(k)})^T$ k-dimenziós vektorparaméter, $\xi(t) = \{\xi^{(i)}(t)\}_{i=1,\ldots,k}$ k-dimenziós stacionárius *Gauss—Markov-folyamat,* amely kielégíti a

$$\xi(j+1) = \mathbf{A}\xi(j) + \varepsilon(j+1) \tag{2.69}$$

sztochasztikus differenciaegyenletet, $\mathbf{A} = \{\alpha_{ij}\}_{\substack{i=1,\ldots,k \\ i=1,\ldots,k}}$ négyzetes mátrix és $\varepsilon(t)$ egy független azonos eloszlású *Gauss-sorozat.*

Tegyük fel, hogy $\varepsilon(t)$ komponensei függetlenek, továbbá $E\varepsilon(t) = \mathbf{0}$,

$$E\varepsilon^{(i)}(t)\varepsilon^{(j)}(t+s) = \begin{cases} \sigma_i^2, & \text{ha} \quad s = 0, \ j = i \\ 0, & \text{ha} \quad s \neq 0 \text{ vagy } s = 0, \ j \neq i. \end{cases}$$

A következőkben vizsgáljuk a $\boldsymbol{\theta}$ vektorparaméterre vonatkozó becslési feladatot, mégpedig $\boldsymbol{\theta}$ valamely becslésének optimalitását.

Az optimalitás fogalma az (1.19)-ben szereplő

$$r(\tilde{\boldsymbol{\theta}}; \boldsymbol{\theta}) = (\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta})(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta})^T \tag{2.70}$$

(négyzetes) veszteségfüggvényhez kapcsolódik.

Legyen a (2.69) egyenletben

$$\mathbf{A} = \begin{pmatrix} \varrho_1 & 0 & \dots & 0 \\ 0 & \varrho_2 & \dots & 0 \\ \vdots & & \ddots & \\ 0 & & & \varrho_k \end{pmatrix}, \tag{2.71}$$

azaz az $\mathbf{A}$ mátrix sajátértékei mind valósak és különbözőek. A (2.69) egyenlet ekkor ekvivalens a következőkkel:

$$\begin{aligned} \xi^{(1)}(j+1) &= \varrho_1 \xi^{(1)}(j) + \varepsilon^{(j)}(j+1), \\ \xi^{(2)}(j+1) &= \varrho_2 \xi^{(2)}(j) + \varepsilon^{(2)}(j+1), \\ &\dots\dots\dots\dots\dots\dots\dots\dots \\ \xi^{(k)}(j+1) &= \varrho_k \xi^{(k)}(j) + \varepsilon^{(k)}(j+1), \end{aligned} \tag{2.72}$$

ahol a $\xi^{(i)}(t)$, $i=1, \dots, k$ folyamatok elsőrendű autoregressziós stacionárius *Gauss-folyamatok.*

Legyenek $\mathbf{x}(1), \dots, \mathbf{x}(n)$ az $\mathbf{x}(t)$ folyamat megfigyelései, továbbá

$$\begin{aligned} \mathbf{X}_1 &= (x^{(1)}(1), \dots, x^{(1)}(n))^T, \\ \mathbf{X}_k &= (x^{(k)}(1), \dots, x^{(k)}(n))^T, \\ &\text{és} \\ \mathbf{X} &= (x^{(1)}(1), \dots, x^{(1)}(n); \dots; x^{(k)}(1), \dots, x^{(k)}(n))^T. \end{aligned} \tag{2.73}$$

Belátható, hogy az $\varepsilon(j)$ komponenseinek függetlenségéből következik, hogy $\mathbf{X}_1, \dots, \mathbf{X}_k$ függetlenek. Tehát az $\mathbf{X}$ kn-dimenziós valószínűségi vektorváltozó sűrűségfüggvénye meghatározható az alábbi reláció segítségével:

(2.74)

$$p_{\mathbf{X}}(x_{11}, \dots, x_{1n}, \dots, x_{k1}, \dots, x_{kn}; \boldsymbol{\theta}) = p_{\mathbf{X}_1}(x_{11}, \dots, x_{1n}; \theta^{(1)}) \dots p_{\mathbf{X}_k}(x_{k1}, \dots, x_{kn}; \theta^{(k)}).$$

(2.68)-ból és (2.72)-ből nyilvánvaló, hogy

$$x^{(i)}(j) = \xi^{(i)}(j) + \theta^{(i)}, \quad i = 1, \dots, k,$$

ahol $\xi^{(i)}(j)$ elsőrendű autoregressziós stacionárius *Gauss-folyamat.*

A 2.1. tételt a $p=1$ esetre alkalmazva (lásd a 2.1. tétel megjegyzését) kapjuk, hogy $n \geqq 3$ esetén az $\hat{l}^{(i)} = \sum_{j=1}^{n} c_{ij}^0 x^{(i)}(j)$ lineáris statisztika legjobb torzítatlan becslése $\theta^{(i)}$-nek és $\mathbf{X}_i = (x^{(i)}(1), \dots, x^{(i)}(n))^T$ sűrűségfüggvénye felírható a következő alakban:

$$p_{\mathbf{X}_i}(x_{i1}, \dots, x_{in}; \theta^{(i)}) = H_i(x_{i1}, \dots, x_{in}) G_i(\theta^{(i)}) \exp \frac{\theta^{(i)}}{c_i} \hat{l}^{(i)}.$$

Ily módon, (2.74)-ből

$$\begin{aligned} p_{\mathbf{X}}(x_{11}, \dots, x_{1n}, \dots, x_{k1}, \dots, x_{kn}; \boldsymbol{\theta}) = \\ = \prod_{i=1}^{k} H_i(x_{i1}, \dots, x_{in}) \prod_{i=1}^{k} G_i(\theta^{(i)}) \exp \sum_{i=1}^{k} \theta^{(i)} \frac{\hat{l}^{(i)}}{c_i}, \end{aligned} \tag{2.75}$$

adódik, ahonnan következik, hogy az $\hat{\mathbf{l}}=(\hat{l}^{(1)}, \ldots, \hat{l}^{(k)})^T$ statisztika elégséges a $P_{\mathbf{X}}(x_{11}, \ldots, x_{1n}, \ldots, x_{k1}, x_{kn}; \boldsymbol{\theta})$ sűrűségfüggvények összességére nézve. Vegyük figyelembe, hogy

$$\hat{\mathbf{l}} = \mathbf{cx}, \tag{2.76}$$

ahol

$$\mathbf{c} = \begin{pmatrix} c_{11}^0 \ldots c_{1n}^0 & 0 \ldots 0 & \ldots & 0 \ldots 0 \\ 0 \ldots 0 & c_{21}^0 \ldots c_{2n}^0 & \ldots & 0 \ldots 0 \\ \ldots\ldots\ldots & \ldots\ldots\ldots & \ldots & \ldots\ldots\ldots \\ 0 \ldots 0 & 0 \ldots 0 & \ldots & c_{k1}^0 \ldots c_{kn}^0 \end{pmatrix} \tag{2.77}$$

alakú $(k\times kn)$-es mátrix, és $\mathbf{X}$ kn-dimenziós normális eloszlású valószínűségi vektorváltozó, így az $\hat{\mathbf{l}}$ statisztika szintén normális eloszlású valószínűségi vektorváltozó, amiből könnyű belátni, hogy $\hat{\mathbf{l}}$ teljes statisztika.

Továbbá

$$E_{\boldsymbol{\theta}}(\hat{\mathbf{l}}) = \boldsymbol{\theta},$$

azaz $\hat{\mathbf{l}}$ torzítatlan becslése $\boldsymbol{\theta}$-nak.

A *Blackwell—Kolmogorov—Rao-tétel* többdimenziós variánsa alapján megállapíthatjuk, hogy az $\hat{\mathbf{l}}=\left(\sum_{j=1}^{n} c_{1j}^0 x^{(1)}(j), \ldots, \sum_{j=1}^{n} c_{kj}^0 x^{(k)}(j)\right)^T$ statisztika legjobb torzítatlan becslése $\boldsymbol{\theta}$-nak.

Megjegyzés. Ha $\mathbf{A}$ nem (2.71) alakú mátrix, de gyökei mind valósak és különbözők, a fenti állítás analogonját megkaphatjuk a következőképpen: Az $\mathscr{R}^k$ k-dimenziós Euklideszi térben az $\mathbf{S}=\{s_{ij}\}_{\substack{i=1,\ldots,k \\ i=1,\ldots,k}}$ nem szinguláris mátrix segítségével olyan lineáris leképezést hajtunk végre, amellyel az $\mathbf{A}$ mátrix a (2.71) alakú $\mathbf{B}=\mathbf{SAS}^{-1}$ mátrixba megy át, és feltesszük, hogy ekkor $\boldsymbol{\zeta}(j)=\mathbf{S}\boldsymbol{\varepsilon}(j)$ komponensei függetlenek lesznek. Az $\mathbf{y}(j)=\mathbf{Sx}(j)$, $\boldsymbol{\eta}(j)=\mathbf{S}\boldsymbol{\xi}(j)$, $\boldsymbol{\zeta}(j)=\mathbf{S}\boldsymbol{\varepsilon}(j)$ új vektorváltozókkal a (2.68) egyenletből az

$$\mathbf{y}(j) = \boldsymbol{\eta}(j)+\boldsymbol{\mu} \tag{2.78}$$

egyenlethez jutunk, ahol $\boldsymbol{\mu}=\mathbf{S}\boldsymbol{\theta}$ és $\boldsymbol{\eta}(j)$ k-dimenziós reguláris stacionárius *Gaus—Markov-folyamat*, mely kielégíti az

$$\boldsymbol{\eta}(j+1) = \mathbf{B}\boldsymbol{\eta}(j)+\boldsymbol{\zeta}(j) \tag{2.79}$$

sztochasztikus egyenletet.

Az előző esetben kapott eredményt a (2.78)—(2.79) sémára alkalmazva kapjuk, hogy az $\hat{\mathbf{m}}=(\hat{m}^{(1)}, \ldots, \hat{m}^{(k)})^T=\left(\sum_{j=1}^{n} d_{1j}^0 y_j^{(1)}, \ldots, \sum_{j=1}^{n} d_{kj}^0 y_j^{(k)}\right)^T$ statisztika legjobb torzítatlan becslése $\boldsymbol{\mu}$-nek.

Legyen

$$\hat{\mathbf{l}} = \mathbf{S}^{-1}\hat{\mathbf{m}}.$$

Bebizonyíthatjuk, hogy az $\hat{\mathbf{l}}$ statisztika elégséges és egyben teljes a $p(\mathbf{X}; \boldsymbol{\theta})$ sűrűségfüggvények összességére nézve, sőt $\hat{\mathbf{l}}$ legjobb torzítatlan becslése $\boldsymbol{\theta}$-nak.

A (2.68)—(2.69) séma esetén szintén érdekes megvizsgálni a $\boldsymbol{\theta}$ vektorparaméter maximum likelihood becslésének optimalitását a $\boldsymbol{\theta}$ torzítatlan becsléseinek osztályában.

Jelölje $\mathbf{B}(s)$ a $\xi(t)$ folyamat, $\mathbf{B}_\varepsilon(s)$ pedig az $\varepsilon(t)$ folyamat korrelációs függvényét.

Vegyük figyelembe, hogy feltevésünk szerint az $\varepsilon(t)$ folyamat komponensei függetlenek, másrészt a $\xi(1), \ldots, \xi(n)$ változók és az $\varepsilon(1), \ldots, \varepsilon(n)$ változók közötti (2.69)-nek megfelelő leképezés determinánsa 1, ily módon

$$p_{\xi(1),\ldots,\xi(n)}(u_{11}, \ldots, u_{1n}; \ldots; u_{k1}, \ldots, u_{kn}) = (2\Pi)^{-\frac{nk}{2}} |\mathbf{R}_{nk}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\mathbf{U}^T \mathbf{R}_{nk}^{-1}\mathbf{U}\right\}, \tag{2.80}$$

ahol

$$\mathbf{U}^T = (u_{11}, \ldots, u_{1n}; \ldots; u_{k1}, \ldots, u_{kn}),$$

és

$$\mathbf{R}_{nk}^{-1} = \begin{pmatrix}
a_{10} & b_{11} & 0 \ldots 0 & a'_{12} & b_{12} & 0 \ldots 0 & \ldots & a'_{1k} & b_{1k} & 0 \ldots 0 \\
b_{11} & a_{11} & b_{11} \ldots 0 & b'_{12} & a_{12} & b_{12} \ldots 0 & \ldots & b'_{1k} & a_{1k} & b_{1k} \ldots 0 \\
\vdots & & & \vdots & & & & & & \\
0 & & a_{11} b_{11} & 0 & & a_{12} b_{12} & \ldots & & & a_{1k} a_{1k} \\
0 & & b_{11} a''_{10} & 0 & & b'_{12} a''_{12} & \ldots & & & b'_{1k} a''_{1k} \\
a'_{12} & b'_{12} & 0 \ldots 0 & a_{20} & b_{22} & 0 \ldots 0 & & & & \\
b_{12} & a_{12} & b'_{12} \ldots 0 & b_{22} & a_{22} & b_{22} \ldots 0 & & & & \\
\vdots & & & & & & & & & \\
0 & & a_{12} b'_{12} & & & a_{22} b_{22} & & & & \\
0 & & b_{12} a''_{12} & & & b_{22} a''_{20} & & & & \\
\vdots & & & & & & & & & \\
a'_{1k} & b'_{1k} & 0 \ldots 0 & & & & & a_{k0} & b_{kk} & 0 \ldots 0 \\
b_{1k} & a_{1k} & b'_{1k} \ldots 0 & & & & & b_{kk} & a_{kk} & b_{kk} \ldots 0 \\
\vdots & & & & & & & \vdots & & \\
0 & & a_{1k} b'_{1k} & & & & & 0 & & a_{kk} b_{kk} \\
0 & & b_{1k} a''_{1k} & & & & & 0 & & b_{kk} a''_{k0}
\end{pmatrix}. \tag{2.81}$$

Az $\mathbf{R}_{nk}^{-1}$ mátrix elemei — $a_{ij}, a'_{ij}, a''_{i0}, b_{ij}, b'_{ij}$ — az $\mathbf{A}$, $\mathbf{B}(0)$ és $\mathbf{B}_\varepsilon(0)$ mátrixok segítségével határozhatók meg (lásd pl. Arató [1]).

(2.80)-ból kapjuk az $\mathbf{x}(1), \ldots, \mathbf{x}(n)$ együttes sűrűségfüggvényét a következő alakban:

$$p(\mathbf{X}, \boldsymbol{\theta}) = p(x_{11}, \ldots, x_{1n}; \ldots; x_{k1}, \ldots, x_{kn}; \boldsymbol{\theta}) = \tag{2.82}$$

$$= (2\Pi)^{-\frac{nk}{2}} |\mathbf{R}_{nk}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{X}-\boldsymbol{\Delta})^T \mathbf{R}_{nk}^{-1}(\mathbf{X}-\boldsymbol{\Delta})\right\},$$

ahol

$$\mathbf{X}-\boldsymbol{\Delta} = (x_{11}-\theta^{(1)}, \ldots, x_{1n}-\theta^{(1)}; \ldots; x_{k1}-\theta^{(k)}, \ldots, x_{kn}-\theta^{(k)})^T.$$

Legyen

$$\hat{\tau}^{(j)} = \mathbf{V}_j \mathbf{X}, \qquad j = 1, \ldots, k,$$
$$\hat{\boldsymbol{\tau}} = (\hat{\tau}^{(1)}, \ldots, \hat{\tau}^{(k)})$$

ahol $(a'_{jj}=a_{j0})$

$$\mathbf{V}_j = \begin{pmatrix} a'_{1j}+b_{1j} \\ b'_{1j}+a_{1j}+b_{1j} \\ \vdots \\ b_{1j}+a''_{1j} \\ \vdots \\ a'_{jj}+b_{jj} \\ b_{jj}+a_{jj}+b_{jj} \\ b_{jj}+a''_{j0} \\ \vdots \\ a'_{jk}+b'_{jk} \\ b_{jk}+a_{jk}+b'_{jk} \\ \vdots \\ b_{jk}+a''_{jk} \end{pmatrix}, \quad j = 1, \ldots, k.$$

Egyszerű számolással belátható, hogy

(2.83)

$$(\mathbf{X}-\boldsymbol{\Delta})^T \mathbf{R}_{nk}^{-1}(\mathbf{X}-\boldsymbol{\Delta}) = -2\sum_{j=1}^{k} \hat{\tau}^{(j)}\theta^{(j)} + \sum_{j=1}^{k} \{a_{j0}+(n-2)a_{jj}+a''_{j0}+2(n-1)b_{jj}\}\theta^{(j)2} +$$

$$+2\sum_{\substack{i,j=1 \\ (i\neq j)}}^{k} \{a'_{ij}+(n-2)a_{ij}+a''_{ij}+(n-1)b_{ij}+(n-1)b'_{ij}\}\theta^{(i)}\theta^{(j)} +$$

$$+f(x_{11}, \ldots, x_{1n}; \ldots; x_{k1}, \ldots, x_{kn}),$$

ahol az $f(x_{11}, \ldots, x_{1n}; \ldots; x_{k1}, \ldots, x_{kn})$ függvény konkrét felírására nem lesz szükségünk. Tehát

$$p(\mathbf{X}, \boldsymbol{\theta}) = H(\mathbf{X})G(\boldsymbol{\theta}) \exp \sum_{j=1}^{k} \hat{\tau}^{(j)}\theta^{(j)} = H(\mathbf{X})G(\boldsymbol{\theta}) \exp \hat{\boldsymbol{\tau}}^T\boldsymbol{\theta} \tag{2.84}$$

alakú, amiből nyilvánvaló, hogy a $\hat{\boldsymbol{\tau}}=(\hat{\tau}^{(1)}, \ldots, \hat{\tau}^{(k)})^T$ statisztika elégséges a $p(\mathbf{x}, \boldsymbol{\theta})$ sűrűségfüggvények összegére nézve.

Megmutatjuk továbbá, hogy $\hat{\boldsymbol{\tau}}$ teljes statisztika. Valóban belátható, hogy

$$\hat{\boldsymbol{\tau}} = (\hat{\tau}^{(1)}, \ldots, \hat{\tau}^{(k)})^T = (\mathbf{V}_1\mathbf{X}, \ldots, \mathbf{V}_k\mathbf{X})^T = \mathbf{V}\mathbf{X}, \tag{2.85}$$

ahol

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_1 \\ \mathbf{V}_2 \\ \vdots \\ \mathbf{V}_k \end{pmatrix}$$

blokkmátrix.

Mivel $\mathbf{X}$ $N_{nk}(\boldsymbol{\Delta}, \mathbf{R}_{nk})$ — normális eloszlású nk-dimenziós vektorváltozó, a normális eloszlás ismert tulajdonságai alapján megállapíthatjuk, hogy a $\hat{\boldsymbol{\tau}}$ statisztika eloszlása

szintén normális, mégpedig k-dimenziós normális eloszlás, $\mathbf{V\Delta}$ várható értékvektorral, $\mathbf{VR}_{nk}\mathbf{V}^T=\mathbf{\Sigma}$ kovariancia mátrixszal.

Így $\hat{\tau}$ sűrűségfüggvénye felírható a következő alakban:

$$\pi(\hat{\tau},\theta)=(2\Pi)^{-k/2}|\mathbf{\Sigma}|^{-1}\exp\left\{-\frac{1}{2}(\hat{\tau}-\mathbf{V\Delta})^T\mathbf{\Sigma}^{-1}(\hat{\tau}-\mathbf{V\Delta})\right\}= \tag{2.86}$$

$$=h(\hat{\tau})g(\theta)\exp\left\{\frac{1}{2}\mathbf{\Delta}^T\mathbf{V}^T\mathbf{\Sigma}^{-1}\hat{\tau}+\frac{1}{2}\hat{\tau}^T\mathbf{\Sigma}^{-1}\mathbf{V\Delta}\right\}=h(\hat{\tau})g(\theta)\exp\sum_{j=1}^{k}q_j(\theta)\hat{\tau}^{(j)}.$$

Lehmann tétele szerint (2.86)-ból következik, hogy a $\hat{\tau}$ statisztika teljes. Most vizsgáljuk meg a θ vektorparaméter maximum likelihood becslését.

Láttuk, hogy

$$p(\mathbf{X},\theta)=H(\mathbf{X})\exp\left\{\sum_{j=1}^{k}\hat{\tau}^{(j)}\theta^{(j)}-\sum_{\substack{i,j=1\\(i\neq j)}}^{k}(a'_{ij}+(n-2)a_{ij}+a''_{ij}+(n-1)b_{ij}+\right.$$

$$\left.+(n-1)b'_{ij})\theta^{(i)}\theta^{(j)}-\frac{1}{2}\sum_{j=1}^{k}(a_{j0}+(n-2)a_{jj}+a''_{j0}+2(n-1)b_{jj})\theta^{(j)^2}\right\}.$$

Az

$$L(\mathbf{X},\theta)=\log p(\mathbf{X},\theta)$$

maximum likelihood függvény bevezetésével a következő egyenletrendszert kapjuk:

$$\frac{\partial L}{\partial\theta^{(j)}}=\hat{\tau}^{(j)}-(a_{j0}+(n-2)a_{jj}+a''_{j0}+2(n-1)b_{jj})\theta^{(j)}- \tag{2.87}$$

$$-\sum_{\substack{i,j=1\\(i\neq j)}}^{k}(a'_{ij}+(n-2)a_{ij}+a''_{ij}+(n-1)b_{ij}+(n-1)b'_{ij})\theta^{(i)}=0,\quad j=1,\ldots,k.$$

Az

$$\omega_{ij}=a'_{ij}+(n-2)a_{ij}+a''_{ij}+(n-1)b_{ij}+(n-1)b'_{ij}, \tag{2.88}$$

$$\omega_j=\omega_{jj}=a_{j0}+(n-2)a_{jj}+a''_{j0}+2(n-1)b_{jj},$$

és

$$\mathbf{\Omega}=\begin{pmatrix}\omega_1 & \omega_{12} & \ldots & \omega_{1k}\\ \omega_{21} & \omega_2 & \ldots & \omega_{2k}\\ \ldots & \ldots & \ldots & \ldots\\ \omega_{k1} & \omega_{k2} & \ldots & \omega_k\end{pmatrix} \tag{2.89}$$

jelölésekkel, a (2.87) egyenletrendszer a

$$\theta^T\mathbf{\Omega}=\hat{\tau}^T,$$

vagy

$$\mathbf{\Omega}\theta=\hat{\tau} \tag{2.90}$$

alakra hozható. (Vegyük észre, hogy az $\mathbf{\Omega}$ mátrix szimmetrikus a főátlóra nézve.) Tegyük fel, hogy $\mathbf{\Omega}$ nem szinguláris mátrix — azaz létezik $\mathbf{\Omega}^{-1}$ ($\mathbf{\Omega}$ inverze),

akkor (2.90)-ből $\boldsymbol{\theta}$ maximum likelihood becslését a következő alakban kapjuk:

$$\hat{\boldsymbol{\theta}}^* = \boldsymbol{\Omega}^{-1}\hat{\boldsymbol{\tau}}. \tag{2.91}$$

Tehát $\boldsymbol{\theta}$ maximum likelihood becslése a $\hat{\boldsymbol{\tau}}$ elégséges teljes statisztikának függvénye. Belátható továbbá, hogy

$$E_{\boldsymbol{\theta}}\hat{\boldsymbol{\tau}} = (E_{\boldsymbol{\theta}}\hat{\tau}^{(1)}, \ldots, E_{\boldsymbol{\theta}}\hat{\tau}^{(k)})^T = \begin{pmatrix} \omega_1 & \omega_{12} & \ldots & \omega_{1k} \\ \omega_{21} & \omega_2 & \ldots & \omega_{2k} \\ \ldots & \ldots & \ldots & \ldots \\ \omega_{k1} & \omega_{k2} & \ldots & \omega_k \end{pmatrix} \begin{pmatrix} \theta^{(1)} \\ \vdots \\ \theta^{(k)} \end{pmatrix} = \boldsymbol{\Omega}\boldsymbol{\theta},$$

amiből adódik, hogy

$$E_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}^*) = \boldsymbol{\Omega}^{-1}E_{\boldsymbol{\theta}}(\hat{\boldsymbol{\tau}}) = \boldsymbol{\theta},$$

azaz $\hat{\boldsymbol{\theta}}^*$ torzítatlan becslése $\boldsymbol{\theta}$-nak.

A *Blackwell—Kolmogorov—Rao-tétel* többdimenziós variánsa szerint $\hat{\boldsymbol{\theta}}^*$ legjobb torzítatlan becslése $\boldsymbol{\theta}$-nak a (2.70) alakú veszteségfüggvénnyel.

Ily módon érvényes az alábbi

2.6. Tétel. A (2.68)—(2.69) séma esetén a $\boldsymbol{\theta}$ vektorparaméter maximum likelihood becslése legjobb torzítatlan becslése $\boldsymbol{\theta}$-nak a (2.70) veszteségfüggvénnyel.

III. FEJEZET

1. Gauss-folyamatok jellemzése az eltolási paraméter Bayes-féle becslésének linearitásával

Az utóbbi évek egyik érdekes matematikai statisztikai problémája a következő:

Legyenek $\theta_1, \ldots, \theta_p$; $x_1, \ldots, x_n$ valószínűségi változók, ahol $x_1, \ldots, x_n$ megfigyelhetők, $\theta_1, \ldots, \theta_p$ pedig ismeretlenek, amelyeket becsülnünk kell az $x_1, \ldots, x_n$ változók értékei alapján.

Legyen

$$\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^T, \quad \mathbf{X} = (x_1, \ldots, x_n)^T.$$

Az $r(\tilde{\boldsymbol{\theta}}, \boldsymbol{\theta})$ veszteségfüggvény esetén a rizikó definíció szerint a következő:

$$R(\tilde{\boldsymbol{\theta}}; \boldsymbol{\theta}) = E_{\boldsymbol{\theta}}r(\tilde{\boldsymbol{\theta}}, \boldsymbol{\theta}).$$

Legyen $\pi(\boldsymbol{\theta})$ egy valószínűségi mérték. Azt a becslést, amely az $E_\pi R(\tilde{\boldsymbol{\theta}}; \boldsymbol{\theta})$ *Bayes-féle veszteséget* minimalizálja, $\boldsymbol{\theta}$ *Bayes-féle becslésének* nevezzük.

Valós θ paraméter és négyzetes veszteségfüggvény esetén, a θ a posteriori középértéke $E(\theta|X)$ — θ-nak X-re vonatkozó regressziója — minimalizálja $E_\pi E_\theta|\tilde{\theta}-\theta|^2$-et, azaz $E(\theta|X)$ *Bayes-féle becslése* θ-nak (lásd pl. RAO [46]).

A θ valós eltolási paraméter becslésének problémáit az

$$x(j) = \varepsilon(j)+\theta, \quad j = 1, \ldots, n \tag{3.1}$$

független megfigyelési séma esetén PITMAN [42] alapvető cikkéből kiindulva többen vizsgálták.

Most vizsgáljuk a problémát a következő függő megfigyelési séma esetén:

$$y(j) = \xi(j) + \theta, \quad j = 1, \ldots, n.$$

Tegyük fel, hogy $E\xi(j)=0$, $0 < E\xi^2(j) < \infty$; $\theta \in \mathscr{R}^1$, $\int_{-\infty}^{+\infty} \theta^2 d\pi(\theta) < \infty$.

Ha $f(u_1, \ldots, u_n)$ a $\xi(1), \ldots, \xi(n)$ valószínűségi változók együttes sűrűségfüggvénye, akkor $y(1), \ldots, y(n)$ együttes sűrűségfüggvénye az $f(u_1-\theta, \ldots, u_n-\theta)$ függvény.

Ismeretes, hogy θ legjobb becslése az a posteriori várható érték:

$$\hat{\theta} = E(\theta | y_1, \ldots, y_n) = \frac{\int_{-\infty}^{+\infty} \theta f(y_1-\theta, \ldots, y_n-\theta)\, d\Pi(\theta)}{\int_{-\infty}^{+\infty} f(y_1-\theta, \ldots, y_n-\theta)\, d\Pi(\theta)}. \tag{3.2}$$

A (3.2) becslés felépítéséhez szükséges ismerni θ eloszlását és $\xi(1), \ldots, \xi(n)$ együttes sűrűségfüggvényét. Felmerül a kérdés, hogy a (3.2) becslés felépítéséhez milyen esetben elegendő ismerni a $\pi(\theta)$ és a $\xi(1), \ldots, \xi(n)$ együttes eloszlásának első- és másodrendű momentumait. Más szóval a probléma a következőképpen fogalmazható: milyen esetben teljesül az

$$E(\theta | y_1, \ldots, y_n) = \hat{E}(\theta | y_1, \ldots, y_n) \tag{3.3}$$

reláció, ahol a jobb oldalon szereplő $\hat{E}(\theta | y_1, \ldots, y_n)$ általánosított várható érték a θ legjobb lineáris becslését jelenti.

A következőkben a (3.3) reláció teljesülésének különböző alakjait vizsgáljuk. Mint látni fogjuk, a (3.3) reláció teljesülése az $y(j)$ folyamat *Gaussi* voltának és a $\pi(\theta)$ eloszlás normalitásának jellemző tulajdonsága. A később ismertetendő eredmények hasonlóak a KAGAN—KARPOV [22] cikkben szereplő eredményekhez, amelyben a problémát a (3.1) független megfigyelési séma esetén vizsgálták.

Megemlítjük, hogy a továbbiakban hacsak nincs külön megjegyzés, a tárgyalás a négyzetes veszteségfüggvényhez kapcsolódik.

Tekintsük az

$$y(j) = \xi(j) + \theta, \quad j = 1, 2, \ldots, n \tag{3.4}$$

sztochasztikus folyamatot.

Tegyük fel, hogy $\xi(j)$ a következő elsőrendű autoregressziós folyamat:

$$\xi(1) = \varepsilon(1)$$
$$\xi(k) = a\xi(k-1) + \varepsilon(k), \quad k = 2, \ldots, n, \tag{3.5}$$

ahol $\varepsilon(j)$ független sorozat, $F_j(x)$ eloszlással.

A θ ($\theta \in \mathscr{R}^1$) paraméter becslésének *Bayes-féle megfogalmazásában* feltesszük, hogy θ olyan valószínűségi változó, amelynek (a priori) eloszlása $\pi(\theta)$, továbbá θ független $\varepsilon(1), \ldots, \varepsilon(n)$-től.

A sztochasztikus folyamatok elméletéből (lásd pl. [12]) ismeretes, hogy θ leg-

obb (torzítatlan) becslése a

$$\hat{\theta} = E(\theta|y(1), \ldots, y(n))$$

statisztika, és θ legjobb lineáris (torzítatlan) becslése

$$\hat{\theta}_1 = \hat{E}(\theta|y(1), \ldots, y(n)) = \hat{E}(\theta|H_y),$$

ahol H_y jelöli az $y(1), \ldots, y(n)$ változók által generált zárt lineáris sokaságot, és $\hat{E}(\theta|H_y)$ θ-nak H_y-ra való vetületét jelenti.

Abban az esetben, amikor a $(\theta, y(1), \ldots, y(n))$ valószínűségi vektorváltozó *Gauss-eloszlású,* akkor θ legjobb becslése megegyezik legjobb lineáris becslésével, azaz

$$E(\theta|y_1, \ldots, y_n) = \hat{E}(\theta|H_y). \tag{3.6}$$

Vizsgáljuk ennek a megfordítását. Megmutatjuk, hogy a (3.4)—(3.5) séma esetén ha (3.6) teljesül, akkor bizonyos egyszerű feltétel mellett, $\Pi(\theta)$, $F_j(x)$ *Gauss-eloszlás-függvény.*

A bizonyítás a KAGAN—KARPOV [22] cikk gondolatmenetén alapszik.

Bevezetjük az

$$\begin{aligned} x(1) &= \varepsilon(1)+\theta \\ x(2) &= \varepsilon(2)+\theta \\ &\cdots\cdots\cdots\cdots \\ x(n) &= \varepsilon(n)+\theta \end{aligned} \tag{3.7}$$

változókat. Belátható, hogy

$$\begin{aligned} y(1) &= x(1) \\ y(2) &= x(2)+a\varepsilon(1) \\ &\cdots\cdots\cdots\cdots\cdots \\ y(n) &= x(n)+a^{n-1}\varepsilon(1)+\ldots+a\varepsilon(n-1). \end{aligned} \tag{3.8}$$

H_x-szel jelöljük az $x(1), \ldots, x(n)$ változók által generált zárt lineáris sokaságot. (3.8)-bál következik, hogy H_y megegyezik H_x-szel, amiből

$$\hat{E}(\theta|H_y) = \hat{E}(\theta|H_x) \tag{3.9}$$

adódik.

Legyenek

$$E\varepsilon(j) = \mu_1^{(j)}, \quad E\theta = \alpha_1,$$

$$E[\varepsilon(j)-\mu_1^{(j)}]^k = \mu_k^{(j)}, \; E(\theta-\alpha_1)^k = \alpha_k, \quad k = 2, 3, \ldots$$

Tegyük fel, hogy

$$0 < \alpha_2 < \infty; \quad 0 < \mu_2^{(j)} < \infty, \quad j = 1, 2, \ldots, n. \tag{3.10}$$

Vegyük figyelembe, hogy $\hat{E}(\theta|H_x)$ θ-nak H_x-re való vetületét jelenti, belátható (lásd [22] is), hogy az $x(j)=\varepsilon(j)+\theta$ séma esetén θ legjobb lineáris becslése

a tágabb értelemben vett *Bayes-féle becslés* és a következő alakú:

$$\hat{E}(\theta|H_x) = c_0 + \sum_{j=1}^{n} c_j x(j), \tag{3.11}$$

ahol

$$c_0 = \frac{\alpha_1 - \sum_{j=1}^{n} \frac{\alpha_2 \mu_1^{(j)}}{\mu_2^{(j)}}}{1 + \sum_{j=1}^{n} \frac{\alpha_2}{\mu_2^{(j)}}}, \quad c_j = \frac{\frac{\alpha_2}{\mu_2^{(j)}}}{1 + \sum_{j=1}^{n} \frac{\alpha_2}{\mu_2^{(j)}}}, \quad j = 1, \ldots, n. \tag{3.12}$$

Ily módon (3.9), (3.11)-ből kapjuk, hogy az $y(j)=\xi(j)+\theta$ folyamat esetén, ahol $\xi(j)$ (3.5) alakú elsőrendű autoregressziós folyamat, és $\varepsilon(j)$-re, θ-ra teljesül a (3.10) feltétel, θ legjobb lineáris becslése

$$\hat{\theta}_1 = \hat{E}(\theta|H_y) = c_0 + \sum_{j=1}^{n} c_j y(j) - (ac_2 + \ldots + a^{n-1}c_n)\varepsilon_1 - \ldots - ac_n\varepsilon_{n-1}, \tag{3.13}$$

ahol $c_0, c_j, j=1, \ldots, n$ együtthatók meghatározhatók (3.12) segítségével.

Legyenek

$$f_j(t) = \int_{-\infty}^{+\infty} e^{itx}\, dF_j(x), \quad g_j(t) = \frac{f_j'(t)}{f_j(t)},$$
$$\varphi(t) = \int_{-\infty}^{+\infty} e^{it\theta}\, d\pi(\theta), \quad \psi(t) = \frac{\varphi'(t)}{\varphi(t)}, \qquad j = 1, \ldots, n$$

ahol a $g_j(t), \psi(t)$ függvények a $t=0$ pont valamilyen környezetében vannak értelmezve.

Tegyük most fel, hogy θ legjobb becslése megegyezik θ legjobb lineáris becslésével, így (3.6)-ból és (3.13)-ból következik, hogy

$$E(\theta|y_1, \ldots, y_n) = c_0 + \sum_{j=1}^{n} c_j y_j - (ac_2 + \ldots + a^{n-1}c_n)\varepsilon_1 - \ldots - ac_n\varepsilon_{n-1}. \tag{3.14}$$

Az általánosság korlátozása nélkül feltehetjük, hogy $\alpha_1=\mu_1^{(j)}=0$, $j=1, \ldots, n$; tehát $c_0=0$ (ha szükséges, használjuk a $\theta^*=\theta-\alpha_1$; $x_j^*=x_j-\alpha_1-\mu_1^{(j)}$ új változókat). Legyenek

$$\begin{aligned} -(ac_2 + \ldots + a^{n-1}c_n) &= r_1, \\ \ldots\ldots\ldots\ldots\ldots\ldots & \\ -ac_n &= r_{n-1}, \end{aligned} \tag{3.15}$$

akkor (3.14) az

$$E(\theta|y_1, \ldots, y_n) = \sum_{j=1}^{n} c_j y_j + \sum_{j=1}^{n-1} r_j \varepsilon_j \tag{3.16}$$

alakra hozható.

A feltételes várható érték jól ismert tulajdonsága és (3.16) alapján kapjuk az

$$E\left(\theta \exp i \sum_{j=1}^{n} t_j y_j\right) = E\left\{\left(\exp i \sum_{j=1}^{n} t_j y_j\right) E(\theta | y_1, \ldots, y_n)\right\} = \tag{3.17}$$

$$= E\left(\sum_{j=1}^{n} c_j y_j \exp i \sum_{j=1}^{n} t_j u_j\right) + E\left(\sum_{j=1}^{n-1} r_j \varepsilon_j \exp i \sum_{j=1}^{n} t_j y_j\right)$$

összefüggést. Mivel

$$\sum_{j=1}^{n} t_j y_j = \sum_{j=1}^{n} t_j (a^{j-1}\varepsilon_1 + \ldots + a\varepsilon_{j-1} + \varepsilon_j + \theta) = \tag{3.18}$$

$$= v_1\varepsilon_1 + v_2\varepsilon_2 + \ldots + v_n\varepsilon_n + \tau\theta =$$

$$= \tau\theta + \sum_{j=1}^{n} v_j \varepsilon_j,$$

ahol

$$\begin{aligned} v_1 &= t_1 + at_2 + \ldots + a^{n-1} t_n \\ v_2 &= t_2 + at_3 + \ldots + a^{n-2} t_n \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ v_n &= t_n \end{aligned} \tag{3.19}$$

és

$$\tau = \sum_{j=1}^{n} t_j = v_1 + (1-a) \sum_{j=2}^{n} v_j,$$

látható, hogy

$$E\left(\theta \exp i \sum_{j=1}^{n} t_j y_j\right) = (E\theta^{i\tau\theta + i \sum_{j=1}^{n} v_j \varepsilon_j}) = \tag{3.20}$$

$$= E(\theta e^{i\tau\theta}) E(e^{i \sum_{j=1}^{n} v_j \varepsilon_j}) =$$

$$= -i\varphi'(\tau) \prod_{j=1}^{n} f_j(v_j).$$

Tekintsük (3.17) jobb oldalát. Egyrészt,

$$E\left(\sum_{j=1}^{n} c_j y_j \exp i \sum_{j=1}^{n} t_j y_j\right) = E\left\{\left(\sum_{j=1}^{n} u_j \varepsilon_j + \gamma\theta\right) \exp i \left(\sum_{j=1}^{n} v_j \varepsilon_j + \tau\theta\right)\right\} = \tag{3.21}$$

$$= E\left\{\sum_{j=1}^{n} u_j \varepsilon_j e^{i \sum_{j=1}^{n} v_j \varepsilon_j + i\tau\theta} + \gamma\theta e^{i \sum_{j=1}^{n} v_j \varepsilon_j + i\tau\theta}\right\} =$$

$$= E\left(\sum_{j=1}^{n} u_j \varepsilon_j e^{i \sum_{j=1}^{n} v_j \varepsilon_j}\right) E(e^{i\tau\theta}) + E(\gamma\theta e^{i\tau\theta}) E(e^{i \sum_{j=1}^{n} v_j \varepsilon_j}) =$$

$$= -i \sum_{j=1}^{n} u_j f_j'(v_j) \prod_{k \neq j} f_k(v_k) \cdot \varphi(\tau) - i\gamma\varphi'(\tau) \prod_{j=1}^{n} f_j(v_j),$$

ahol

$$u_1 = c_1 + ac_2 + \ldots + a^{n-1}c_n$$
$$u_2 = c_2 + ac_3 + \ldots + a^{n-2}c_n \tag{3.22}$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$
$$u_n = c_n$$

és

$$\gamma = c_1 + c_2 + \ldots + c_n.$$

Másrészt

$$E\left(\sum_{j=1}^{n-1} r_j \varepsilon_j \exp i \sum_{j=1}^{n} t_j y_j\right) = E\left(\sum_{j=1}^{n-1} r_j \varepsilon_j e^{i\left(\sum_{j=1}^{n} v_j \varepsilon_j + \tau\theta\right)}\right) = \tag{3.23}$$

$$= E\left(\sum_{j=1}^{n-1} r_j \varepsilon_j e^{i \sum_{j=1}^{n} v_j \varepsilon_j}\right) E(e^{i\tau\theta}) = E\left(\sum_{j=1}^{n-1} r_j \varepsilon_j e^{i \sum_{j=1}^{n-1} v_j \varepsilon_j}\right) E(e^{iv_n \varepsilon_n}) E(e^{i\tau\theta}) =$$

$$= -i \sum_{j=1}^{n-1} r_j f_j'(v_j) \prod_{\substack{k=1 \\ (k \neq j)}}^{n-1} f_k(v_k) f_n(v_n) \varphi(\tau).$$

Ily módon (3.17), (3.20), (3.21) és (3.23) alapján kapjuk a következő egyenletet:

$$\varphi'(\tau) \prod_{j=1}^{n} f_j(v_j) = \varphi(\tau) \sum_{j=1}^{n} u_j f_j'(v_j) \prod_{k \neq j} f_k(v_k) + \tag{3.24}$$

$$+ \gamma \varphi'(\tau) \prod_{j=1}^{n} f_j(v_j) + \varphi(\tau) f_n(v_n) \sum_{j=1}^{n-1} r_j f_j'(v_j) \prod_{\substack{k=1 \\ k \neq j}}^{n-1} f_k(v_k).$$

Mivel $\varphi(\tau) \prod_{j=1}^{n} f_j(v_j)$ folytonos a $(v_1, \ldots, v_n) = (0, \ldots, 0)$ pontban és $\varphi(0) \prod_{j=1}^{n} f_j(0) = 1$, létezik olyan ε-környezet, amelyben $\varphi(\tau) \prod_{j=1}^{n} f_j(v_j)$ különbözik 0-tól.

(3.24) mindkét oldalát $\varphi(\tau) \prod_{j=1}^{n} f_j(v_j)$-vel osztva, kapjuk az

$$(1-\gamma) \Psi(\tau) = \sum_{j=1}^{n} u_j g_j(v_j) + \sum_{j=1}^{n-1} r_j g_j(v_j)$$

összefüggést, amiből $u_j + r_j = c_j$, $j = 1, \ldots, n-1$; $u_n = c_n$ miatt

$$(1-\gamma) \Psi(\tau) = \sum_{j=1}^{n} c_j g_j(v_j), \quad \text{ahol} \quad \tau = v_1 + (1-a) \sum_{j=1}^{n} v_j. \tag{3.25}$$

(3.25)-ből következik, hogy

$$(1-\gamma) \Psi[v_1 + (1-a)v_2 + \ldots + (1-a)v_n] = c_1 g_1(v_1) + \ldots + c_n g_n(v_n). \tag{3.26}$$

A (3.10) feltétel teljesülése esetén

$$c_j > 0, \quad 1 - \gamma = \frac{1}{1 + \sum_{j=1}^{n} \frac{\alpha_2}{\mu_2^{(j)}}} > 0.$$

Tegyük fel, hogy $1-a\neq 0$. Ily módon (3.26)-ból kapjuk a

(3.27) $\Psi[v_1+(1-a)v_2+\ldots+(1-a)v_n]=\bar{g}_1(v_1)+\bar{g}_2((1-a)v_2)+\ldots+\bar{g}_n((1-a)v_n)$

egyenletet, amely nem más, mint a *Pexider-féle egyenlet,* (lásd ACZÉL [0]), ahol

$$\bar{g}_1(t)=\frac{c_1}{1-\gamma}g_1(t),$$

$$\bar{g}_j(t)=\frac{c_j}{1-\gamma}g_j\left(\frac{t}{1-a}\right),\quad j=2,\ldots,n.$$

(3.27) megoldása lineáris függvénye t-nek, azaz $\psi(t)$ lineáris függvénye t-nek, és így $g_j(t), j=1,\ldots,n$ is lineáris függvénye t-nek a $t=0$ pont valamilyen környezetében. (A (3.25) egyenlet csak a 0 valamilyen környezetében érvényes.) Így láthatjuk, hogy

$$\varphi(t)=\exp Q(t),\quad |t|<\varepsilon \text{ esetén},$$
$$f_j(t)=\exp Q_j(t),\quad j=1,\ldots,n,$$

ahol $Q(t), Q_j(t)$ másodfokú polinomok

$$\left(\Psi(t)=\frac{\varphi'(t)}{\varphi(t)}=\frac{d}{dt}\log\varphi(t);\quad g_j(t)=\frac{f_j'(t)}{f_j(t)}=\frac{d}{dt}\log f_j(t)\right).$$

Az analitikus folytatás elvével látjuk, hogy

$$\left.\begin{aligned}\varphi(t)&=\exp Q(t),\\ f_j(t)&=\exp Q_j(t),\end{aligned}\right\}\quad \text{minden } t\text{-re},$$

és a *Marcinkiewicz-tétel* szerint $\varphi(t), f_j(t)$ normális eloszlás karakterisztikus függvénye, ahonnan adódik, hogy $\pi(\theta), F_j(x)$ *Gauss-eloszlás.* Tehát igazoltuk a következő állítást:

3.1. Tétel. Tegyük fel, hogy a (3.4)—(3.5) sémában $n\geqq 2$, $1-a>0$ és teljesül a (3.10) feltétel.

A θ paraméter legjobb becslése akkor és csak akkor egyezik meg a legjobb lineáris becsléssel, ha $\pi(\theta)$ normális eloszlás és $y(j)$ *Gauss-folyamat.*

Most megmutatjuk, hogy a fenti állítás — bizonyos egyszerű feltételek teljesülése esetén — érvényben marad abban az esetben is, amikor (3.4)-ben $\xi(j)$ a következő p-edrendű autoregressziós folyamat:

(3.28)
$$\begin{aligned}\xi(1)&=\varepsilon(1)\\ \xi(2)+a_1\xi(1)&=\varepsilon(2)\\ &\cdots\cdots\cdots\cdots\cdots\cdots\\ \xi(p)+a_1\xi(p-1)+\ldots+a_{p-1}\xi(1)&=\varepsilon(p)\\ \xi(k)+a_1\xi(k-1)+\ldots+a_p\xi(k-p)&=\varepsilon(k),\quad k=p+1,\ldots,n,\end{aligned}$$

ahol $\varepsilon(j), j=1,2,\ldots,n,$ független sorozat, $F_j(x)$ eloszlással és $(\varepsilon_1,\ldots,\varepsilon_n)$ független θ-tól.

Ismeretes (lásd (2.39) formulát), hogy a (3.28) p-edrendű autoregressziós folyamat esetén $\xi(j)$ előállítható az $\varepsilon(j)$ sorozat segítségével a

$$\xi(j) = \sum_{k=0}^{j-1} b_k \varepsilon(j-k) \tag{3.29}$$

alakban, ahol a b_k együtthatók $a_1, \ldots, a_p$ segítségével meghatározhatók.

Ily módon

$$\begin{aligned} y(1) &= \varepsilon(1)+\theta \\ y(2) &= \varepsilon(2)+b_1\varepsilon(1)+\theta \\ &\cdots\cdots\cdots\cdots\cdots \\ y(n) &= \varepsilon(n)+b_1\varepsilon(n-1)+\ldots+b_{n-1}\varepsilon(1)+\theta. \end{aligned} \tag{3.30}$$

Az

$$\begin{aligned} x(1) &= \varepsilon(1)+\theta \\ x(2) &= \varepsilon(2)+\theta \\ &\cdots\cdots\cdots\cdots \\ x(n) &= \varepsilon(n)+\theta \end{aligned} \tag{3.31}$$

új változók bevezetésével könnyű belátni, hogy

$$\begin{aligned} y(1) &= x(1) \\ y(2) &= x(2)+b_1\varepsilon(1) \\ &\cdots\cdots\cdots\cdots\cdots \\ y(n) &= x(n)+b_1\varepsilon(n-1)+\ldots+b_{n-1}\varepsilon(1), \end{aligned} \tag{3.32}$$

amiből következik, hogy H_y és H_x megegyeznek egymással, azaz fennáll

$$\hat{E}(\theta|H_y) = \hat{E}(\theta|H_x). \tag{3.33}$$

Tehát (3.33), (3.11) és (3.32) alapján

$$\hat{E}(\theta|H_y) = c_0 + \sum_{j=1}^{n} c_j x(j) = c_0 + \sum_{j=1}^{n} y(j) + \sum_{j=1}^{n-1} r_j^* \varepsilon(j), \tag{3.34}$$

ahol

$$\begin{aligned} r_1^* &= -(c_2 b_1 + c_3 b_2 + \ldots + c_n b_{n-1}) \\ r_2^* &= -(c_3 b_1 + \ldots + c_n b_{n-2}) \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ r_{n-1}^* &= -c_n b_1, \end{aligned} \tag{3.35}$$

és c_j (3.12)-ből meghatározható.

Tegyük most fel, hogy teljesül az

$$E(\theta|y_1, \ldots, y_n) = \hat{E}(\theta|H_y)$$

reláció.

Ebből kiindulva, az $\alpha_1=\mu_1^{(j)}=0$ feltevéssel, a korábban használt módszerhez hasonlóan $\varphi(t)$-re és $f_j(t)$-re a

$$\varphi'(t)\prod_{j=1}^{n}f_j(v_j^*)=\sum_{j=1}^{n}u_j^*f_j'(v_j^*)\prod_{k\neq j}f_k(v_j^*)\varphi(\tau)+\gamma\varphi'(\tau)\prod_{j=1}^{n}f_j(v_j^*)+ \tag{3.36}$$

$$+\varphi(\tau)f_n(v_n^*)\sum_{j=1}^{n-1}r_j^*f_j'(v_j^*)\prod_{\substack{k=1\\k\neq j}}^{n-1}f_k(v_k^*)$$

egyenletet kapjuk, ahol

$$\begin{aligned}
\gamma &= c_1+\ldots+c_n,\\
\tau &= t_1+\ldots+t_n,\\
u_1^* &= c_1+b_1c_2+\ldots+b_{n-1}c_n = c_1-r_1^*\\
u_2^* &= c_2+b_1c_3+\ldots+b_{n-2}c_n = c_2+r_2^*\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
u_n^* &= c_n \qquad\qquad\qquad = c_n-r_n^* \quad (r_n^*=0),\\
v_1^* &= t_1+b_1t_2+\ldots+b_{n-1}t_n\\
v_2^* &= t_2+b_1t_3+\ldots+b_{n-2}t_n\\
&\ldots\ldots\ldots\ldots\ldots\ldots\\
v_n^* &= t_n.
\end{aligned}$$

Egyszerűsítés után a következő egyenlethez jutunk:

$$(1-\gamma)\Psi(\tau)=\sum_{j=1}^{n}c_jg_j(v_j^*), \tag{3.37}$$

ahol $\tau=\sum_{j=1}^{n}t_j=(1+v_1+\ldots+v_{n-1})v_n^*+(1+v_1+\ldots+v_{n-2})v_{n-1}^*+\ldots+(1+v_1)v_2^*+v_1^*$.

(v_j, $j=1,\ldots,n$ meghatározható $a_1,\ldots,a_p$ segítségével, lásd a (2.46) formulát.) (3.10) feltétel teljesülése esetén

$$c_j>0, \quad 1-\gamma>0.$$

Tegyük fel, hogy

$$(1+v_1)\ldots(1+v_1+\ldots+v_{n-1})\neq 0.$$

A (3.37) egyenletből következik, hogy

$$(1-\gamma)\Psi[v_1^*+(1+v_1)v_2^*+\ldots+(1+v_1+\ldots+v_{n-1})v_n^*]=c_1g_1(v_1^*)+\ldots+c_ng_n(v_n^*),$$

vagy ezzel ekvivalens

$$\Psi[v_1^*+(1+v_1)v_2^*+\ldots+(1+v_1+\ldots+v_{n-1})v_n^*]= \tag{3.38}$$

$$=g_1^*(v_1^*)+g_2^*((1+v_1)v_2^*)+\ldots+g_n^*((1+v_1+\ldots+v_{n-1})v_n^*),$$

ahol

$$g_1^*(t) = \frac{c_1}{1-\gamma} g_1(t),$$

$$g_j^*(t) = \frac{c_j}{1-\gamma} g_j\left(\frac{t}{1+v_1+\ldots+v_{j-1}}\right), \quad j = 2, \ldots, n.$$

A (3.38) egyenlet nem más, mint *Pexider-egyenlet,* melynek megoldása lineáris függvénye t-nek. Ily módon láthatjuk, hogy $\varphi(t), f_j(t)$ normális eloszlás karakterisztikus függvénye.

A fentieket összefoglalva, a következő tételt kapjuk:

3.2. Tétel. Tegyük fel, hogy a (3.4)—(3.28) sémában $n \geqq 2$, és teljesül a (3.10) és az $1+v_1+\ldots+v_{j-1}>0$ $(j=2, \ldots, n)$ feltétel. Az

$$E(\theta|y_1, \ldots, y_n) = \hat{E}(\theta|H_y)$$

reláció akkor és csak akkor áll fenn, ha $y(j)$ *Gauss-folyamat* és $\pi(\theta)$ normális eloszlás.

Megjegyzés. Megemlítjük, hogy feltevés szerint θ olyan valószínűségi változó, amely független $(\varepsilon_1, \ldots, \varepsilon_n)$-től. Ebből belátható, hogy a (3.4)—(3.5) és (3.4)—(3.28) sémák esetén a $(\theta, y_1, \ldots, y_n)$ valószínűségi vektorváltozó normalitása azzal ekvivalens, hogy θ normális eloszlású valószínűségi változó és $y(j)$ *Gauss-folyamat.*

2. A Bayes-féle becslés linearitásának egy feltétele

Tekintsük most a (3.4)—(3.5) sémát azzal a feltevéssel, hogy $\varepsilon(j)$, $j=1, \ldots, n$, független sorozat, azonos $F(x)$ eloszlással.

Ebben az esetben az

$$E\varepsilon(j) = \mu_1, \quad E\theta = \alpha_1,$$

$$E[\varepsilon(j)-\mu_1]^k = \mu_k, \quad E(\theta-\alpha_1)^k = \alpha_k, \quad k \geqq 2,$$

jelölésekkel belátható, hogy a

(3.39)
$$0 < \alpha_2 < \infty,$$

$$0 < \mu_2 < \infty$$

feltétel teljesülése esetén θ-nak (3.13) legjobb lineáris becslése a következő alakú:

(3.40)
$$\hat{\theta}_1 = \hat{E}(\theta|H_y) = c_0 + c\sum_{j=1}^{n} y(j) + c\sum_{j=1}^{n-1} \varrho_j \varepsilon(j),$$

ahol

$$c_0 = \frac{\alpha_1 - n\alpha_2 \dfrac{\mu_1}{\mu_2}}{1+n\dfrac{\alpha_2}{\mu_2}}, \quad c = \frac{\dfrac{\alpha_2}{\mu_2}}{1+n\dfrac{\alpha_2}{\mu_2}} = \frac{\alpha_2}{\mu_2+n\alpha_2},$$

és

$$\varrho_1 = -(a+\ldots+a^{n-1})$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$
$$\varrho_{n-1} = -a.$$

Legyen $l=\sum_{j=1}^{n} l_j y(j)$ egy általános alakú lineáris becslés.

Vizsgáljuk meg azt a problémát, hogy milyen esetben áll fenn a következő reláció:

$$(3.41)\quad E\left(\theta \,\middle|\, \sum_{j=1}^{n} y(j)+\sum_{j=1}^{n-1}\varrho_j\varepsilon(j),\ \sum_{j=1}^{n} l_j y(j)\right) = \hat{\theta}_1 = c_0 + c\left(\sum_{j=1}^{n} y(j)+\sum_{j=1}^{n-1}\varrho_j\varepsilon(j)\right).$$

Az általánosság korlátozása nélkül feltehetjük, hogy $\alpha_1=\mu_1=0$, ekkor $c_0=0$ és a (3.41) formula az

$$(3.42)\quad E\left(\theta \,\middle|\, \sum_{j=1}^{n} y(j)+\sum_{j=1}^{n-1}\varrho_j\varepsilon_j,\ \sum_{j=1}^{n} l_j y_j\right) = c\left(\sum_{j=1}^{n} y_j+\sum_{j=1}^{n-1}\varrho_j\varepsilon_j\right)$$

alakra hozható.

Mindenekelőtt bebizonyítunk egy lemmát, amelyre a következőkben szükségünk lesz. Ez a lemma RAO nevéhez fűződik és megtalálható [43]-ban.

Tekintsük a

$$(3.43)\quad d_1 g(\delta_1 t)+\ldots+d_n g(\delta_n t) = 0$$

funkcionál egyenletet, amely érvényes ha $|t|<\varepsilon$, $\varepsilon>0$.

Tegyük fel, hogy δ_j, $j=1, \ldots, n$, értékek közül pontosan egy olyan δ_k létezik, mondjuk δ_n, hogy

$$(3.44)\quad |\delta_n| > \max\{|\delta_1|, \ldots, |\delta_{n-1}|\}.$$

Legyen

$$\delta_j^* = \frac{\delta_j}{\delta_n}, \quad j = 1, \ldots, n-1.$$

(3.44) miatt $|\delta_j^*|<1$, $j=1, \ldots, n-1$.

A (3.43) egyenlet így a következő alakra hozható:

$$(3.43)\quad g(t) = -\{d_1^* g(\delta_1^* t)+\ldots+d_{n-1}^* g(\delta_{n-1}^* t)\}.$$

Igaz az alábbi

3.1. Lemma (Rao). Ha $g(t)$ kielégíti (3.43′)-t, deriváltja az origóban folytonos (vagy általánosabban, $g(t)$ felírható $th(t)$ alakban, ahol $h(t)$ folytonos az origóban), és $\sum_{j=1}^{n-1} d_j^*\delta_j^* = -1$, $d_j^*\cdot\delta_j^*<0$, $j=1, \ldots, n-1$, akkor $g(t)=qt$ (q állandó).

Bizonyítás. $g(t)$ helyett $th(t)$-t helyettesítve (3.43′)-be, majd a (3.43′) egyenlet mindkét oldalát $t\neq 0$-val osztva kapjuk, hogy

$$(3.45)\quad h(t) = -\{d_1^*\delta_1^* h(\delta_1^* t)+\ldots+d_{n-1}^*\delta_{n-1}^* h(\delta_{n-1}^* t)\} =$$
$$= p_1 h(\delta_1^* t)+\ldots+p_{n-1} h(\delta_{n-1}^* t),$$

ahol a feltevés miatt minden p_j pozitív, és $\sum p_j=1$. (3.45)-ből belátható, hogy

$$h(\delta_j^* t) = p_1 h(\delta_1^* \delta_j^* t)+\ldots+p_{n-1} h(\delta_{n-1}^* \delta_j^* t).$$

Tehát

$$h(t) = \sum_j p_j h(\delta_j^* t) = \sum_j \sum_k p_j p_k h(\delta_j^* \delta_k^* t) = \sum_j \sum_k q_{jk} h(\delta_j^* \delta_k^* t),$$

ahol

$$\sum_j \sum_k q_{jk}=1,$$

amiből

$$h(t)-h(0) = \sum_j \sum_k q_{jk}[h(\delta_j^* \delta_k^* t)-h(0)] \tag{3.46}$$

adódik.

Így folytatva kapjuk, hogy

$$h(t)-h(0) = \sum q_{j_1 \ldots j_m}[h(\delta_{j_1}^* \ldots \delta_{j_m}^* t)-h(0)], \tag{3.47}$$

ahol $\sum q_{j_1 \ldots j_m}=1$, az összegezés minden $1 \leqq j_i \leqq n-1$, $i=1, \ldots, m$-re értendő. Fix $t \neq 0$-ra és adott $\varepsilon>0$-ra válasszunk olyan nagy m-et, hogy $(\max_j |\delta_j^*|)^m < \frac{\varrho}{|t|}$, ahol ϱ olyan pozitív szám, amelyre $|h(x)-h(0)|<\varepsilon$, ha $|x|<\varrho$.

Ily módon (3.47) jobb oldalának abszolút értéke ε-nál kisebb, tehát $|h(t)-h(0)|<\varepsilon$ minden $\varepsilon>0$-ra, azaz $h(t)=h(0)=q$. Így $g(t)=qt$. Ezzel a lemma bizonyítását befejeztük.

Megjegyezzük, hogy ha a 3.1. lemma bizonyításában használjuk azt a feltevést, hogy $g(t)$ deriváltja folytonos az origóban, akkor (3.45)-ben $h(t)$ helyett $g'(t)$-t -runk, és ugyanígy kapjuk, hogy $g'(t)=g'(0)=q$, amiből $g(t)=qt$ adódik (feltételezve, hogy $g(0)=0$).

Ilyen esetben a $\sum_{j=1}^{n-1} d_j^* \delta_j^* = -1$ feltétel helyettesíthető a $g'(0)=q \neq 0$ feltétellel.

Most visszatérünk a vizsgált problémára. Tegyük fel, hogy (3.42) teljesül. Ebben az esetben

$$E\left\{\theta \exp i\left(t \sum_{j=1}^{n} y_j + t \sum_{j=1}^{n-1} \varrho_j \varepsilon_j + t \sum_{j=1}^{n} l_j y_j\right)\right. = \tag{3.48}$$

$$= E\left\{\exp i\left(t \sum_{j=1}^{n} y_j + t \sum_{j=1}^{n-1} \varrho_j \varepsilon_j + z \sum_{j=1}^{n} l_j y_j\right) E\left(\theta \,\middle|\, \sum_{j=1}^{n} y_j + \sum_{j=1}^{n-1} \varrho_j \varepsilon_j,\ \sum_{j=1}^{n} l_j y_j\right)\right\} =$$

$$= E\left\{c\left(\sum_{j=1}^{n} y_j + \sum_{j=1}^{n-1} \varrho_j \varepsilon_j\right) \exp i\left(t \sum_{j=1}^{n} y_j + t \sum_{j=1}^{n-1} \varrho_j \varepsilon_j + z \sum_{j=1}^{n} l_j y_j\right)\right\}.$$

Mivel

$$\sum_{j=1}^{n} y_j + \sum_{j=1}^{n-1} \varrho_j \varepsilon_j = n\theta + \sum_{j=1}^{n} \varepsilon_j,$$

$$\sum_{j=1}^{n} l_j y_j = \beta_1 \varepsilon_1 + \ldots + \beta_n \varepsilon_n + (l_1 + \ldots + l_n)\theta = \sum_{j=1}^{n} \beta_j \varepsilon_j + n\beta\theta,$$

ahol

$$\beta_1 = l_1 + al_2 + \ldots + a^{n-1} l_n$$
$$\beta_2 = l_2 + al_3 + \ldots + a^{n-2} l_n$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$
$$\beta_n = l_n$$

és

$$\beta = \frac{1}{n}(l_1 + \ldots + l_n),$$

(3.48)-ból kapjuk az

$$E\left\{\theta \exp i\left[t\left(n\theta + \sum_{j=1}^{n} \varepsilon_j\right) + z\left(n\beta\theta + \sum_{j=1}^{n} \beta_j \varepsilon_j\right)\right]\right\} = \tag{3.49}$$

$$= E\left\{c\left(n\theta + \sum_{j=1}^{n} \varepsilon_j\right) \exp i\left[t\left(n\theta + \sum_{j=1}^{n} \varepsilon_j\right) + z\left(n\beta\theta + \sum_{j=1}^{n} \beta_j \varepsilon_j\right)\right]\right\}$$

összefüggést.

Belátható, hogy egyrészt

$$E\left\{\theta \exp i\left[t\left(n\theta + \sum_{j=1}^{n} \varepsilon_j\right) + z\left(n\beta\theta + \sum_{j=1}^{n} \beta_j \varepsilon_j\right)\right]\right\} = E\{\theta \cdot e^{i(tn+zn\beta)\theta + i\sum_{j=1}^{n}(t+z\beta_j)\varepsilon_j}\} = \tag{3.50}$$

$$= E\{\theta \cdot e^{i(tn+zn\beta)\theta}\} \cdot E\{e^{i\sum_{j=1}^{n}(t+z\beta_j)\varepsilon_j}\} = -i\varphi'(nt+n\beta z) \prod_{j=1}^{n} f(t+z\beta_j),$$

ahol $\varphi(t) = \int_{-\infty}^{+\infty} e^{it\theta} d\pi(\theta)$, $f(t) = \int_{-\infty}^{+\infty} e^{itx} dF(x)$,

másrészt,

$$E\left\{c\left(n\theta + \sum_{j=1}^{n} \varepsilon_j\right) e^{i(nt+n\beta z)\theta + i\sum_{j=1}^{n}(t+z\beta_j)\varepsilon_j}\right\} = \tag{3.51}$$

$$= E\{cn\theta e^{i(nt+n\beta z)\theta}\} \cdot E\{e^{i\sum_{j=1}^{n}(t+z\beta_j)\varepsilon_j}\} + E\left\{c\sum_{j=1}^{n} \varepsilon_j e^{i\sum_{j=1}^{n}(t+z\beta_j)\varepsilon_j}\right\} E\{e^{i(nt+n\beta z)\theta}\} =$$

$$= -icn\varphi'(nt+n\beta z) \prod_{j=1}^{n} f(t+z\beta_j) - ic\varphi(nt+n\beta z) \sum_{j=1}^{n} f'(t+\beta_j z) \prod_{k\neq j} f(t+\beta_k z).$$

Ily módon (3.49), (3.50) és (3.51)-ből következik

$$-i\varphi'(nt+n\beta z) \prod_{j=1}^{n} f(t+z\beta_j) = -icn\varphi'(nt+n\beta z) \prod_{j=1}^{n} f(t+z\beta_j) -$$

$$- ic\varphi(nt+n\beta z) \sum_{j=1}^{n} f'(t+\beta_j z) \prod_{k\neq j} f(t+\beta_k z),$$

ahonnan

$$\frac{\varphi'(nt+n\beta z)}{\varphi(nt+n\beta z)}(1-cn) = c\sum_{j=1}^{n} \frac{f'(t+\beta_j z)}{f(t+\beta_j z)} \tag{3.52}$$

adódik, ha $|t|$ és $|z|$ eléggé kicsiny.

$$\psi(t) = \frac{\varphi'(t)}{\varphi(t)},\ g(t) = \frac{f'(t)}{f(t)} \quad \text{jelölésekkel,} \quad \frac{c}{1-nc} = \frac{\alpha_2}{\mu_2} > 0$$

miatt, (3.52)-ből kapjuk a

$$\psi(nt+n\beta z) = \frac{c}{1-nc}\sum_{j=1}^{n} g(t+\beta_j z) \tag{3.53}$$

egyenletet.

Legyen $t = -\beta z$, így (3.53)-ból adódik

$$\sum_{j=1}^{n} g(\delta_j z) = 0, \quad \text{ha} \quad |z| < \varepsilon,\ \varepsilon > 0 \tag{3.54}$$

eléggé kicsiny, ahol $\delta_j = \beta_j - \beta$.

Tegyük fel, hogy $\delta_j, j=1, \ldots, n$ közül létezik pontosan egy δ_k abszolút értékben legnagyobb, mondjuk δ_n, azaz

$$|\delta_n| > \max\{|\delta_1|, \ldots, |\delta_{n-1}|\}, \tag{3.55}$$

továbbá

(3.56) $g(t)$ deriváltja az origóban folytonos (vagy ami ezzel ekvivalens, $f''(t)$ az origóban folytonos),

$$\frac{\delta_j}{\delta_n} < 0, \quad j = 1, \ldots, n-1 \tag{3.57}$$

(figyelembe véve, hogy $g'(0) = -\mu_2 \neq 0$.)

A 3.1. lemmát alkalmazva kapjuk, hogy

$$g(t) = qt,$$

(q állandó) az origó valamely környezetében, amiből könnyű belátni (lásd 1.3. tétel szükségességének bizonyítását), hogy $f(t)$ normális eloszlás karakterisztikus függvénye.

Ilyen esetben (3.53)-ból következik, hogy

$$\psi(t) = pt$$

(p állandó) ha $|t|$ elég kicsiny, ami azt jelenti, hogy $\pi(\theta)$ normális eloszlás.

Megfordítva, $y(j)$ *Gauss-folyamat* és $\pi(\theta)$ normális eloszlás esetén megmutattuk, hogy

$$E(\theta|y_1, \ldots, y_n) = c\left(\sum_{j=1}^{n} y(j) + \sum_{j=1}^{n-1} \varrho_j \varepsilon(j)\right)$$

és még inkább

$$E\left(\theta \middle| \sum_{j=1}^{n} y(j) + \sum_{j=1}^{n-1} \varrho_j \varepsilon(j),\ \sum_{j=1}^{n} l_j y(j)\right) = c\left(\sum_{j=1}^{n} y(j) + \sum_{j=1}^{n-1} \varrho_j \varepsilon(j)\right).$$

Tehát igazoltuk a következő állítást:

3.3. Tétel. Tegyük fel, hogy $n \geqq 2$ és teljesülnek a (3.39), (3.55), (3.56) és (3.57)

feltételek. Az

$$E\left(\theta \middle| \sum_{j=1}^{n} y(j)+\sum_{j=1}^{n-1} \varrho_j \varepsilon(j),\ \sum_{j=1}^{n} l_j y(j)\right) = c_0 + c\left(\sum_{j=1}^{n} y(j)+\sum_{j=1}^{n-1} \varrho_j \varepsilon(j)\right)$$

összefüggés akkor és csak akkor áll fenn, ha $y(j)$ *Gauss-folyamat* és $\pi(\theta)$ normális eloszlás.

3. Bayes-féle polinomiális becslések linearitásának feltétele

Tekintsük a (3.4)—(3.5) sémát.

Tegyük fel, hogy van olyan k, amelyre teljesül

$$\alpha_{2k} < \infty, \quad \mu_{2k}^{(j)} < \infty, \quad j = 1, \ldots, n. \tag{3.58}$$

Jelöljük H_y^k-val az $y(1), \ldots, y(n)$ valószínűségi változók k-nál nem magasabb fokú polinomjainak zárt lineáris sokaságát. Ebben az esetben θ-nak az $y(1), \ldots, y(n)$ mintán alapuló becslésére használhatjuk a

$$\hat{\theta}_k = \hat{E}(\theta|H_y^k) \tag{3.59}$$

statisztikát, ahol $\hat{E}(\theta|H_y^k)$ jelöli θ-nak H_y^k-ra való vetületét. A (3.59) becslés nem más, mint θ-nak négyzetes középben legjobb k-adfokú polinomiális becslése, és az ilyen becslés csak az $\alpha_1, \alpha_{2k}; \mu_1^{(j)}, \ldots, \mu_{2k}^{(j)}$ $(j=1, \ldots, n)$ momentumoktól függ.

Mivel $H_y^{k-1} \subset H_y^k$, könnyen belátható, hogy

$$\hat{\theta}_{k-1} = \hat{E}(\theta|H_y^{k-1}) = \hat{E}(\hat{E}(\theta|H_y^k)|H_y^{k-1}) = \hat{E}(\hat{\theta}_k|H_y^{k-1}), \tag{3.60}$$

ahonnan

$$E(\hat{\theta}_{k-1}-\theta)^2 = E(\hat{\theta}_k-\theta)^2 + E(\hat{\theta}_k-\hat{\theta}_{k-1})^2 \tag{3.61}$$

adódik.

(3.61)-ből következik, hogy

$$E(\hat{\theta}_{k-1}-\theta)^2 \geqq E(\hat{\theta}_k-\theta)^2 \tag{3.62}$$

és egyenlőség akkor és csak akkor áll fenn, ha $\hat{\theta}_k = \hat{\theta}_{k-1}$. Felmerül a kérdés, hogy milyen esetben teljesül a következő reláció

$$\hat{\theta}_k = \hat{E}(\theta|H_y^k) = \hat{E}(\theta|H_y^1) = \hat{\theta}_1 \quad (H_y^1 = H_y). \tag{3.63}$$

Az előzőekhez hasonlóan, az $x(j)=\varepsilon(j)+\theta$ $(j=1, \ldots, n)$ új változók bevezetésével azt kapjuk, hogy

$$\begin{aligned} y(1) &= x(1) \\ y(2) &= x(2)+a\varepsilon(1) \\ &\cdots\cdots\cdots\cdots \\ y(n) &= x(n)+a^{n-1}\varepsilon(1)+\ldots+a\varepsilon(n-1). \end{aligned} \tag{3.64}$$

Az $x(1), \ldots, x(n)$ valószínűségi változók k-nál nem magasabb fokú polinomjainak zárt lineáris sokaságát H_x^k-val jelölve, (3.64)-ből belátható, hogy

$$H_y^k = H_x^k, \quad k = 1, 2, \ldots,$$

amiből

$$\hat{E}(\theta|H_y^k) = \hat{E}(\theta|H_x^k) = \hat{\theta}_k, \quad k = 1, 2, \ldots, \tag{3.65}$$

következik.

Ily módon a probléma arra vezethető vissza, hogy az $x(j)=\varepsilon(j)+\theta$, $j=1, \ldots, n$, sémára milyen esetben teljesül az

$$\hat{E}(\theta|H_x^k) = \hat{E}(\theta|H_x^1) \quad (H_x^1 = H_x)$$

reláció.

Erre válaszol a KAGAN—KARPOV [22]-cikkben szereplő következő tétel, amelyet a teljesség kedvéért be is bizonyítunk.

3.4. Tétel. (Kagan—Karpov). A (3.58) feltételt kielégítő $x(j)=\varepsilon(j)+\theta$ séma esetén valamilyen $n\geqq 2$-re az

$$\hat{E}(\theta|H_x^k) = \hat{E}(\theta|H_x^1) \tag{3.66}$$

összefüggés akkor és csak akkor áll fenn, ha az $\alpha_1, \ldots, \alpha_{k+1}; \mu_1^{(j)}, \ldots, \mu_{k+1}^{(j)}$ momentumok megegyeznek valamely normális eloszlás megfelelő momentumaival, azaz

$$\alpha_m = \begin{cases} 0, & \text{ha } m \text{ páratlan } 1 \leqq m \leqq k+1, \\ (m-1)!!\sigma^m, & \text{ha } m \text{ páros } \quad 1 \leqq m \leqq k+1, \end{cases} \tag{3.67}$$

$$\mu_m^{(j)} = \begin{cases} 0, & \text{ha } m \text{ páratlan } 1 \leqq m \leqq k+1, \\ (m-1)!!\sigma_j^m, & \text{ha } m \text{ páros } \quad 1 \leqq m \leqq k+1, \end{cases} \tag{3.68}$$

ahol $\sigma^2>0$, $\sigma_j^2>0$.

Bizonyítás. 1°. *Elégségesség.* Az általánosság korlátozása nélkül feltehetjük, hogy $\alpha_1=\mu_1^{(j)}=0$. A (3.66) reláció ekkor a következő alakú:

$$\hat{E}(\theta|H_x^k) = \sum_{j=1}^{n} c_j x_j, \tag{3.69}$$

ahol c_j meghatározható (3.12) segítségével.

Vegyük észre, hogy (3.69) ekvivalens azzal, hogy

$$E\left\{\left(\theta - \sum_{j=1}^{n} c_j x_j\right) x_1^{m_1} \ldots x_n^{m_n}\right\} = 0, \quad 0 \leqq m_1+\ldots+m_n \leqq k. \tag{3.70}$$

Tehát (3.69) teljesülése csak az $\alpha_1, \ldots, \alpha_{k+1}; \mu_1^{(j)}, \ldots, \mu_{k+1}^{(j)}$ momentumoktól függ. Ismeretes, hogy a $(\theta, \varepsilon_1, \ldots, \varepsilon_n)$ normális eloszlású valószínűségi vektor változó esetén fennáll a (3.66)-nál erősebb

$$E(\theta|x_1. \ldots, x_n) = \hat{E}(\theta|H_x^1)$$

reláció is. Ebből már könnyen belátható, hogy ha az $\alpha_1, \ldots, \alpha_{k+1}; \mu_1^{(j)}, \ldots, \mu_{k+1}^{(j)}$ momentumok a (3.67)—(3.68) alakúak, azaz ezek megegyeznek valamely normális eloszlás megfelelő momentumaival, akkor (3.66) teljesül.

2°. *Szükségesség.* Először megmutatjuk, hogy (3.70) teljesüléséből adódik, hogy $\alpha_3, \ldots, \alpha_{k+1}; \mu_3^{(j)}, \ldots, \mu_{k+1}^{(j)}$ egyértelműen kifejezhetők $\alpha_1, \alpha_2; \mu_1^{(1)}, \mu_2^{(1)}, \ldots, \mu_1^{(n)}, \mu_2^{(n)}$ segítségével.

Valóban, tegyük fel, hogy $\alpha_3, \ldots, \alpha_m; \mu_3^{(j)}, \ldots, \mu_m^{(j)}$, $m\leqq k$, már kifejezhetők $\alpha_1, \alpha_2; \mu_1^{(1)}, \mu_2^{(1)}, \ldots, \mu_1^{(n)}, \mu_2^{(n)}$ segítségével.

Figyelembe véve, hogy $x(j)=\varepsilon(j)+\theta$, $j=1, \ldots, n$, az

$$E\left\{\left(\theta-\sum_{j=1}^{n} c_j x_j\right) x_1 x_2^{m-1}\right\}=0 \tag{3.71}$$

reláció felhasználásával ki tudjuk fejezni α_{m+1}-et az előbbi momentumok segítségével a következő alakban:

$$\alpha_{m+1}=\alpha_{m+1}(\alpha_1, \alpha_2; \mu_1^{(1)}, \mu_2^{(1)}, \ldots, \mu_1^{(n)}, \mu_2^{(n)}). \tag{3.72}$$

Továbbá, használjuk fel az

$$E\left\{\left(\theta-\sum_{j=1}^{n} c_j x_j\right) x_j^m\right\}=0$$

formulát, és vegyük figyelembe, hogy α_{m+1} (3.72) alakú, a

$$\mu_{m+1}^{(j)}=\mu_{m+1}^{(j)}(\alpha_1, \alpha_2; \mu_1^{(j)}, \mu_2^{(j)}, \ldots, \mu_1^{(n)}, \mu_2^{(n)}) \tag{3.73}$$

kifejezést kapjuk.

Láttuk, hogy a (3.67)—(3.68) alakú momentumok esetén érvényes a (3.46) reláció. Így az $\alpha_3, \ldots, \alpha_{k+1}; \mu_3^{(j)}, >, \mu_{k+1}^{(j)}$ momentumok $\alpha_1, \alpha_2; \mu_1^{(1)}, \mu_2^{(1)}, \ldots, \mu_1^{(n)}, \mu_2^{(n)}$ segitségével való egyértelmü kifejezhetőségből következik a kívánt állítás.

Ezzel a 3.4. tétel bizonyítását befejeztük.

A korábbi tárgyalásból és a 3.4. tételből kapjuk a következő eredményt a (3.4)—(3.5) séma esetére:

3.5. Tétel. Tegyük fel, hogy a (3.4)—(3.5) séma esetén teljesül a (3.58) feltétel és $n \geqq 2$.

Annak szükséges és elegendő feltétele, hogy

$$\hat{E}(\theta|H_y^k)=\hat{E}(\theta|H_y^1), \quad k=1, 2, \ldots$$

legyen az, hogy az $\alpha_1, \ldots, \alpha_{k+1}; \mu_k^{(j)}, \ldots, \mu_{k+1}^{(j)}$ momentumokra teljesüljenek a (3.67)—(3.68) relációk, azaz $\alpha_1, \ldots, \alpha_{k+1}; \mu_1^{(j)}, \ldots, \mu_{k+1}^{(j)}$ megegyezzenek valamely normális eloszlás megfelelő momentumaival.

1. Megjegyzés. Könnyű belátni, hogy a fenti állítás érvényben marad abban az esetben is, amikor a $y(j)=\xi(j)+\theta$ sémában $\xi(j)$ a következő p-edrendű autoregressziós folyamat (lásd (2.17) és (3.29)):

$$\begin{aligned}
\xi(1) &= \varepsilon(1) \\
&\cdots\cdots\cdots\cdots\cdots\cdots \\
\xi(p)+a_1\xi(p-1)+\ldots+a_{p-1}\xi(1) &= \varepsilon(p) \\
\xi(k)+a_1\xi(k-1)+\ldots+a_p\xi(1) &= \varepsilon(k) \quad k \geqq p+1,
\end{aligned}$$

ahol $\varepsilon(j), j=1, \ldots, n$, független sorozat, $F_j(x)$ eloszlással.

2. Megjegyzés. A 3.1. és 3.5. tételekből belátható, hogy ha a (3.4)—(3.5) sémában $y(j)$ *Gauss-folyamat* és $\pi(\theta)$ normális eloszlás, akkor θ legjobb becslése *(Bayes-féle becslése)* megegyezik θ legjobb k-adfokú ($k=1, 2, \ldots$) polinomiális becslésével, azaz teljesül

$$\hat{\theta}=E(\theta|y_1, \ldots, y_n)=\hat{\theta}_k=\hat{\theta}_1, \quad k=2, 3, \ldots .$$

IV. FEJEZET

1. Stacionárius folyamatok szórásnégyzetének szabályos becslései. Pitman-féle becslés

Legyen $\xi(t)$ valós $(0 \leqq t \leqq T_0)$ stacionárius folyamat, $E\xi(t)=0$; $D^2\xi(t)= E\xi^2(t)=\sigma^2=s$ ismeretlen. Adott s esetén a folyamathoz tartozó vlalószínűségi mértéket, ill. várható értéket jelölje P_s, ill. E_s.

Sztochasztikus folyamatok esetén az $E\xi(n)=\theta$ paraméter korrekt becslésével foglalkozik ARATÓ [2] dolgozata. Független megfigyeléssorozat esetén a skála paraméter szabályos becslésének definícióját KAGAN és RUHIN adta meg [17], [23]. A továbbiakban megvizsgáljuk e fogalom hasznosságát sztochasztikus folyamatokra.

Szabályos becslések.

4.1. Definíció. Az $\hat{s}(\xi(t))$ funkcionált szabályos becslésnek nevezzük, ha tetszőleges $-\infty<\lambda<+\infty$ esetén

$$\hat{s}(\lambda\xi(t)) = \lambda^2\hat{s}(\xi(t)). \tag{4.1}$$

Könnyen belátható, hogy pl. az $\hat{s}=\frac{1}{T_0}\int_0^{T_0}\xi^2(t)\,dt$; $\hat{s}=\frac{1}{n}\sum_{i=1}^{n}\xi^2(t_i)$, $0\leqq t_1<t_2<\ldots$ $<t_n\leqq T_0$; $\hat{s}=\xi^2(0)$ funkcionálok szabályos becslések.

Jelölje $\mathscr{S}$ a szabályos becslések osztályát. Ha $\hat{s}\in\mathscr{S}$, akkor

$$E_s(\hat{s}-s)^2 = s^2E_1(\hat{s}-1)^2. \tag{4.2}$$

Nyilvánvaló ugyanis, hogy

$$E_s(\hat{s}-s)^2 = E_s(\hat{s}(\xi(t))-s)^2 = s^2E_s\left[\hat{s}\left(\frac{\xi(t)}{\sqrt{s}}\right)-1\right]^2 = s^2E_1[\hat{s}(\xi(t))-1]^2.$$

Jelöljük $\mathscr{B}_0^{T_0}$-val a $\frac{\xi(t)}{\xi(0)}$ $(0\leqq t\leqq T_0)$ változók által generált σ-algebrát. (Tegyük fel, hogy $\xi(0)\neq 0 \pmod{P_1}$).

Legyen $\hat{s}\in\mathscr{S}$ és

$$\hat{u}(\xi(t)) = \hat{s}(\xi(t))\frac{E_1(\hat{s}(\xi(t))|\mathscr{B}_0^{T_0})}{E_1(\hat{s}^2(\xi(t))|\mathscr{B}_0^{T_0})}. \tag{4.3}$$

Nyilván $\hat{u}\in\mathscr{S}$ és megmutatjuk, hogy igaz a következő

4.1. Lemma. Tetszőleges $\hat{s}_1, \hat{s}_2\in\mathscr{S}$-re

$$\hat{s}_1\frac{E_1(\hat{s}_1|\mathscr{B}_0^{T_0})}{E_1(\hat{s}|\mathscr{B}_0^{T_0})} = \hat{s}_2\frac{E_1(\hat{s}_2|\mathscr{B}_0^{T_0})}{E_1(\hat{s}_2^2|\mathscr{B}_0^{T_0})} \pmod{P_1}, \tag{4.4}$$

azaz tetszőleges $\hat{s}\in\mathscr{S}$-re az $\hat{u}=\hat{s}\frac{E_1(\hat{s}|\mathscr{B}_0^{T_0})}{E_1(\hat{s}^2|\mathscr{B}_0^{T_0})}$ becslés ugyanaz.

Bizonyítás. Valóban,

$$\frac{\hat{s}_1(\xi(t))}{\hat{s}_2(\xi(t))} = \frac{\hat{s}_1\left(\frac{\xi(t)}{\xi(0)}\right)}{\hat{s}_2\left(\frac{\xi(t)}{\xi(0)}\right)} = f\left(\frac{\xi(t)}{\xi(0)}\right)$$

és így $\frac{\hat{s}_1}{\hat{s}_2}$ a $\frac{\xi(t)}{\xi(0)}$ funkcionálja, melyet f-fel jelölünk. Másrészt $\frac{\hat{s}_1}{\hat{s}_2}$ $\mathscr{B}_0^{T0}$-mérhetősége miatt a feltételes várható érték jól ismert tulajdonsága alapján

$$E_1(\hat{s}_1|\mathscr{B}_0^{T0}) = E_1\left(\hat{s}_1\frac{\hat{s}_2}{\hat{s}_2}\middle|\mathscr{B}_0^{T0}\right) = \frac{\hat{s}_1}{\hat{s}_2}E_1(\hat{s}_2|\mathscr{B}_0^{T0}),$$

$$E_1(\hat{s}_1^2|\mathscr{B}_0^{T0}) = E_1\left(\hat{s}_1^2\frac{\hat{s}_2^2}{\hat{s}_1^2}\middle|\mathscr{B}_0^{T0}\right) = \frac{\hat{s}_1^2}{\hat{s}_2^2}E_1(\hat{s}_2^2|\mathscr{B}_0^{T0}),$$

ahonnan következik, hogy

$$\hat{s}_1\frac{E_1(\hat{s}_1|\mathscr{B}_0^{T0})}{E_1(\hat{s}_1^2|\mathscr{B}_0^{T0})} = \hat{s}_1\frac{\frac{\hat{s}_1}{\hat{s}_2}E_1(\hat{s}_2|\mathscr{B}_0^{T0})}{\frac{\hat{s}_1^2}{\hat{s}_2^2}E_1(\hat{s}_2^2|\mathscr{B}_0^{T0})} = \hat{s}_2\frac{E_1(\hat{s}_2|\mathscr{B}_0^{T0})}{E_1(\hat{s}_2^2|\mathscr{B}_0^{T0})} \pmod{P_1}$$

és ezzel a 4.1. lemmát bebizonyítottuk.

4.2. Lemma. A szabályos becslések osztályában az

$$\hat{u} = \hat{s}\frac{E_1(\hat{s}|\mathscr{B}_0^{T0})}{E_1(\hat{s}^2|\mathscr{B}_0^{T0})}$$

becslés, ahol $\hat{s}\in\mathscr{S}$ minimális négyzetes eltérésű becslése σ^2-nek.

Bizonyítás. Legyen $h\left(\frac{\xi(t)}{\xi(0)}\right) = \frac{E_1(\hat{s}|\mathscr{B}_0^{T0})}{E_1(\hat{s}^2|\mathscr{B}_0^{T0})}$, akkor $\hat{u}=h\hat{s}$, továbbá

$$E_s(\hat{s}-s)^2 = E_s(\hat{s}-\hat{u}+\hat{u}-s)^2 = E_s(\hat{s}-\hat{u})^2+E_s(\hat{u}-s)^2+2E_s(\hat{s}-\hat{u})(\hat{u}-s).$$

Vegyük észre, hogy

$$\begin{aligned}
E_s(\hat{s}-\hat{u})(\hat{u}-s) &= E_s\{\hat{s}(\xi(t))-h\hat{s}(\xi(t))\}\{\hat{u}(\xi(t))-s\} = \\
&= s^2E_s\left\{\hat{s}\left(\frac{\xi(t)}{\sqrt{s}}\right)-h\hat{s}\left(\frac{\xi(t)}{\sqrt{s}}\right)\right\}\left\{\hat{u}\left(\frac{\xi(t)}{\sqrt{s}}\right)-1\right\} = \\
&= s^2E_1(\hat{s}-h\hat{s})(\hat{u}-1) = \\
&= s^2E_1\{\hat{s}(1-h)(\hat{u}-1)\} = \\
&= s^2E_1\{(1-h)E_1[\hat{s}(\hat{u}-1)|\mathscr{B}_0^{T0}]\} = \\
&= s^2E_1\{(1-h)E_1(\hat{s}\hat{u}-\hat{s}|\mathscr{B}_0^{T0})\} = \\
&= s^2E_1\{(1-h)E_1(h\hat{s}^2-\hat{s}|\mathscr{B}_0^{T0})\} = \\
&= s^2E_1\{(1-h)[hE_1(\hat{s}^2|\mathscr{B}_0^{T0})-E_1(\hat{s}|\mathscr{B}_0^{T0})\} = 0,
\end{aligned}$$

amiből

$$E_s(\hat{s}-s)^2 = E_s(\hat{s}-\hat{u})^2+E_s(\hat{u}-s)^2 \geqq E_s(\hat{u}-s)^2 \tag{4.5}$$

adódik. A 4.1. lemma alapján (4.5)-ből következik a 4.2. lemma helyessége.

4.3. Lemma. Legyen $\hat{s}\in\mathcal{S}$ és $\hat{s}$ torzítatlan becslése s-nek. Legyen továbbá

$$\hat{u}^* = c\hat{s}\frac{E_1(\hat{s}|\mathcal{B}_0^{T0})}{E_1(\hat{s}^2|\mathcal{B}_0^{T0})} = c\hat{u} = ch\hat{s}, \tag{4.6}$$

ahol a c állandó úgy határozható meg, hogy

$$cE_1\left\{\frac{E_1^2(\hat{s}|\mathcal{B}_0^{T0})}{E_1(\hat{s}^2|\mathcal{B}_0^{T0})}\right\} = 1.$$

Akkor $\hat{u}^*\in\mathcal{S}$ és $\hat{u}^*$ torzítatlan becslés s-re.

Bizonyítás. Definíció szerint $\hat{u}^*\in\mathcal{S}$, továbbá

$$E_s(\hat{u}^*) = cE_s(h\hat{s}) = csE_1(h\hat{s}) =$$

$$= csE_1\{hE_1(\hat{s}|\mathcal{B}_0^{T0})\} = csE_1\left\{\frac{E_1^2(\hat{s}|\mathcal{B}_0^{T0})}{E_1(\hat{s}^2|\mathcal{B}_0^{T0})}\right\} = s.$$

Megjegyzés. A 4.3. lemma bizonyításából azt is kapjuk, hogy

$$cE_1(\hat{u}) = 1. \tag{4.7}$$

Pitman-féle becslés

Legyen $\hat{s}_1, \hat{s}_2\in\mathcal{S}$ és $\hat{u}_1^*=c_1h_1\hat{s}_1=c_1\hat{u}_1$, $\hat{u}_2^*=c_2h_2\hat{s}_2=c_2\hat{u}_2$. A 4.1. lemma szerint

$$\hat{u}_1 = \hat{s}_1\frac{E_1(\hat{s}_1|\mathcal{B}_0^{T0})}{E_1(\hat{s}_2^2|\mathcal{B}_0^{T0})} = \hat{s}_2\frac{E_1(\hat{s}_2|\mathcal{B}_0^{T0})}{E_1(\hat{s}_2^2|\mathcal{B}_0^{T0})} = \hat{u}_2 \pmod{P_1}.$$

Másrészt (4.7)-ből $c_1E_1(\hat{u}_1)=1$, $c_2E_1(\hat{u}_2)=1$, amikből következik, hogy $c_1=c_2$ és

$$\hat{u}_1^* = \hat{u}_2^*, \tag{4.8}$$

tehát minden $\hat{s}\in\mathcal{S}$-re $\hat{u}^*$ ugyanaz.

A 4.2. lemma bizonyításához hasonló módon belátható, hogy minden $\hat{s}\in\mathcal{S}$-re

$$E_s(\hat{s}-s)^2 \geqq E_s(\hat{u}^*-s)^2. \tag{4.9}$$

A 4.3. lemmából, (4.8)-ból és (4.9)-ből adódik a következő állítás:

4.1. Tétel. Ha $\hat{s}\in\mathcal{S}$ és $\hat{s}$ torzítatlan becslése s-nek, akkor $\hat{u}^*$ legjobb torzítatlan becslése s-nek a szabályos becslések osztályában.

4.2. Definíció. Pitman-féle becslésnek nevezzük azt az $\hat{u}^*$ becslést, amely legkisebb szórású, torzítatlan becslése $s=\sigma^2$-nek a szabályos becslések $\mathcal{S}$-osztályában. Tekintsünk egy speciális esetet. Legyen $\hat{s}=\xi^2(0)$. Nyilvánvaló, hogy $\hat{s}\in\mathcal{S}$ és $E_s(\xi^2(0))=\sigma^2=s$, azaz $\xi^2(0)$ torzítatlan becslése s-nek.

A 4.1. tétel szerint az $\hat{u}^*=c\xi^2(0)\frac{E_1(\xi^2(0)|\mathcal{B}_0^{T0})}{E_1(\xi^4(0)|\mathcal{B}_0^{T0})}$ becslés, ahol c meghatá-

rozható a $c \cdot E_1\left\{\frac{E_1^2(\xi^2(0)|\mathscr{B}_0^{T_0})}{E_1(\xi^4(0)|\mathscr{B}_0^{T_0})}\right\}=1$ összefüggésből, legjobb torzítatlan becslése s-nek az $\mathscr{S}$ osztályban. Tehát

$$\hat{u}^* = c\xi^2(0)\frac{E_1(\xi^2(0)|\mathscr{B}_0^{T_0})}{E_1(\xi^4(0)|\mathscr{B}_0^{T_0})}$$

Pitman-féle becslés.

Tegyük most fel, hogy $\xi_1, \ldots, \xi_n$ azonos eloszlásúak, de nem szükségképpen függetlenek, és $\xi_1=\xi(0)\neq 0 \pmod{P_1}$.

Legyen $p_0(x_1, \ldots, x_n)$ a $\xi_1, \ldots, \xi_n$ változók együttes sűrűségfüggvénye, legyen továbbá $\eta=\xi_1, \eta_2=\frac{\xi_2}{\xi_1}, \ldots, \eta_n=\frac{\xi_n}{\xi_1}$, akkor az $\eta, \eta_2, \ldots, \eta_n$ együttes sűrűségfüggvénye a következő:

$$p(z, y_2, \ldots, y_n) = p_0(z, zy_2, \ldots, zy_n)z^{n-1}.$$

Innen belátható, hogy

$$E_1\left(\xi_1^2\middle|\frac{\xi_2}{\xi_1}, \ldots, \frac{\xi_n}{\xi_1}\right) = E_1(\eta^2|\eta_2, \ldots, \eta_n) = \frac{\int z^{n+1}p_0(z, z\eta_2, \ldots, z\eta_n)\,dz}{\int z^{n-1}p_0(z, z\eta_2, \ldots, z\eta_n)\,dz} \tag{4.10}$$

és

$$E_1\left(\xi_1^4\middle|\frac{\xi_2}{\xi_1}, \ldots, \frac{\xi_n}{\xi_1}\right) = E_1(\eta^4|\eta_2, \ldots, \eta_n) = \frac{\int z^{n+3}p_0(z, z\eta_2, \ldots, z\eta_n)\,dz}{\int z^{n-1}p_0(z, z\eta_2, \ldots, z\eta_n)\,dz}. \tag{4.11}$$

Jelölje $\mathscr{B}_2^n$ a $\frac{\xi_k}{\xi_1}$ $(2\leqq k\leqq n)$ változók által generált σ-algebrát. Mivel az $\hat{s}=\xi^2(0)=\xi_1^2$ statisztika szabályos, torzítatlan becslése s-nek, a *Pitman-féle becslés* a következő:

$$\hat{u}^* = c\xi^2(0)\frac{E_1(\xi^2(0)|\mathscr{B}_2^n)}{E_1(\xi^4(0)|\mathscr{B}_2^n)}. \tag{4.12}$$

(4.10), (4.11), (4.12) alapján $z=t\xi_1$ helyettesítéssel adódik, hogy

$$\hat{u}^* = c\xi_1^2\frac{\int t^{n+1}\xi_1^{n+1}p_0(t\xi_1, t\xi_2, \ldots, t\xi_n)\,dt}{\int t^{n+3}\xi_1^{n+3}p_0(t\xi_1, t\xi_2, \ldots, t\xi_n)\,dt} = \tag{4.13}$$

$$= c\frac{\int t^{n+1}p_0(t\xi_1, t\xi_2, \ldots, t\xi_n)\,dt}{\int t^{n+3}p_0(t\xi_1, t\xi_2, \ldots, t\xi_n)\,dt},$$

ahol

$$cE_1\frac{E_1^2(\xi^2(0)|\mathscr{B}_2^n)}{E_1(\xi^4(0)|\mathscr{B}_2^n)} = 1.$$

2. Autoregressziós folyamatok szórásnégyzetének Pitman-féle becslése

Vegyük észre, hogy a fenti eredmények érvényesek maradnak abban az esetben is, amikor $\xi(t)$ stacionrius, $E\xi(t)=0$, $E\xi^2(t)=\beta\sigma^2=\beta s$, ahol β egy ismert állandó, és $s=\sigma^2$ ismeretlen paraméter (jelentését alább megadjuk).

Tekintsük most a

$$\xi(t)+a_1\xi(t-1)+\ldots+a_p\xi(t-p)=\varepsilon(t) \tag{4.14}$$

sztochasztikus differenciaegyenletnek eleget tevő p-edrendű autoregressziós folyamatot, ahol $\varepsilon(t)$, $t=0, \pm1, \pm2, \ldots$ független azonos eloszlású *Gauss-sorozat*, $E\varepsilon(t)=0$, $E\varepsilon^2(t)=\sigma^2$ (lásd (2.3)). Ismeretes (lásd [8]), hogy $\xi(t)$ felírható $\varepsilon(t)$ segítségével a

$$\xi(t)=\sum_{k=0}^{\infty}\beta_k\varepsilon(t-k) \tag{4.15}$$

alakban, ahol a β_k-együtthatók meghatározhatók $a_1, \ldots, a_p$ segítségével a következőképpen. Legyen

$$A(z)=\sum_{k=0}^{p}a_k z^k, \quad (a_0=1)$$

és tegyük fel, hogy $z_1, \ldots, z_p$ az $A(z)$ polinom p különböző gyöke, továbbá

$$\frac{a_0\beta_0}{A(z)}=\frac{\omega_1}{z_1-z}+\ldots+\frac{\omega_p}{z_p-z} \quad (a_0=\beta_0=1),$$

akkor

$$\beta_k=\omega_1 z_1^{-k-1}+\ldots+\omega_p z_1^{-k-1}. \tag{4.16}$$

Tegyük fel, hogy $|z_j|>1$, $j=1, \ldots, p$, amiből következik, hogy $\beta=\sum_{k=0}^{\infty}\beta_k^2<\infty$ és a $\xi(t)$ folyamat stacionárius, $E\xi(t)=0$, $E\xi^2(t)=\sigma^2\sum_{k=0}^{\infty}\beta_k^2$. A következőkben a $\sigma^2=s$ paraméter becslésével kapcsolatos problémát vizsgáljuk.

σ^2 maximum likelihood becslése.

Ismeretes (lásd [1]), hogy $\xi(1), \ldots, \xi(n)$ együttes sűrűségfüggvénye a következő

$$p_{\xi(1),\ldots,\xi(n)}(x_1, \ldots, x_n; \sigma)=p(\mathbf{X}; \sigma)= \tag{4.17}$$

$$=(2\pi)^{-n/2}|\mathbf{Q}_p|^{-1/2}\sigma^{-n}\exp\left\{-\frac{1}{2\sigma^2}\left[\mathbf{X}_p^T\mathbf{Q}_p^{-1}\mathbf{X}_p+\sum_{i=p+1}^{n}(x_i+a_1x_{i-1}+\ldots+a_px_{i-p})^2\right]\right\},$$

ahol

$$\mathbf{Q}_p^{-1}=\begin{pmatrix} a_0^2 & a_0a_1 & a_0a_2 & \ldots & a_0a_{p-1} \\ a_0a_1 & a_0^2+a_1^2 & a_0a_1+a_1a_2 & & \\ a_0a_2 & a_0a_1+a_1a_2 & a_0^2+a_1^2+a_2^2 & & \\ \vdots & & & \ddots & \\ a_0a_{p-1} & & & & a_0^2+a_1^2+\ldots+a_{p-1}^2 \end{pmatrix} \tag{4.18}$$

$$\mathbf{X}=(x_1, \ldots, x_n)^T, \quad \mathbf{X}_p=(x_1, \ldots, x_p)^T.$$

Jelölje $\mathbf{R}_p^{-1}$ $\xi(1), \ldots, \xi(p)$ kovarianciamátrixának inverzét, akkor $\mathbf{R}_p^{-1}=\frac{1}{\sigma^2}\mathbf{Q}_{p-1}^{p}$.

Legyen

$$L(\mathbf{X}, \sigma) = \log p(\mathbf{X}, \sigma) =$$

$$= \Omega_n - n \log \sigma - \frac{1}{2\sigma^2}\left[\mathbf{X}_p^T \mathbf{Q}_p^{-1} \mathbf{X}_p + \sum_{i=p+1}^{n} (x_i + a_1 x_{i-1} + \ldots + a_p x_{i-p})^2\right],$$

ahol

$$\Omega_n = \log [(2\pi)^{-n/2} |\mathbf{Q}_p|^{-1/2}].$$

A maximum likelihood egyenlet a következő

$$\frac{\partial L(\mathbf{X}, \sigma)}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\left[\mathbf{X}_p^T \mathbf{Q}_p^{-1} \mathbf{X}_p + \sum_{i=p+1}^{n} (x_i + a_1 x_{i-1} + \ldots + a_p x_{i-p})^2\right] = 0,$$

innen adódik $\sigma^2 = s$ maximum likelihood becslése

$$\hat{s}^* = \hat{\sigma}^{*2} = \frac{1}{n}\left[\mathbf{X}_p^T \mathbf{Q}_p^{-1} \mathbf{X}_p + \sum_{i=p+1}^{n} (x_i + a_1 x_{i-1} + \ldots + a_p x_{i-p})^2\right]. \tag{4.19}$$

Pitman-féle becslés.

Megmutatjuk, hogy a (4.19) maximum likelihood becslés szabályos, egyben torzítatlan. Valóban, (4.19)-ből következik, hogy

$$\hat{s}^*(\lambda \xi(t)) = \lambda^2 \hat{s}^*(\xi(t)),$$

továbbá

$$E(\hat{s}^*) = \frac{1}{n}\left[E(\mathbf{X}_p^T \mathbf{Q}_p^{-1} \mathbf{X}_p) + \sum_{i=p+1}^{n} E(\varepsilon_i^2)\right] =$$

$$= \frac{1}{n}[\mathrm{Tr}\, \mathbf{Q}_p^{-1} \mathbf{R}_p + E(\mathbf{X}_p^T) \mathbf{Q}_p^{-1} E(\mathbf{X}_p) + (n-p)\sigma^2] =$$

$$= \frac{1}{n}[\mathrm{Tr}\, \sigma^2 \mathbf{I}_p + (n-p)\sigma^2] =$$

$$= \frac{1}{n}[p\sigma^2 + (n-p)\sigma^2] = \sigma^2,$$

ahol $\mathrm{Tr}\, \mathbf{A}$ az $\mathbf{A}$ mátrix nyomát, $\mathbf{I}_p$ pedig a $(p \times p)$-egységmátrixot jelöli. A (4.17) összefüggésből belátható, hogy az

$$\hat{s}^* = \frac{1}{n}\left[\mathbf{X}_p^T \mathbf{Q}_p^{-1} \mathbf{X}_p + \sum_{i=p+1}^{n} (x_i + a_1 x_{i-1} + \ldots + a_p x_{i-p})^2\right]$$

statisztika elégséges a $p_{\xi(1),\ldots,\xi(n)}(x_1, \ldots, x_n; \sigma)$ sűrűségfüggvények összegére nézve.

Tekintsük most az

$$\hat{s}^* = \frac{1}{n}\left[\mathbf{X}_p^T \mathbf{Q}_p^{-1} \mathbf{X}_p + \sum_{i=p+1}^{n} (x_i + a_1 x_{i-1} + \ldots + a_p x_{i-p})^2\right] =$$

$$= \frac{1}{n}\left(\sigma^2 \mathbf{X}_p^T \mathbf{R}_p^{-1} \mathbf{X}_p + \sum_{i=p+1}^{n} \varepsilon_i^2\right) =$$

$$= \frac{\sigma^2}{n}\left(\mathbf{X}_p^T \mathbf{R}_p^{-1} \mathbf{X}_p + \sum_{i=p+1}^{n} \zeta_i^2\right)$$

statisztikát, ahol $\zeta_i = \frac{\varepsilon_i}{\sigma}$ független, $N(0, 1)$ normális eloszlású sorozat.

Figyelembe véve, hogy $\mathbf{X}_p^T \mathbf{R}_p^{-1} \mathbf{X}_p$, illetve $\sum_{i=p+1}^{n} \zeta_i^2$ $\varkappa^2(p, 0)$, illetve $\varkappa^2(n-p, 0)$ eloszlású, és $\mathbf{X}_p^T \mathbf{R}_p^{-1} \mathbf{X}_p$ független $\sum_{i=p+1}^{n} \zeta_i^2$-től, $\hat{s}^*$ sűrűségfüggvényét a következő alakban kapjuk:

$$\pi(\hat{s}^*, \sigma) = \frac{n}{\sigma^2} \frac{1}{\Gamma\left(\frac{n}{2}\right)} \frac{1}{2^{n/2}} \left(\frac{n}{\sigma^2} \hat{s}^*\right)^{\frac{n}{2}-1} e^{-\frac{n}{2\sigma^2}\hat{s}^*}. \tag{4.20}$$

Innen a *Lehmann-tétel* (lásd [31]) alapján következik, hogy $\hat{s}^*$ teljes elégséges statisztika. Ily módon a (4.19) maximum likelihood becslés az egyetlen minimális szórású torzítatlan becslés.

Az

$$\hat{s}^* = \frac{1}{n}\left[\mathbf{X}_p^T \mathbf{Q}_p^{-1} \mathbf{X}_p + \sum_{i=p+1}^{n} (x_i + a_1 x_{i-1} + \ldots + a_p x_{i-p})^2\right]$$

becslés szabályos, torzítatlan. Egyrészt a 4.1. tétel szerint az

$$\hat{u}^* = c\hat{s}^* \frac{E_1(\hat{s}^* | \mathscr{B}_0^{T_0})}{E_1(\hat{s}^{*2} | \mathscr{B}_0^{T_0})},$$

(ahol c meghatározható a $cE_1\left\{\frac{E_1^2(\hat{s}^* | \mathscr{B}_0^{T_0})}{E_1(\hat{s}^{*2} | \mathscr{B}_0^{T_0})}\right\} = 1$ összefüggésből) *Pitman-féle becslés* legkisebb szórású torzítatlan becslés az $\mathscr{S}$ osztályban, másrészt $\hat{s}^*$ az egyetlen legkisebb szórású torzítatlan becslés, ebből adódik, hogy

$$\hat{U}^* = \hat{s}^*,$$

tehát igaz a következő állítás:

4.2. Tétel. A (4.14) sztochasztikus differenciaegyenletnek eleget tevő $\xi(t)$ p-edrendű autoregressziós folyamat esetén σ^2-nek maximum likelihood becslése és *Pitman-féle becslése* megegyezik egymással, és az egyetlen legjobb torzítatlan becslést adják σ^2-re.

IRODALOM

[0] ACZÉL, J., *On Applications and Theory of Functional Equations* (J. Wiley, 1969).
[1] ARATÓ, M., „Folytonos állapotú Markov-folyamatok statisztikai vizsgálatáról, IV.", *A Magy. Tud. Akad. III. Oszt. Közleményei* **15** (1965) 107—124.
[2] ARATÓ, M., „Racionális spektrál sűrűségfüggvényű stacionárius folyamatok várható értékének megengedhető becsléséről", *A Magy. Tud. Akad. III. Oszt. Közleményei* **19** (1969) 89—99.
[3] ARATÓ, M., „Elemi Gauss-folyamatok statisztikai problémái", Doktori disszertáció. Budapest, 1969.
[4] BLACKWELL, D, "Conditional expectation and unbiased sequential estimation", *Ann Math. Stat.* **18** (1947) 105—110.
[5] BLYTH, C., „On minimax statistical decision procedures and their admissibility", *Ann. Math. Stat.* **22** (1951) 22—42.
[6] DOOB, J. L., "The elementary Gaussian processes", *Ann. Math. Stat.* **15** (1944) 229—281.
[7] FARRELL, R., "Estimators of a location parameter in the absolutely continuous case", *Ann. Math. Stat.* **35** (1964) 949—998.
[8] FELLER, W., *An Introduction to Probability Theory and Its Appliacations, Volume II.* (John Wiley, New York, 1966).
[9] FIEGER, W., „Eine statistische Charakterisierung der Normalverteilung", *Z. Wahrscheinlichkeitstheor. und verw. Geb.* **19** (1971) 330—344.
[10] Финтушал, С. М., «О допустимости оценки Питмэна для многомерного параметра сдвига», *Теория вероят. и её примен.* **XVI.** 4 (1971) 718—723.
[11] FOX, M. and RUBIN, H., "Admissibility of quantile estimates of a single location parameter", *Ann. Math. Stat.* **35** (1964) 1019—1030.
[12] Гихман, И. И. и Скороход, А. В., *Введение в теорию случайных процессов* (Москва, 1965).
[13] GIRSHICK, M. A. and SAVAGE, L. J., "Bayes and minimax estimates for quadratic loss function", *Proc. Second Berkeley Symp. Math. Statist. and Prob.* 1951, 53—73.
[14] HODGES, J. L. and LEHMANN, E. L., "Some applications of the Cramér—Rao inequality", *Proc. Second Berkeley Symp. Math. Statist. and Prob.* 1951, 13—22.
[15] Ибрагимов, И. А. и Хасьминский, Р. З., «О допустимости оценок Питмэна для параметра сдвига», *Записки научных семинаров ЛОМИ АН СССР* **29** 1972 57—61.
[16] JOSHI, V. M., "Admissibility of confidence intervals", *Ann. Math. Stat.* **37** (1966) 629—638.
[17] Каган, А. М., «Теория оценивания для семейств с параметром сдвига, масштаба и экспоненетных», *Труды Матем. ин-та им. Стеклова АН СССР* **104** (1968) 19—87.
[18] KAGAN, A. M., "On the estimation theory of location parameter", *Sankhyā, Series A.* **28** (1966) 335—352.
[19] KAGAN, A. M., "On ε-admissibility of the sample mean as an estimator of location parameter", *Sankhyā, Series A* **32** (1970) 37—40.
[20] KAGAN, A. M., Linnik, YU. V. and RAO, C. R., "On a characterization of the normal law based on a property of the sample average", *Sankhyā, Ser. A* **27** (1965) 405—406.
[21] Каган, А. М., Линник, Ю. В. и Рао, С. Р. *Характеризационные задачи математической статистики* (Изд-во «Наука», Москва, 1972).
[22] Каган, А. М. и Карпов, Ю. Н., «Байесовская постановка задачи оценивания параметра сдвига», *Зап. Науч. семинаров Ленингр. отд. мат. ин-та АН СССР* **29** (1972) 62—73.
[23] Каган, А. М. и Рухин, А. Л., «К теории оценивания параметра масштаба», *Теория вероятн. и её примен.* **XII** 4 (1967) 735—741.
[24] KAGAN, A. M. and ZINGER, A. A., "Sample mean as an estimator of location parameter. Case of nonquadratic loss functions", *Sankhy$_9$, Ser. A* **33** (1971) 351—358.
[25] KIEFER, J., "Invariance, minimax sequential estimation, and continuous time processes", *Ann. Math. Statist.* **28** (1957) 573—601.
[26] Клебанов, Л. Б., «Допустимость выборочного среднего как оценки параметра сдвига при полиномиальных ущербах», *ДАН СССР* **194** (1970) 508—509.
[27] Клебанов, Л. Б., «Допустимость выборочного среднего как оценки параметра сдвига при неквадратических ущербах», *Теория вероятн. и её прим.* **XVIII** 2 (1973) 339—349
[28] Клебанов, Л. Б., Линник, Ю. В. И Рухин, А. Л., «Несмещенное оценивание и матричные функции потерь», *ДАН СССР* **200** (1971) 1024—1025.
[29] Колмогоров, А. Н., «Несмещённые оценки», *Известия АН СССР сер. матем.* **14** (1950) 303—326.
[30] KUDO, H., "On minimax invariant estimators of the transformation parameter", *Nat. Sci. Rep. Ochanomizu Univ.* **6** (1955) 31—73.

[31] LEHMANN, E. L., *Testing Statistical Hypotheses* (New York, Wiley, 1959).
[32] LEHMANN, E. L. and SCHEFFÉ, H., "Completness, similar regions and unbased estimation", *Sankhyā* **10** (1950) 305—313.
[33] Линник, Ю. В., *Разложение вероятностных законов* (Изд-во. Ленинград. ин-та, 1960).
[34] Линник, Ю. В. и Рухин, А. Л., «Выпуклые функции потерь в теории несмещённого оценивания», *ДАН СССР* **198** (1971) 527—529.
[35 LINNIK, YU. V. and RUHIN, A. L., "Matrix loss function admitting the Rao-Black wellizaton", *Sankhy*$_9$, *Ser. A.* **34** (1972) 1—4.
[36] MARCINKIEWICZ, I., "Sur une propriété de la loi de Gauss", *Math. Zeitschrift* **44** (1938) 622—638.
[37] PARTHASARATHY, K. P., RANGA RAO, R. and VARADHAN, S. R. S., "Probability distributions on locally compact abelian groups", *Illinois Journal of Mathematics* **7** (1963) 337—369.
[38] PHAM NGOC PHUC, „Gauss-folyamatok egy statisztikai problémájáról", *MTA Számítástechnikai és Automatizálási Kutató Intézet Közlemények* **10** (1973) 45—57.
[39] PHAM NGOC PHUC, „Stacionárius folyamatok paraméterének becsléséről", *MTA Számítástechnikai és Automatizálási Kutató Intézet Közelmények* **10** (1973) 59—68.
[40] PHAM NGOC PHUC, „Gauss-folyamatok eltolási paraméterének Bayes-féle becslése", *MTA Számítástechnikai és Automatizálási Kutató Intézet Közlemények* **11** (1973) (Sajtó alatt.)
[41] PHAM NGOC PHUC, "On some problems of the estimation of location parameter in the case of Gaussian processes" *Proceedings of the Computer Science Conference — Hungary — Székesfehérvár — May* 21—26. (1973). (Sajtó alatt.)
[42] PITMAN, E., "The estimantion of location and scale parameter of a continuous population of any given form", *Biometrika* **30** (1938) 391—421.
[43] RAO, C. R., "On some characterisations of the normal law", *Sankhyā, Ser. A.* **29** (1967) 1—14.
[44] RAO, C. R., "Characterization of the distribution of random variables in linear structural relations", *Sankhyā, Ser. A.* **28** (1966) 251—260.
[45] RAO, C. R., "Information and accuracy attainable in estimation of statistical parameters", *Bull. Calcutta Math. Soc.* **37** (1945) 81—91.
[46] RAO, C. R., *Linear Statisctical Inference and Its Appliactions* (New York, John Wiley, 1965).
[47] Rao, C. R., "Some theorems on minimum variance estimation", *Sankhyā* **12** (1952) 27—42.
[48] Рухин, А. Л., «Некоторые статистические и вероятностные задачи на группах», *Труды. Матем. ин-та им. Стеклова АН СССР* (1970) 52—109.
[49] SARKADI, K., "On testing for normality", *Proc. Fifth. Berkeley Sympos. Math. Statist. and Prob.* **1** (1966) 373—387.
[50] STEIN., C., "The admissibility of Pitman estimator of a single location parameter", *Annals. Math. Stat.* **30** (1959) 970—978.
[51] STEIN, C., Inadmissibility of the usual estimator for the mean of a multivariate normal distribution", *Proc. Third Berkeley Sympos. Math. Statist. and Prob.* 1 (1956) 197—206.
[52] STEIN, C. and JAMES, W., "Estimation with quadratic loss", *Proc. Forth Berkeley Sympos. Math. Statist. and Prob.* 1 (1961) 361—379.
[53] STEIN, C., "Unbasied estimates with minimum variance", *Annals Math. Stat.* **21** (1950) 406—415.
[54] Зингер, А. А., Кадач, А. М. и Клебанов, Л. Б., «Выборочное среднее как оценка параметра сдвига при некоторых ущербах, отличных от квадратического», *Д АН СССР* **189** (1969) 29—30.

(Beérkezett: 1979. május 20.)

PHAM NGOC PHUC
MTA SZÁMÍTÁSTECHNIKAI ÉS AUTOMATIZÁLÁSI KUTATÓ INTÉZET
1502 BUDAPEST, KENDE U. 13—17.

ON CHARACTERIZATION PROBLEMS OF AUTOREGRESSION TYPE GAUSSIAN PROCESSES

PHAM NGOC PHUC

In the paper some results of KAGAN, LINNIK, RAO's book for independent random variables are extended to stationary autoregressive type processes. E.g. let $\xi(t)$ denote an autoregressive stochastic process, i.e.

$$\xi(t)+a_1\xi(t-1)+\ldots+a_p\xi(t-p)=\varepsilon(t),$$

where $\varepsilon(t)$ is white noise. If $x(t)=\xi(t)+\theta$, then (see Theorem 2.3) under some simple conditions on characteristic function of $\varepsilon(t)$ and if the best linear estimator of θ in the class of unbiased estimators is admissible $x(t)$ has to be *Gaussian.*

The optimal properties of estimators of multidimensional elementary *Gaussian processes* are also discussed in the paper. The *Pitman estimator* is analyzed for autoregressive type processes in case of location parameter and scale parameter as well.

Alkalmazott Matematikai Lapok **5** *(1979)* 375—384

MÁTRIXOK SZINGULÁRIS FELBONTÁSA

TUSNÁDY GÁBOR
Budapest

A dolgozat a lineáris leképezések szinguláris felbontását ismerteti. Megmutatja, hogyan lehet ezt a felbontást a főtengely-transzformáció előállításánál szokásos gondolattal, de az arra történő explicit hivatkozás nélkül származtatni. A szinguláris felbontás alapján több, elsősorban a többváltozós statisztikai vizsgálatokban felmerülő szélsőérték-feladat igen egyszerűen és szemléletesen oldható meg. Így többek között képet kapunk a faktor analízis, kanonikus korreláció és cluster analízis alapfeladatáról.

1. Lineáris leképezések szinguláris felbontása

Legyen $\mathbf{A}$ az n-dimenziós E_n euklideszi térnek a p-dimenziós E_p térre való tetszőleges lineáris leképezése. Az E_n tér $\mathbf{U}=(\mathbf{u}_1, \ldots, \mathbf{u}_n)$ ortonormált bázisát $\mathbf{A}$ saját bázisának nevezzük, ha

$$(\mathbf{A}\mathbf{u}_i, \mathbf{A}\mathbf{u}_j) = s_i^2 \delta_{ij}, \quad 1 \leqq i, j \leqq n, \tag{1.1}$$

ahol $(.,.)$ a skaláris szorzást, és δ_{ij} a *Kronecker-deltát* jelöli. Az s_i számok $\mathbf{A}$ szinguláris értékei, amelyek sorrendjéről feltesszük, hogy

$$s_1 \geqq s_2 \geqq \ldots \geqq s_n \geqq 0. \tag{1.2}$$

1.1. Tétel. Tetszőleges lineáris leképezésnek van saját bázisa.

Bizonyítás. Legyen $\mathbf{u}_1$ az E_n olyan eleme, melyre $\|\mathbf{u}_1\|=1$, $\|\mathbf{A}\mathbf{u}_1\| = \max_{\|\mathbf{u}\|=1} \|\mathbf{A}\mathbf{u}\|$, és legyen $\mathbf{u}$ az E_n tetszőleges eleme, melyre $(\mathbf{u}_1, \mathbf{u})=0$, $\|\mathbf{u}\|=1$, ahol $\|\mathbf{u}\|^2=(\mathbf{u}, \mathbf{u})$. Az

$$f(\alpha) = \|\mathbf{A}(\mathbf{u}_1 \cos\alpha + \mathbf{u} \sin\alpha)\|^2 =$$
$$= \|\mathbf{A}\mathbf{u}_1\|^2 \cos^2\alpha + 2(\mathbf{A}\mathbf{u}_1, \mathbf{A}\mathbf{u}) \cos\alpha \sin\alpha + \|\mathbf{A}\mathbf{u}\|^2 \sin^2\alpha$$

függvénynek $\mathbf{u}_1$ megválasztása miatt $\alpha=0$ mellett lokális maximuma van, ezért $f'(0)=2(\mathbf{A}\mathbf{u}_1, \mathbf{A}\mathbf{u})=0$. Ez azt jelenti, hogy $\mathbf{A}$ az E_n tér $\mathbf{u}_1$-re merőleges alterét az E_p tér $\mathbf{A}\mathbf{u}_1$-re merőleges alterére képezi le. (Feltehetjük, hogy $\|\mathbf{A}\mathbf{u}_1\|>0$, különben tételünk állítása nyilvánvaló.) Meggondolásunkat $\mathbf{A}$-nak ezekre az alterekre való megszorítására megismételve lépésről lépésre előállíthatjuk $\mathbf{U}$ elemeit.

Az $F \subset E_n$ alteret $\mathbf{A}$-ra nézve izotrópnak nevezzük, ha van olyan s szám, hogy

$$\|\mathbf{A}\mathbf{x}\| = s\|\mathbf{x}\| \tag{1.3}$$

teljesül F minden $\mathbf{x}$ elemére. Az $\mathbf{A}$-ra nézve izotróp F alteret maximálisnak mondjuk, ha (1.3) csak F elemeire teljesül. Bontsuk az $s_i=\|\mathbf{A}\mathbf{u}_i\|$, $i=1, 2, \ldots, n$

számsort egyenlő elemekből álló maximális blokkokra, akkor a megfelelő $\mathbf{u}_i$ bázis-elemek által kifeszített alterek $\mathbf{A}$ maximális izotróp alterei lesznek. Jelöljük ezeket az altereket $F_1, F_2, \ldots, F_n$-nel. (Az F_i alterek tehát nem feltétlenül különböznek, F_i pontosan akkor azonos F_j-vel, ha $s_i = s_j$.)

1.2. Tétel. Legyen $\tilde{\mathbf{U}}$ az $\mathbf{A}$ lineáris leképezés tetszőleges saját bázisa $\tilde{s}_1, \ldots, \tilde{s}_n$ szinguláris értékekkel, és jelölje s, $\mathbf{U}$, F az előbb megkonstruált mennyiségeket. Akkor $\tilde{\mathbf{u}}_i \in F_i$ és $\tilde{s}_i = s_i$, $i = 1, 2, \ldots, n$.

Bizonyítás. Legyen $1 \leqq i \leqq n$ tetszőleges,

$$\tilde{\mathbf{u}}_i = \sum_{j=1}^{n} x_j \mathbf{u}_j,$$

és legyenek az $1 \leqq \alpha < \beta \leqq n$ indexek ismét tetszőlegesek. Mivel a $\mathbf{v} = x_\beta \mathbf{u}_\alpha - x_\alpha \mathbf{u}_\beta$ vektor merőleges $\tilde{\mathbf{u}}_i$-ra, az $\mathbf{Av} = s_\alpha x_\beta \mathbf{v}_\alpha - s_\beta x_\alpha \mathbf{v}_\beta$ vektornak is merőlegesnek kell lennie az $\mathbf{A}\tilde{\mathbf{u}}_i = \sum_{j=1}^{n} s_j x_j \mathbf{v}_j$ vektorra, ahol $\mathbf{v}_i = \mathbf{Au}_i / s_i$ amíg $s_i > 0$, különben tetszőleges, a már definiált bázishoz csatlakozó ortonormált rendszer. Ebből következik, hogy

$$(s_\alpha^2 - s_\beta^2) x_\alpha x_\beta = 0.$$

Ez azt jelenti, hogy $\tilde{\mathbf{u}}_i$ nem-nulla koordinátáihoz egyenlő s_j szinguláris értékek tartoznak, vagyis $\tilde{\mathbf{u}}_i$ benne van valamelyik F_j altérben. Mivel ekkor $\tilde{s}_i = s_j$, és az $\tilde{s}_i$ számsor is fogyó, ez csak $s_i = s_j$ mellett teljesülhet. Ekkor viszont $\tilde{\mathbf{u}}_i \in F_i$ és $\tilde{s}_i = s_i$, amint azt bizonyítani akartuk.

Tételünk azt mondja ki, hogy ha $\mathbf{U}$ nem is egyértelmű, de csak az F_i altereken belüli megválasztása tetszőleges, maga az s_i számsor és az F_i alterek egyértelműen meghatározottak.

Az $\mathbf{A}$ leképezés adjungáltja az az E_p-t E_n-be vivő $\mathbf{A}'$ leképezés, melyre tetszőleges $\mathbf{x} \in E_n$, $\mathbf{y} \in E_p$ mellett

$$(\mathbf{Ax}, \mathbf{y}) = (\mathbf{x}, \mathbf{A}'\mathbf{y}). \tag{1.4}$$

(Azt, hogy (1.4) $\mathbf{A}'$-t egyértelműen definiálja, *Riesz tétele* alapján tudjuk.)

Legyen ismét $\mathbf{v}_i = \mathbf{Au}_i / s_i$ azokra az indexekre, amelyekre még $s_i > 0$ a többi indexre pedig válasszuk tetszőlegesen a $\mathbf{V}$ ortonormált bázis elemeit. Az $(\mathbf{U}, \mathbf{V})$ párt együtt $\mathbf{A}$ saját bázispárjának nevezzük. Arra az $\mathbf{A}'$ leképezésre, amely a $\mathbf{v}_i$ vektorokhoz $s_i \mathbf{u}_i$-t rendeli (ahol az esetleg még meg nem határozott s_i számok értéke 0) nyilván teljesül (1.4), ez tehát az $\mathbf{A}$ adjungáltja. Jelöljük az egyenlő s_i-knek megfelelő $\mathbf{v}_i$-k által kifeszített altereket G_i-vel, a legnagyobb indexet pedig, amelyre még $s_i > 0$ teljesül, jelöljük r-rel, ez utóbbi az $\mathbf{A}, \mathbf{A}'$ leképezések rangja. Ha $i \leqq r$, az $\mathbf{A}, \mathbf{A}'$ leképezések az F_i, G_i alterek elemeit rendelik egymáshoz mégpedig úgy, hogy az $\frac{1}{s_i}\mathbf{A}, \frac{1}{s_i}\mathbf{A}'$ leképezéseknek ezekre az alterekre vett megszorításai normatartóak (unitérek). Az $F_{r+1, r+1}$ alterek elemeihez pedig az $\mathbf{A}, \mathbf{A}'$ leképezések rendre a nullvektort rendelik. Ha az $\mathbf{A}$ és $\mathbf{A}'$ leképezések azonosak, $\mathbf{A}$-t önadjungáltnak mondjuk.

Lemma. Ha **A** unitér, és önadjungált, akkor az

$$E_n^+ = \{\mathbf{x}: \mathbf{Ax} = \mathbf{x}\}, \quad E_n^- = \{\mathbf{x}: \mathbf{Ax} = -\mathbf{x}\} \tag{1.5}$$

alterek ortogonálisak és direkt összegük maga az E_n tér.

Bizonyítás. Feltehetjük, hogy E_n^+ nem azonos E_n-nel, ez esetben legyen $\mathbf{x}$ az E_n tetszőleges, E_n^+-ra merőleges nem nulla eleme, $\mathbf{y}=\mathbf{Ax}$, $\mathbf{z}=\mathbf{x}+\mathbf{y}$. Mivel **A** unitér, $\mathbf{y}$ is merőleges E_n^+-ra, így $\mathbf{z}$ is az. Ha megmutatjuk, hogy

$$\mathbf{Ay} = \mathbf{x}, \tag{1.6}$$

készen is vagyunk, hiszen így $\mathbf{z}$ eleme E_n^+-nak is, tehát csak nullvektor lehet, vagyis $\mathbf{x}\in E_n^-$. Mivel **A** önadjungált,

$$(\mathbf{x}, \mathbf{Ay}) = (\mathbf{Ax}, \mathbf{y}) = \|\mathbf{y}\|^2,$$

és mivel **A** unitér, $\|\mathbf{x}\| = \|\mathbf{y}\| = \|\mathbf{Ay}\|$. Eszerint $\mathbf{x}$ és $\mathbf{Ay}$ normája egyenlő és skaláris szorzatuk egyenlő a normájuk négyzetével, ami csak úgy lehetséges, ha egyenlőek.

Ha tehát **A** önadjungált, akkor

$$F_i = F_i^+ \oplus F_i^-, \quad i = 1, 2, \ldots, n,$$

és az F_i^+ altér elemeihez **A** az s_i-szeresüket, az F_i^- altér elemeihez pedig a $(-s_i)$-szeresüket rendeli:

$$\mathbf{Au}_i = \varepsilon_i s_i \mathbf{u}_i,$$

ahol $\varepsilon_i = 1$ vagy $\varepsilon_i = -1$, feltéve, hogy **U** elemei vagy valamelyik F_i^+-ban, vagy valamelyik F_i^--ban vannak, amikor is ezek **A** saját vektorai, az $\varepsilon_i s_i$ számok pedig **A** saját értékei.

A lineáris leképezések szinguláris felbontását általában az önadjungált operátorok spektrálfelbontására vezetik vissza, miközben valamilyen módon „négyzetgyököt kell vonni" az **AA'**, **A'A** leképezésekből. Ezt az utat követi például FORSYTHE és MOLER az egyetlen (általam ismert) magyar nyelvű könyvben, amelyben a szinguláris felbontásról szó esik. A nem önadjungált operátorok vizsgálatát szokás az ún. polár tétellel is bevezetni, mely szerint ezek mindig előállíthatóak egy unitér és egy önadjungált operátor szorzataként. Ezt az utat követi például GOHBERG és KREIN, akik elég részletesen tárgyalják a szinguláris felbontást.

2. Kvadratikus és bilineáris alakok

Legyen **A** az E_n tér E_p-re való lineáris leképezése, és legyen $k \leqq n$, $k \leqq p$, továbbá

$$\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_k), \quad \mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_k)$$

az E_n, illetve E_p terek ortonormált vektorrendszerei. Sok feladatban szükség van a

$$Q(\mathbf{X}) = \sum_{i=1}^{k} (\mathbf{Ax}_i, \mathbf{x}_i), \quad Q(\mathbf{X}, \mathbf{Y}) = \sum_{i=1}^{k} (\mathbf{Ax}_i, \mathbf{y}_i)$$

ún. kvadratikus, illetve bilineáris alakok vizsgálatára. Ha $(\mathbf{U}, \mathbf{V})$ az $\mathbf{A}$ saját bázispárja, és $s_1, \ldots, s_n$ az $\mathbf{A}$ szinguláris értékei, akkor

$$\sum_{i=1}^{k} (\mathbf{A}\mathbf{u}_i, \mathbf{v}_i) = \sum_{i=1}^{k} s_i,$$

ha pedig $\mathbf{A}$ önadjungált, és $\mathbf{U}$ elemei saját vektorok, akkor

$$\sum_{i=1}^{k} (\mathbf{A}\mathbf{u}_i, \mathbf{u}_i) = \sum_{i=1}^{k} \varepsilon_i s_i.$$

Rendezzük nagyság szerint fogyó sorba az $\varepsilon_i s_i$ számokat, és jelöljük őket ebben az új sorrendjükben λ_1-gyel, λ_2-vel, ..., λ_n-nel, a megfelelő saját vektorokat pedig $\mathbf{e}_1$-gyel, $\mathbf{e}_2$-vel, ..., $\mathbf{e}_n$-nel.

2.1. Tétel. A fenti jelölések mellett $Q(\mathbf{X}, \mathbf{Y})$ maximuma $\sum_{i=1}^{k} s_i$, ha pedig $\mathbf{A}$ önadjungált, akkor $Q(\mathbf{X})$ maximuma $\sum_{i=1}^{k} \lambda_i$.

Bizonyítás. Ha $\mathbf{A}$ önadjungált, és a saját értékei nem negatívak, a második állítás az első következménye. Különben legyen $\mathbf{I}$ az E_n-beli identitás, és $\mathbf{B} := \mathbf{A} - \lambda_n \mathbf{I}$. Akkor

$$Q(\mathbf{X}) = \sum_{i=1}^{k} ((\mathbf{B} + \lambda_n \mathbf{I})\mathbf{x}_i, \mathbf{x}_i) = \sum_{i=1}^{k} (\mathbf{B}\mathbf{x}_i, \mathbf{x}_i) + k\lambda_n$$

miatt az általános eset visszavezethető a nem negatív sajátértékek esetére. Elég tehát az első állítást belátni.

Jelöljük $\mathbf{x}_i$, $\mathbf{y}_i$ $\mathbf{U}$-beli, $\mathbf{V}$-beli koordinátáit x_{ij}, y_{ij}-vel, akkor

$$Q(\mathbf{X}, \mathbf{Y}) = \sum_{i=1}^{k} \sum_{j=1}^{r} s_j x_{ij} y_{ij} = \sum_{j=1}^{r} s_j z_j, \tag{2.1}$$

ahol $z_j = \sum_{i=1}^{k} x_{ij} y_{ij}$.

Lemma. A

$$\sum_{j=1}^{r} s_j z_j \leqq \sum_{j=1}^{r} s_j w_j \tag{2.2}$$

egyenlőtlenség akkor és csakis akkor teljesül tetszőleges s_j, $j = 1, \ldots, r$ számsorra, melyre

$$s_1 \geqq \ldots \geqq s_r \geqq 0, \tag{2.3}$$

ha

$$\sum_{j=1}^{m} z_j \leqq \sum_{j=1}^{m} w_j; \quad m = 1, \ldots, r. \tag{2.4}$$

A mondott (2.4) feltétel ugyanis (2.2) speciális esete $s_1 = \ldots = s_i = 1$, $s_{i+1} = \ldots =$

$\ldots = s_r = 0$ mellett, tehát valóban szükséges. Elégségessége pedig abból következik, hogy (2.2) jobb és bal oldalának a különbsége

$$\sum_{m=1}^{r} \left(\sum_{j=1}^{m} w_j - \sum_{j=1}^{m} z_j \right) (s_m - s_{m+1})$$

alakra hozható, ahol $s_{r+1} = 0$, ez pedig (2.3) és (2.4) teljesülése esetén valóban nem negatív.

Visszatérve a tétel bizonyítására, legyen z_j a (2.1) alatti számsor, és legyen

$$w_1 = \ldots = w_k = 1, \quad w_{k+1} = \ldots = w_r = 0.$$

(Ha $k \geqq r$, legyen az összes w_j értéke 1.) Lemmánk alapján azt kell belátnunk, hogy ezekre teljesül (2.4). Mivel

$$z_j^2 \leqq \sum_{i=1}^{k} x_{ij}^2 \sum_{i=1}^{k} y_{ij}^2,$$

ha $m \leqq k$, (2.4) abból következik, hogy

$$\sum_{i=1}^{k} x_{ij}^2 \leqq 1, \quad \sum_{i=1}^{k} y_{ij}^2 \leqq 1, \tag{2.5}$$

ha pedig $m > k$, akkor abból, hogy

$$\left(\sum_{j=1}^{m} x_{ij} y_{ij} \right)^2 \leqq 1, \quad i = 1, \ldots, k.$$

Végül (2.5) abból következik, hogy tetszőleges $\mathbf{X}, \mathbf{Y}$ ortonormált rendszer bázissá egészíthető ki, és egy unitér operátor adjungáltja is unitér.

Jelöljük tetszőleges E_n, E_p-beli koordinátarendszerben az eddig használt vektorrendszerekből, mint oszlopvektorokból összeállított mátrixokat ugyanúgy, mint a vektorrendszereket. Akkor egy tetszőleges E_n-beli $\mathbf{z}$ vektor $\mathbf{U}$-beli koordinátáit $\mathbf{U'z}$, magát az $\mathbf{Az}$ transzformált vektort a $\mathbf{VSU'z}$ szorzat adja, ahol $\mathbf{S}$ az a $p \times n$-es mátrix, melynek a diagonálisában az s_i számok állnak, és a többi eleme 0. A $\mathbf{VSU'}$ alakot az $\mathbf{A}$ leképezés szinguláris felbontásának nevezzük. Jelöljük egy négyzetes $\mathbf{A}$ mátrix diagonálisában álló elemeinek összegét $\mathrm{tr}\,(\mathbf{A})$-val. Tételünk ebben az alakban azt mondja ki, hogy

$$\mathrm{tr}\,(\mathbf{Y'VSU'X}) \leqq \sum_{i=1}^{k} s_i \tag{2.6}$$

tetszőleges $n \times k$, illetve $p \times k$ méretű $\mathbf{X}, \mathbf{Y}$ mátrixokra, amelyekre $\mathbf{X'X}$ és $\mathbf{Y'Y}$ a k-dimenziós egységmátrixszal egyenlő. Látható, hogy itt $\mathbf{U}, \mathbf{V}$ feleslegesek, és ha $\mathbf{XY'}$ diagonális elemeit továbbra is z_j-vel jelöljük, akkor a lemma alapján (2.6) azzal ekvivalens, hogy

$$\sum_{j=1}^{m} z_j = \sum_{j=1}^{m} \sum_{i=1}^{k} x_{ij} y_{ij} \leqq \min(m, k).$$

Az történik tehát, hogy kivágunk az $\mathbf{X}, \mathbf{Y}$ mátrixokból egy-egy $k \times m$-es téglalapot, egymásra tesszük őket, és a megfelelő elemeket összeszorozzuk. Emiatt szim-

metrikus m és k szerepe, hiszen a két téglalapban a sorok skaláris szorzata is, oszlopok skaláris szorzata is becsülhető 1-gyel.

2.2. Tétel. Jelöljük az $\mathbf{A}$ leképezés szinguláris értékeit $s_i(\mathbf{A})$-val. és ha $\mathbf{A}$ önadjungált, a saját értékeit $\lambda_i(\mathbf{A})$-val, Tetszőleges $\mathbf{A}: E_n \to E_p$, $\mathbf{B}: E_n \to E_q$, $k \leqq n, p, q$ mellett

$$\sum_{i=1}^{k} s_i(\mathbf{AB}') \leqq \sum_{i=1}^{k} s_i(\mathbf{A}) s_i(\mathbf{B}); \quad \sum_{i=1}^{k} \lambda_i(\mathbf{AB}) \leqq \sum_{i=1}^{k} \lambda_i(\mathbf{A}) \lambda_i(\mathbf{B}). \tag{2.7}$$

Bizonyítás. Az $\tilde{\mathbf{A}} = \mathbf{A} + \Delta\mathbf{I}$, $\tilde{\mathbf{B}} = \mathbf{B} + \Delta\mathbf{I}$ leképezések segítségével mindig elérhető, hogy a saját értékek nem negatívak legyenek. Ekkor viszont a második állítás ekvivalens az elsővel, így elég azt bizonyítani. Legyen $\mathbf{A}$ szinguláris felbontása $\mathbf{VSU}'$, és legyen $\mathbf{X}, \mathbf{Y}$ tetszőleges E_q, illetve E_p-beli ortonormált rendszer. A 2.1. tétel alapján elég azt belátni, hogy

$$\operatorname{tr}(\mathbf{Y}'\mathbf{VSU}'\mathbf{B}'\mathbf{X}) = \operatorname{tr}(\mathbf{SU}'\mathbf{B}'\mathbf{XY}'\mathbf{V}) \leqq \sum_{i=1}^{k} s_i(\mathbf{A}) s_i(\mathbf{B}). \tag{2.8}$$

(A tr függvény definíciója alapján könnyen látható, hogy benne az argumentum tényezői ciklikusan szabadon átrendezhetőek.) Ha itt $\mathbf{U}$-t a $\mathbf{B}$-be, $\mathbf{V}$-t az $\mathbf{Y}$-ba beolvasztjuk és alkalmazzuk a lemmát, az állítás arra redukálódik, hogy a $\mathbf{B}'\mathbf{XY}'$ szorzat diagonálisában a bal felső m elem összege legfeljebb $\sum_{i=1}^{\min(k,m)} s_i(\mathbf{B})$, ahol m tetszőleges. Ha $m \geqq k$, ezt már tudjuk, ha pedig $m < k$, akkor a következő lemma alapján fejezhetjük be a bizonyítást.

Lemma. Legyen $\mathbf{B}$ tetszőleges $q \times n$-es mátrix, $m < n$ és legyen $\tilde{\mathbf{B}}$ a $\mathbf{B}$ első m oszlopából álló mátrix. Akkor

$$\sum_{i=1}^{m} s_i(\tilde{\mathbf{B}}) \leqq \sum_{i=1}^{m} s_i(\mathbf{B}).$$

Ez ismét a 2.1. tétel alapján látható be, hiszen e tétel alapján a bal oldal értéke egyenlő $\operatorname{tr}(\mathbf{Y}'\mathbf{BX})$ maximumával, ahol $\mathbf{Y}$ az E_p tetszőleges, $\mathbf{X}$ pedig E_n-nek az első m koordináta-irány alapján meghatározott alterébe eső m elemű ortonormált rendszere.

Ha $\mathbf{A}$ négyzetes, de nem szimmetrikus, az általa generált kvadratikus alak azonos a szimmetrikus $\frac{1}{2}(\mathbf{A}+\mathbf{A}')$ mátrix által generált kvadratikus alakkal, ezért elegendő a szimmetrikus kvadratikus alakokat vizsgálni. A 2.2. tétel következménye, hogy ha $\mathbf{A}, \mathbf{B}$ szimmetrikus $n \times n$-es mátrixok, akkor

$$\operatorname{tr}(\mathbf{AB}) \leqq \sum_{i=1}^{n} \lambda_i(\mathbf{A}) \lambda_i(\mathbf{B}). \tag{2.9}$$

Meglepő, hogy milyen nehéz erre az ártatlannak látszó állításra közvetlen bizonyítást találni. Ne feledjük, hogy a sajátértékek monoton fogyó sorba vannak rendezve. Ha $\pi(i)$ az első n természetes szám tetszőleges permutációja, akkor

$$\sum_{i=1}^{n} \lambda_i(\mathbf{A}) \lambda_{\pi(i)}(\mathbf{B}) \leqq \sum_{i=1}^{n} \lambda_i(\mathbf{A}) \lambda_i(\mathbf{B}), \tag{2.10}$$

amint az közvetlenül látható, de (2.8)-ból is kiolvasható. Belátható (2.9) úgy is, hogy előbb azt mutatjuk meg, hogy tr (**AB**) egyenlő a (2.10) bal oldalán álló, különböző $\pi(i)$ permutációk konvex kombinációjával.

A fenti egyenlőtlenségek A. HORNtól és KY FANtól származnak, általánosításuk megtalálható GOHBERG és KREIN már említett könyvében.

3. Többdimenziós normális eloszlás

Egydimenzióban az ún. standard normális eloszlás a $\varphi(t)=(2\pi)^{-1/2}e^{-t^2/2}$ sűrűségfüggvénnyel van meghatározva, nevezzük ennek mintájára n-dimenziós standard normális eloszlásúnak a $\zeta'=(\zeta_1, \ldots, \zeta_n)$ véletlen vektort, ha koordinátái függetlenek, és standard normális eloszlásúak. Könnyen látható, hogy ez az eloszlás független a koordináta-rendszer megválasztásától, például azért, mert sűrűségfüggvénye csak $\|\zeta\|$-tól függ. Tetszőleges $\mathbf{A}: E_n \to E_p$, $\mathbf{a} \in E_p$, $\mathbf{B}: E_n \to E_q$, $\mathbf{b} \in E_q$ mellett a

$$\xi = \mathbf{A}\zeta + \mathbf{a}, \quad \eta = \mathbf{B}\zeta + \mathbf{b}$$

véletlen vektorok többdimenziós normális eloszlásúak, és együttes eloszlásuk is többdimenziós normális. Ismét könnyen látható, hogy $E\xi=\mathbf{a}$, $E\eta=\mathbf{b}$, $\mathbf{\Sigma}_{11}=E(\xi-\mathbf{a})(\xi-\mathbf{a})'=\mathbf{AA}'$, $\mathbf{\Sigma}_{12}=E(\xi-\mathbf{a})(\eta-\mathbf{b})'=\mathbf{AB}'$, $\mathbf{\Sigma}_{22}=E(\eta-\mathbf{b})(\eta-\mathbf{b})'=\mathbf{BB}'$, továbbá ξ, η együttes eloszlását az $\mathbf{a}, \mathbf{b}, \mathbf{\Sigma}_{11}, \mathbf{\Sigma}_{12}, \mathbf{\Sigma}_{22}$ mennyiségek egyértelműen meghatározzák. Mi a továbbiakban feltesszük, hogy $\mathbf{a}=\mathbf{0}$, $\mathbf{b}=\mathbf{0}$. (Itt, és a továbbiakban E a várható érték jele.)

Ha a ξ, η változók közül η aktuális értékei könnyebben, olcsóbban, vagy hamarabb beszerezhetőek, kézenfekvő igény, hogy ξ értékeit ezek alapján kíséreljük meghatározni. Olyan $\mathbf{C}: E_q \to E_p$ leképezést keresünk tehát, melyre

$$(3.1) \qquad E\|\xi-\mathbf{C}\eta\|^2 = \operatorname{tr}\left((\mathbf{A}-\mathbf{CB})(\mathbf{A}-\mathbf{CB})'\right) = \operatorname{tr}\left(\mathbf{\Sigma}_{11}-2\mathbf{\Sigma}_{12}\mathbf{C}'+\mathbf{C}\mathbf{\Sigma}_{22}\mathbf{C}'\right)$$

minimális. Ha a p, q dimenziók értéke nagy, az η méréseivel egyenrangú költségként jelentkezhet a $\mathbf{C}\eta$ transzformáció végrehajtásának a költsége is, ilyenkor korlátozó feltételként előírhatjuk $\mathbf{C}$ rangját.

Ha $\mathbf{B}$ szinguláris felbontása $\mathbf{B}=\mathbf{WTZ}'$ és rangja r, a (3.1) alatti kifejezés minimuma csak a $\mathbf{Z}=(\mathbf{z}_1, \ldots, \mathbf{z}_r)$ vektorok által kifeszített altértől függ, hiszen ζ-nak erre az altérre eső vetületét, és csakis ezt a vetületet határozhatjuk meg η alapján

3.1. Tétel. Legyen az $\mathbf{A}\tilde{\mathbf{Z}}'$ leképezés szinguláris felbontása $\mathbf{VSU}'$, akkor tetszőleges $k \leqq n$ mellett a (3.1) alatti kifejezést minimalizáló legfeljebb k rangú leképezés a következő:

$$(3.2) \qquad \mathbf{C} = \mathbf{V}\tilde{\mathbf{S}}\mathbf{U}'\tilde{\mathbf{T}}\mathbf{W}',$$

ahol $\tilde{\mathbf{S}}$ első k diagonálisbeli eleme megegyezik $\mathbf{S}$ megfelelő elemével, a többi 0, $\tilde{\mathbf{T}}$ első r diagonálisbeli eleme egyenlő $\mathbf{T}$ megfelelő elemének a reciprokával, és a többi 0.

Bizonyítás. Mint már mondtuk, eleve feltehetjük, hogy a $\tilde{\mathbf{T}}\mathbf{W}'$ transzformációt elvégeztük. A feladatunk olyan $\tilde{\mathbf{C}}=\mathbf{PMQ}'$ legfeljebb k-adrangú leképezést találni, amelyre

$$(3.3) \qquad \operatorname{tr}\left((\tilde{\mathbf{A}}-\tilde{\mathbf{C}})(\tilde{\mathbf{A}}-\tilde{\mathbf{C}})'\right)$$

minimális, ahol $\tilde{\mathbf{A}}=\mathbf{A}\tilde{\mathbf{Z}}'$. Adott $\mathbf{M}$ mellett ez a feladat ekvivalens $\tilde{\mathbf{A}}\tilde{\mathbf{C}}'$ maximalizálásával, ami a 2.2. tétel alapján a $\tilde{\mathbf{C}}=\mathbf{VMU}'$ választással érhető el, amikor is a maximum értéke $\sum_{i=1}^{n} s_i m_i$, tehát adott $\mathbf{M}$ mellett (3.3) minimuma

$$\sum_{i=1}^{n} (s_i - m_i)^2. \tag{3.4}$$

Ha $\tilde{\mathbf{C}}$ rangja legfeljebb k, akkor itt a 0-tól különböző m_i-k száma legfeljebb k, és a (3.4) alatti kifejezés nyilván

$$m_i = \begin{cases} s_i, & \text{ha } i \leqq k, \\ 0, & \text{ha } i > k \end{cases}$$

mellett minimális.

Tekintve, hogy a tétel regressziószámításról szól, benne ξ és η szerepe különböző, a $\mathbf{\Sigma}_{11}$ kovariancia mátrix struktúrája szerepet kap a legjobb közelítést adó leképezésben, $\mathbf{\Sigma}_{22}$ struktúrája nem. Az utóbbi akár egységmátrix is lehetne, és helyreáll a szimmetria, ha $\mathbf{\Sigma}_{11}$ is egyenlő az egységmátrixszal. Ha már a $\begin{pmatrix}\xi\\ \eta\end{pmatrix}$ vektor kovarianciáját $\begin{pmatrix}\mathbf{I} & \mathbf{\Sigma}_{12}\\ \mathbf{\Sigma}_{12}' & \mathbf{I}\end{pmatrix}$ alakra hoztuk, ebben az alakban $\mathbf{\Sigma}_{12}$ szinguláris felbontása adja az ún. kanonikus korreláció kérdésére a választ. Erről részletesen ír LENGYEL TAMÁS, így itt nem térünk ki rá, csak megjegyezzük, hogy a fenti tétel akkor is alkalmazható, ha ξ és η azonosak, amikor is a faktoranalízis feladatára jutunk.

Eddigi vizsgálatunkban feltettük, hogy ξ és η együttes eloszlása adott, és azt kerestük, hogy ennek ismeretében hogyan lehet a két változó közül az egyik alapján a másikra becslést adni. Tegyük most fel, hogy $p=q$, és vizsgáljuk meg, adott $\mathbf{\Sigma}_{11}, \mathbf{\Sigma}_{22}$ mellett melyik az a $\mathbf{\Sigma}_{12}$ keresztkovariancia mátrix, amelyre

$$E\|\xi-\eta\|^2 = \operatorname{tr}\left((\mathbf{A}-\mathbf{B})(\mathbf{A}-\mathbf{B})'\right) \tag{3.5}$$

minimális. Ez a minimum ugyanis azt fejezi ki, hogy milyen mértékben tér el egymástól ξ és η kovarianciamátrixa.

3.2. Tétel. Adott $\mathbf{AA}'=\mathbf{\Sigma}_{11}$ $\mathbf{BB}'=\mathbf{\Sigma}_{22}$ mellett (3.5) akkor minimális, ha $\mathbf{A}$ és $\mathbf{B}$ saját bázisai megegyeznek.

Bizonyítás. Legyen $\mathbf{A}=\mathbf{VSU}'$, $\mathbf{B}=\mathbf{WTZ}'$, akkor $\mathbf{\Sigma}_{11}$ és $\mathbf{\Sigma}_{22}$ meghatározza a $\mathbf{V}, \mathbf{W}$ bázisokat és az $\mathbf{S}, \mathbf{T}$ szinguláris értékeket, de nem szól bele az $\mathbf{U}, \mathbf{Z}$ saját bázisok megválasztásába. Ismét elég $\operatorname{tr}(\mathbf{AB}')$ értékét maximalizálnunk, és a 2.2. tételből következik, hogy ehhez valóban arra van szükség, hogy $\mathbf{U}$ és $\mathbf{Z}$ megegyezzen.

GEORGES HUPETTŐL származik az ötlet, hogy a 3.2. tételt a következő lemma alapján bizonyítsuk.

Lemma. Ha $\mathbf{A}$ tetszőleges $n\times n$ méretű mátrix és $\mathbf{P}$ tetszőleges $n\times n$ méretű ortonormált mátrix, akkor $\operatorname{tr}(\mathbf{AP})$ akkor maximális, ha $\mathbf{AP}$ szimmetrikus.

Legyen ugyanis $\mathbf{A}=\mathbf{VSU}'$, akkor $\operatorname{tr}(\mathbf{AP})=\operatorname{tr}(\mathbf{SU}'\mathbf{PV})$, és itt $\mathbf{U}'\mathbf{PV}$ ortonormált, tehát diagonálisbeli elemeinek az értéke külön-külön legfeljebb 1. Ha pedig az, akkor $\mathbf{U}'\mathbf{PV}=\mathbf{I}$, $\mathbf{AP}=\mathbf{USU}'$, amint azt igazolni akartuk.

4. Cluster analízis

Az $\mathbf{A}: E_n \to E_p$ lineáris leképezés a gyakorlatban a legtöbbször egy adat-mátrixot jelent, amelynek a sorai az esetek, oszlopai a mérések (tehát n mérést végzünk p esetben). Előfordulhat, a mérés csupán abból áll, hogy valamilyen tulajdonság meglétét vagy hiányát regisztráljuk. Ilyenkor $\mathbf{A}$ ún. 0—1 mátrix, ami persze formálisan csak egy konkrét speciális eset, mivel azonban a többdimenziós analízis hallgatólagosan feltételezi a normális eloszlást, általában mint a nem-normalitás legfrappánsabb példája szerepel, és külön bánásmódot igényel. Ez a külön bánásmód általában az euklideszi metrikától eltérő metrikák használatában merül ki, és mindkét esetet (a normálist is, a nem-normálist is) általában értékelt gráfok vizsgálatára szokás visszavezetni. A csúcsok vagy az esetek, vagy a mérések aszerint, hogy melyik mennyiségben gyanakszunk nagyobb heterogenitásra, vagy a még nem ismert csoportosulásokat melyik mennyiségen keresztül kívánjuk jellemezni. A vizsgálat során tehát a vizsgálat tárgya három különböző dolog lehet:

a) adat-mátrix,

b) távolság (hasonlóság, különbözőség) -mátrix,

c) esetek vagy mérések csoportosítása,

és a vizsgálat célja az éppen kézben levő állapotból valamelyik másik állapotba való átmenet.

Nevezzük beágyazásnak azt a feladatot, amikor a távolság-mátrixból akarjuk az eredeti adat-mátrixot meghatározni. Erre nem azért van szükség, mert „elvesztek" az adatok, hanem mert csökkenteni akarjuk a dimenziót, az a meggyőződésünk, hogy ha látnánk az adatokat, biztosan tudnánk csoportosítani őket. Ilyenkor az eredeti (magas dimenziós) adatok távolság-mátrixát szeretnénk alacsony dimenziós adatok távolság-mátrixával közelíteni, de az is lehet, hogy nem távolság-mátrixunk van, mert a hasonlóságot vagy különbözőséget nem euklideszi metrikával mértük, ennek ellenére valamilyen *a*-típusú adatmezőre kialakított módszert szeretnénk alkalmazni.

4.1. Tétel. Ha $\mathbf{A}=\mathbf{VSU}'$ és $\mathbf{B}$ tetszőleges, legfeljebb k rangú mátrix, $\operatorname{tr}((\mathbf{A}-\mathbf{B})(\mathbf{A}-\mathbf{B})')$ akkor minimális, ha $\mathbf{B}=\mathbf{V}\tilde{\mathbf{S}}\mathbf{U}'$, ahol $\tilde{\mathbf{S}}$ diagonálisában az első k elem megegyezik $\mathbf{S}$ megfelelő elemével, a többi 0.

Ez a tétel a 3.1. tétel speciális esete, így nem bizonyítjuk be. A szemléletes jelentése a következő. Adott mondjuk az n-dimenziós térben p pont, és ezeket olyan pontokkal szeretnénk közelíteni, hogy a közelítő pontok által kifeszített altér dimenziója legfeljebb k legyen, és a megfelelő pontpárok távolságnégyzetének összege minimális legyen. Ez a tétel önmagában nem oldja meg az eredeti feladatot, mely szerint az $n\times n$-es $\mathbf{A}$ mátrixhoz olyan k-dimenziós $\mathbf{x}_1, \ldots, \mathbf{x}_n$ pontokat kell keresnünk, melyre

$$\sum_{i=1}^{n}\sum_{j=1}^{n}(a_{ij}-\|\mathbf{x}_i-\mathbf{x}_j\|^2)^2$$

minimális. Ennek a feladatnak a megoldása igen hasznos volna a cluster analízis számára.

Végül megköszönöm KRÁMLI ANDRÁSnak, MAJOR PÉTERnek és SZÉP ANDRÁSnak a dolgozatom megírásában nyújtott segítségüket.

IRODALOM

[1] AMIR—MOÉZ, A. R. és HORN, A., "Singular values of a matrix", *Amer, Math. Monthly* **65** (1958) 742—748.
[2] ANDERSON, T. W., *An Introduction to Multivariate Statistical Analysis* (Wiley, New York, 1958).
[3] BJÖRCK, A. and GOLUB, G. H., "Numerical methods for computing angles between linear subspaces", *Mathematics of Computation* **27** (1973) 579—594.
[4] BUSINGER, P. and GOLUB, G. H., "Linear least squares solutions by *Householder* transformations", *Numer. Math.* **7** (1965) 269—276.
[5] FORSYTHE, G. E. és MOLER, C. B., *Lineáris algebrai problémák megoldása számítógéppel* (Műszaki Könyvkiadó, Bp., 1976).
[6] GOHBERG, I. C. and KREIN, M. G., *Introduction to the Theory of Linear Nonselfadjoint Operators* (Nauka, Moszkva, 1965, Providence, 1969).
[7] GOOD, I. J., "Some applications of the singular decomposition of a matrix", *Technometrix* **11** (1969) 823—831.
[8] GOLUB, G. H., "Numerical methods for solving linear least square problems", *Numer. Math.* **7** (1965) 206—216.
[9] GOLUB, G. H. and KAHAN, W., "Calculating the singular values and pseudoinverse of a matrix", *J. SIAM Numer. Anal. Ser. B* **2** (1965) 205—224.
[10] GOLUB, G. H. and REINSCH, C., "Singular values decompositions and least squares solutions", *Numer. Math.* **14** (1970) 403—420.
[11] HORN, A., "On the singular values of a product of completely continuous operators", *Proc. Nat. Acad. Sci. USA* **36** (1950) 374—375.
[12] HORN, A., "On the eigenvalues of a matrix with prescribed singular values", *Proc. Amer. Math. Soc.* **5** (1954) 4—7.
[13] KY FAN, "Maximum properties and inequalities for the eigenvalues of completely continuous operators", *Proc. Nat. Acad. Sci. USA* **37** (1951) 760—766.
[14] LANCZOS, C., "Linear systems in self-adjoint form", *Amer. Math. Monthly* **65** (1958) 665—679.
[15] LENGYEL, T., „A kanonikus korrelációanalízis és néhány kapcsolódó probléma", *Alkalmazott Matematikai Lapok* **5** (1979).
[16] MARSAGLIA, G. and STYAN, G. P. H., "Equalities and inequalities for ranks of matrices", *Linear and Multilinear Algebra* **2** (1974) 269—292.
[17] RAO, C. R., *Linear Statistical Inference and its Applications* (Second edition, J. Wiley, 1973).
[18] RIESZ, F. és SZŐKEFALVI-NAGY, B., *Lecons d'Analyse fonctionelle* (Akadémiai Kiadó, Bp., 1955).
[19] RÓZSA, P., *Lineáris algebra és alkalmazásai* (Műszaki Könyvkiadó, Bp., 1976).
[20] SESHADRI, V. and STYAN, G. P. H., "Canonical correlations, rank additivity and characterizations of multivariate normality", Bolyai Kollokvium, 1978.
[21] TUSNÁDY, G., "Strong invariance principles", *Recent developments in statistics*, ed. *J. R. Barra*, (1977) 289—300.
[22] WHITTLE, P., "On principal components and least square methods of factor analysis", *Skand. Aktuar.* **35** (1952) 223—239.

(Beérkezett: 1979. szeptember 7.)

TUSNÁDY GÁBOR
MTA MATEMATIKAI KUTATÓ INTÉZETE
1053 BUDAPEST, REÁLTANODA U. 13—15.

ON THE SINGULAR DECOMPOSITION OF MATRICES

G. TUSNÁDY

In the paper a new approach is given to the singular decomposition of matrices. The basic idea is similar to that what is used in the proof of the existence of the eigen-vectors of a self-adjoint operator, but there is no explicit reference to this fact. The singular decomposition turns out to be an effective tool in proving extremal properties of matrices. The investigated questions appeared in factor analysis, canonical correlation and cluster analysis.

Alkalmazott Matematikai Lapok **5** *(1979)* 385—393

A KANONIKUS KORRELÁCIÓANALÍZIS ÉS NÉHÁNY KAPCSOLÓDÓ PROBLÉMA

LENGYEL TAMÁS
Budapest

Ebben a cikkben a kanonikus korrelációanalízis módszerére és alkalmazási lehetőségeire szeretnénk felhívni a figyelmet.

A kanonikus korrelációs együtthatók és a kanonikus faktorok néhány, a szórás- és a kovarianciamagyarázat[1] szempontjából optimalizáló tulajdonságát ismertetjük.

1. Bevezetés

Gyakorlati problémák megoldása során gyakran merül fel az igény 2 valószínűségi vektorváltozó közötti viszony mérésére. A következő példa [3] segítségével illusztrálom az ilyen vizsgálatok szükségességét.

Szívkoszorúér-megbetegedések előrejelzése volt a feladatunk. 5 évenként 3 alkalommal összesen 1082 ember egészségi állapotára jellemző adatokat, dohányzási szokásokat regisztráltak. Abból indultunk ki, hogy ezek az adatok sok információt tartalmazhatnak nemcsak az aktuális egészségi állapotról, hanem annak közeljövőben várható alakulásáról is. A feladat éppen a jövőbeli állapot előrejelzése és értékelése volt. Ennek megfelelően a különböző alkalmakkor mért adatok közötti összefüggésre voltunk kíváncsiak, valamint arra, hogyan lehet ezt az összefüggést felhasználni egyiküknek a másikkal történő becslése céljából. A feladatot a kanonikus korrelációanalízis segítségével oldottuk meg. Eredményeink alapján megállapíthattuk, hogy az egészséges emberek beteggé válása során nagyobb az összefüggés az aktuális és az 5 év múlva mért adatok között, mint akkor, amikor egészségesek maradnak. A várható állapotra vonatkozó becsléseket diszkriminancia analízis segítségével osztályozni is sikerült.

A feladatot megfogalmazhatjuk általánosabban is.

Tegyük fel, hogy az $\mathbf{U}_1$ és $\mathbf{U}_2$ valószínűségi változók egyszerre nem mérhetők, például időben nem egyszerre lezajló eseményeket írnak le. Ha azonban szoros kapcsolat van a 2 esemény között, akkor úgy gondolhatjuk, hogy az egyik mérésének a felhasználásával a másikat is jól leírhatjuk, becsülhetjük.

A példa terminológiáját használva, ha $\mathbf{U}_1$-et mérjük és az $\mathbf{U}_2$-re szeretnénk becslést kapni, akkor szerepüknek megfelelően $\mathbf{U}_1$-et szokás prediktornak, $\mathbf{U}_2$-t pedig predikátumnak nevezni.

Egy másik alkalmazási lehetőségre utal a következő példa.

[1] Szórásmagyarázaton valamilyen valószínűségi vektorváltozó koordinátánkénti szórásnégyzeteinek, kovarianciamagyarázaton pedig kovarianciamátrixa minden elemének (tehát nemcsak az átlójában levő szórásnégyzeteknek) valamilyen normában vett vizsgálatát értjük.

Ha egy sok dimenziós, bonyolult jelenséget leíró valószínűségi változók bizonyos csoportjait reprezentáló $\mathbf{U}_1$ és $\mathbf{U}_2$ valószínűségi vektorváltozók között szoros a kapcsolat, akkor nyilván felesleges a mindkét csoportra vonatkozó megfigyelési értékeket figyelembe venni a probléma analizálása során, hiszen bármelyik ilyen csoport a másikhoz hasonló mennyiségű információval szolgálhat. Ezzel bizonyos tárolási redundanciák kiküszöbölhetők. Ha az egyes csoportokra vonatkozó mérési költségek lényegesen különböznek, akkor csak a „legolcsóbb" csoport változóinak a mérésével lényeges mérési költségmegtakarítás is elérhető.

Az első példa jól rávilágít a probléma lényegére. Két valószínűségi vektorváltozó közötti kapcsolatra vagyunk kíváncsiak és ha ezt szorosnak ítéljük, akkor ennek a hipotézisnek a felhasználásával az egyik változónak a másikkal való lineáris regressziós becslését szeretnénk előállítani. A második példa a faktoranalízissel való kapcsolatra utal. Később látni fogjuk, hogy a faktoranalízis a kanonikus korrelációanalízis speciális esetének tekinthető.

A kanonikus korrelációanalízis alkalmas eszköz a fentiekhez hasonló feladatok megoldására.

Megjegyezzük, hogy két nominális típusú valószínűségi változó közötti összefüggés nagyságának a mérésére is lehetőség van a kanonikus korrelációanalízis felhasználásával [1].

2. A kanonikus korrelációanalízis elve

Jelölje $\mathbf{U}_1$ és $\mathbf{U}_2$ azt a két valószínűségi vektorváltozót, amelyek közötti kapcsolatot szeretnénk vizsgálni. Tegyük fel, hogy az összes komponensük standardizált, azaz várható értékük nullával, szórásuk eggyel egyenlő. Jelölje $\operatorname{cov}(\boldsymbol{\xi}, \boldsymbol{\eta})$ a $\boldsymbol{\xi}$ és $\boldsymbol{\eta}$ valószínűségi vektorváltozók kovarianciamátrixát, azaz az (i, j) eleme $\operatorname{cov}(\xi_i, \eta_j)$-vel egyenlő, ahol ξ_i a $\boldsymbol{\xi}$ változó i-edik komponense. Legyen $\boldsymbol{\Sigma}_{ij} = \operatorname{cov}(\mathbf{U}_i, \mathbf{U}_j)$ $(i, j = 1, 2)$. A feltételek miatt most a kovariancia- és a korrelációmátrix megegyezik. Tegyük fel, hogy az $\mathbf{U}_1$ q-dimenziós, az $\mathbf{U}_2$ $p-q$-dimenziós valószínűségi vektorváltozó, $p > q > 0$ egészek, valamint $\operatorname{rang}(\boldsymbol{\Sigma}_{11}) = q$, $\operatorname{rang}(\boldsymbol{\Sigma}_{22}) = p-q$. Jelölje $n = \min\{q, p-q\}$, $m = \operatorname{rang}(\boldsymbol{\Sigma}_{12})$; $\mathbf{A}'$, illetve $\mathbf{A}^{-1}$ az $\mathbf{A}$ mátrix transzponáltját, illetve inverzét.

Két, egydimenziós valószínűségi változó közötti összefüggés mérésének egy általánosan használt mérőszáma a korrelációs együttható. Ismeretes, hogy a két változó egymásra vonatkoztatott regressziós egyenesének az együtthatója és a korrelációs együttható között milyen kapcsolat van. Jelölje r az U_1 és U_2 korrelációs együtthatóját. Ekkor: $\hat{U}_2 = rU_1$ a regressziós egyenes és $E(U_2 - \hat{U}_2)^2 = 1 - r^2$, amit úgy is mondhatunk, hogy az U_2 szórásnégyzetéből r^2-nyit magyaráz az U_1 változó.

Ugyancsak közismert mérőszám az egydimenziós és a többdimenziós valószínűségi változók közötti kapcsolat mérése az ún. többszörös korreláció. Ez a regresszióanalízisben éppen az U_2 egydimenziós változónak és az $\mathbf{U}_1$ vektorváltozó lineáris függvényével való legkisebb négyzetes becslésének $(\hat{U}_2)$ a korrelációs együtthatója: $r_{U_2 \cdot \mathbf{U}_1}$. A többszörös korreláció ugyanakkor a maximális korreláció az $\mathbf{U}_1$ és U_2 lineáris függvényei között. Most

$$E(U_2 - \hat{U}_2)^2 = 1 - r^2_{U_2 \cdot \mathbf{U}_1}.$$

Tehát U_2 szórásnégyzetéből $r^2_{U_2 \cdot \mathbf{U}_1}$-nyit magyaráz az $\mathbf{U}_1$.

Általában is az $\mathbf{U}_1$ q-dimenziós és az $\mathbf{U}_2$ $p-q$-dimenziós valószínűségi vektorváltozók közötti kapcsolatot jellemezhetjük a lineáris függvényeik közötti maximális korrelációval. Ezt a számot nevezzük a két változó kanonikus korrelációjának. Analitikus úton bevezethetünk n darab kanonikus korrelációs együtthatót (vagy röviden kanonikus korrelációt), illetve faktort. Ennek segítségével lehetőség van áttérni egy olyan koordináta-rendszerre (vagy más szóval faktortérre), amelyben az $\mathbf{U}_1$ és $\mathbf{U}_2$ komponensei korrelálatlanok, kivéve $\mathbf{U}_1$ és $\mathbf{U}_2$ ugyanolyan sorszámú koordinátáit, melyek viszont „jól" korreláltak. Ebben a térben az $\mathbf{U}_1$-nek az i-edik komponense szórásnégyzetéből az $\mathbf{U}_2$ éppen annyit magyaráz, amennyit az i-edik koordinátája.

A két változó első kanonikus korrelációja alatt tehát a következő értéket értjük:

$$\varrho_1 := \max r(\mathbf{L}_1'\mathbf{U}_1, \mathbf{M}_1'\mathbf{U}_2),$$

$$\mathbf{L}_1 \in R^q, \quad \mathbf{M}_1 \in R^{p-q}$$

$$D^2(\mathbf{L}_1'\mathbf{U}_1) = D^2(\mathbf{M}_1'\mathbf{U}_2) = 1,$$

ahol $r(\xi, \eta)$ a ξ és η egydimenziós valószínűségi változók korrelációs együtthatóját, míg $D^2(\xi)$ a ξ szórásnégyzetét jelöli.

Könnyen látható, hogy a fenti maximum eléretik. $\mathbf{L}_1'\mathbf{U}_1$-et az első bal oldali, $\mathbf{M}_1'\mathbf{U}_2$-t az első jobb oldali kanonikus faktornak nevezzük; $\mathbf{L}_1$-et, illetve $\mathbf{M}_1$-et (az eredeti koordináta-rendszerre vonatkozó) bal-, illetve jobb oldali kanonikus együtthatóknak nevezzük.

Az i-edik kanonikus korrelációt $(1 < i \leqq n)$, illetve bal és jobb oldali faktorokat a következőképpen definiáljuk:

$$\varrho_i := \max r(\mathbf{L}_i'\mathbf{U}_1, \mathbf{M}_i'\mathbf{U}_2)$$

$$\mathbf{L}_i \in R^q, \quad \mathbf{M}_i \in R^{p-q},$$

$$D^2(\mathbf{L}_i'\mathbf{U}_1) = D^2(\mathbf{M}_i'\mathbf{U}_2) = 1,$$

$$\operatorname{cov}(\mathbf{L}_i'\mathbf{U}_1, \mathbf{L}_j'\mathbf{U}_1) = \operatorname{cov}(\mathbf{M}_i'\mathbf{U}_2, \mathbf{M}_j'\mathbf{U}_2) = 0, \quad 1 \leqq j < i.$$

Nyilván $\varrho_1 \geqq \varrho_2 \geqq \ldots \geqq \varrho_n \geqq 0$.

Az eredeti változók komponenseinek a megfelelő oldali kanonikus faktorokra vonatkozó lineáris regressziós együtthatóit faktorsúlyoknak nevezzük. Mivel a bal, illetve jobb oldali kanonikus faktorok korrelálatlanok, ezért ezek az együtthatók megegyeznek az eredeti változók komponenseinek a megfelelő oldali kanonikus faktorokkal vett korrelációival.

Az eredeti $\mathbf{U}_1$, illetve $\mathbf{U}_2$ változókra végrehajtott tetszőleges reguláris lineáris transzformációra nézve a kanonikus korrelációk és faktorok invariánsak, ugyanakkor a megfelelő kanonikus együtthatók, illetve faktorsúlyok egyszerűen adódnak.

A *Lagrange-féle multiplikátoros eljárásból* adódik, hogy [7]:

$$(\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} - \varrho_i^2\boldsymbol{\Sigma}_{22})\mathbf{M}_i = \mathbf{0},$$

és

$$(\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} - \varrho_i^2\boldsymbol{\Sigma}_{11})\mathbf{L}_i = \mathbf{0} \quad (1 \leqq i \leqq n).$$

Ugyancsak belátható, hogy

$$\varrho_i = 0, \quad \text{ha} \quad m < i \leqq n,$$

és

$$\operatorname{cov}(\mathbf{l}_i' \mathbf{U}_1, \mathbf{M}_j' \mathbf{U}_2) = \varrho_i \delta_{ij} \quad (i, j = 1, 2, \ldots, n),$$

ahol δ_{ij} a *Kronecker-delta.*

3. A kapcsolat mérésének néhány lehetősége

Tegyük fel, hogy $\mathbf{U}_1$ és $\mathbf{U}_2$ együttes eloszlása p-dimenziós normális.

Aligha várható, hogy 2 valószínűségi vektorváltozó közötti kapcsolat egyetlen számmal, pl.: a legnagyobb kanonikus korrelációval jellemezhető lenne. A következő példával jól illusztrálhatjuk ezt.

Ha $\mathbf{U}_1$ és $\mathbf{U}_2$ olyan, hogy első komponensük megegyezik, de az összes többi korrelálatlan, akkor $\varrho_1 = 1$, és $\varrho_i = 0$ $(i = 2, 3, \ldots, n)$.

Nehezen interpretálható az összefüggés akkor is, ha az összes nem nulla kanonikus korrelációt, mint különálló számot próbáljuk vizsgálni.

BARTLETT (1974) vezette be a következő eljárást, annak eldöntésére, hogy vajon hány kanonikus korreláció különbözik szignifikánsan a nullától, azaz hány kanonikus faktor szükséges a két valószínűségi vektorváltozó közötti összefüggés kifejezéséhez. Ha úgy találjuk, hogy a kanonikus korrelációk összességükben különböznek a nullától, akkor a következő lépésben azt teszteljük, hogy a legnagyobb korrelációtól eltekintve, a többi különbözik-e nullától, azaz hogy az első kanonikus korreláció elegendő-e a 2 vektorváltozó közötti kapcsolat leírásához. Ennek a hipotézisnek az elutasítása esetén addig folytatjuk ezt az eljárást, amíg először nem fogadjuk el azt a feltevést, hogy a megmaradó korrelációk már szignifikánsan nem különböznek a nullától. BARTLETT szerint a gyakorlati értékkel rendelkező kanonikus faktorok száma legfeljebb akkora, mint a szignifikánsan nem nulla korrelációk száma.

Az első lépésben

$$H_0^{(1)}: \quad \mathbf{U}_1 \text{ és } \mathbf{U}_2 \text{ független},$$

$$H_1^{(1)}: \quad \mathbf{U}_1 \text{ és } \mathbf{U}_2 \text{ nem független}.$$

Az $x_1^2 = -(N-1-0{,}5(p+1)) \ln \Lambda^{(1)}$ közelítőleg χ^2 eloszlást követ $q(p-q)$ szabadsági fokkal a $H_0^{(1)}$ fennállása esetén, ahol N a megfigyelések száma, és

$$\Lambda^{(j)} = \prod_{i=j}^{n} (1 - \varrho_i^2) \quad (j = 1, 2, \ldots, n).$$

Ha $H_0^{(1)}$-et elutasítjuk, akkor az első kanonikus faktorok hatását elhagyjuk és a maradékok függetlenségét teszteljük az

$$x_j^2 = -(N-1-0{,}5(p+1)) \ln \Lambda^{(j)}$$

teszt segítségével, ahol a megfelelő hipotézisek

$$H_0^{(j)}: \quad \mathbf{U}_1^{(j)} \text{ és } \mathbf{U}_2^{(j)} \text{ független},$$

$$H_1^{(j)}: \quad \mathbf{U}_1^{(j)} \text{ és } \mathbf{U}_2^{(j)} \text{ nem független}$$

és

$$\mathbf{U}_1^{(j)} = \mathbf{U}_1 - \boldsymbol{\Sigma}_{11}\tilde{\mathbf{L}}_{j-1}\tilde{\mathbf{L}}'_{j-1}\mathbf{U}_1,$$

$$\mathbf{U}_2^{(j)} = \mathbf{U}_2 - \boldsymbol{\Sigma}_{22}\tilde{\mathbf{M}}_{j-1}\tilde{\mathbf{M}}'_{j-1}\mathbf{U}_2 \quad (j = 2, 3, \ldots, n),$$

azaz az eddig szignifikánsan nem nulla „hatású" kanonikus faktorok által magyarázott részek elhagyása után adódó maradékok (vö.: 5. pont). A $H_0^{(j)}$ fennállása esetén az x_j^2 közelítőleg χ^2 eloszlást követ $(q-j+1)(p-q-j+1)$ szabadsági fokkal.

A 2. pontban láttuk, hogy milyen szoros kapcsolat van a szórásmagyarázat és a változók összefüggését mérő korreláció között. A szórásmagyarázat általánosításával az összefüggés mérésének általánosításához juthatunk [2], [5]. Két valószínűségi vektorváltozó közötti kapcsolat nagyságának a mértékét azzal a két számmal jellemezhetjük, amely azt mutatja, hogy az egyik változó a másik össz-szórásnégyzetéből átlagosan mennyit magyaráz az előbbi kanonikus faktoraira vonatkozó legkisebb négyzetes lineáris becslések révén. Attól függően, hogy az $\mathbf{U}_2$-t és az $\mathbf{U}_1$-et milyen koordináta-rendszerben írjuk fel, ezek a számok különböző értéket vehetnek fel. Pl.: ha a kanonikus faktorok terében írjuk fel őket, akkor

$$R_1^2 = \frac{1}{q}\sum_{i=1}^{m}\varrho_i^2$$

$$R_2^2 = \frac{1}{p-q}\sum_{i=1}^{m}\varrho_i^2.$$

Ezek a mennyiségek (illetve tagjaik) a két vektorváltozó egyidejű tárolásával keletkező, az egymás szórásnégyzetéből magyarázott redundáns információ nagyságát mérik. Az ilyen jellegű vizsgálatokat redundanciaanalízisnek nevezzük. Úgy gondoljuk, hogy az eredeti változókból az ellenkező oldali kanonikus faktorok által magyarázott mennyiségek (redundanciák) a kanonikus korrelációnál komplexebb magyarázatát adják a 2 eredeti változó közötti összefüggésnek. Ennek részleteit illetően a [2], [3], [4], [5] munkákban található utalás.

Megjegyezzük, hogy a két nominális típusú valószínűségi változó közötti összefüggés mérésére használatos C^2 *Cramér-mérték* éppen a megfelelő indikátor vektorváltozókra vonatkozó nem nulla kanonikus korrelációk négyzetes közepével egyenlő [1], azaz $q=p-q=m$ esetén pontosan az itt definiált mérőszámmal, a redundanciával $(C^2=R_1^2=R_2^2)$.

4. A kanonikus korrelációanalízis egyenleteinek megoldása

Az optimumot szolgáltató $\mathbf{L}, \mathbf{M}$ vektorok kielégítik a következő egyenletrendszert [7]:

$$(\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} - \varrho^2\boldsymbol{\Sigma}_{22})\mathbf{M} = \mathbf{0}, \tag{4.1}$$

$$(\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} - \varrho^2\boldsymbol{\Sigma}_{11})\mathbf{L} = \mathbf{0}, \tag{4.2}$$

ahol $\varrho = r(\mathbf{L}'\mathbf{U}_1, \mathbf{M}'\mathbf{U}_2)$.

Ha ezt az egyenletrendszert megoldjuk, akkor az $\mathbf{L}, \mathbf{M}$ vektorok ezen megoldások közül kiválaszthatók.

Tegyük fel, hogy rang $(\mathbf{\Sigma}_{11})=q$ és rang $(\mathbf{\Sigma}_{22})=p-q$. Ismertetjük a (4.1) feladat megoldását (a (4.2) megoldása ehhez hasonlóan történhet).

Állítás: Legyen $\mathbf{A}$ $m\times m$-es valós szimmetrikus, $\mathbf{B}$ $m\times m$-es pozitív definit szimmetrikus mátrix. Ekkor létezik olyan $m\times m$-es $\mathbf{R}$ mátrix, hogy

$$\mathbf{R}'\mathbf{A}\mathbf{R} = \mathbf{\Lambda} \quad \text{diagonális},$$

$$\mathbf{R}'\mathbf{B}\mathbf{R} = \mathbf{I},$$

$$\mathbf{A}\mathbf{R}_i = \lambda_i \mathbf{B}\mathbf{R}_i,$$

ahol $\mathbf{R}_i$ az $\mathbf{R}$ mátrix i-edik oszlopát jelöli, λ_i pedig a $\mathbf{\Lambda}$ i-edik diagonális eleme.

Bizonyítás. Közismert tény, hogy $\mathbf{B}$-hez található olyan $\mathbf{P}_0$ mátrix, hogy

$$\mathbf{P}_0'\mathbf{B}\mathbf{P}_0 = \mathbf{\Lambda}_0 \quad \text{diagonális}$$

és

$$\mathbf{P}_0'\mathbf{P}_0 = \mathbf{I}$$

(ahogy ezt a főkomponens analízisnél is csináljuk). Legyen $\mathbf{T}=\mathbf{\Lambda}_0^{1/2}\mathbf{P}_0'$, ahol $\mathbf{\Lambda}_0^{1/2}$ egy olyan diagonális mátrix, melynek az elemei a $\mathbf{\Lambda}_0$ diagonálisában álló (nem negatív) számok négyzetgyökei. Ekkor

$$\mathbf{B} = \mathbf{P}_0\mathbf{\Lambda}_0\mathbf{P}_0' = \mathbf{P}_0\mathbf{\Lambda}_0^{1/2}\mathbf{\Lambda}_0^{1/2}\mathbf{P}_0' = \mathbf{T}'\mathbf{T},$$

$$\mathbf{T}^{-1} = \mathbf{P}_0\mathbf{\Lambda}_0^{-1/2}.$$

Tekintsük a következő sajátérték problémát:

$$|\mathbf{T}'^{-1}\mathbf{A}\mathbf{T}^{-1} - \lambda\mathbf{I}| = 0,$$

és vegyük észre, hogy itt már a $\mathbf{C}=\mathbf{T}'^{-1}\mathbf{A}\mathbf{T}^{-1}$ szimmetrikus mátrix, hiszen (mivel $\mathbf{A}$ szimmetrikus) bármely $m\times m$-es $\mathbf{D}$ mátrix esetén $\mathbf{D}'\mathbf{A}\mathbf{D}$ szimmetrikus. Mint előbb $\mathbf{B}$-hez, most $\mathbf{C}$-hez is található olyan $m\times m$-es mátrix, hogy

$$\mathbf{P}'\mathbf{C}\mathbf{P} = \mathbf{\Lambda} \quad \text{diagonális},$$

és

$$\mathbf{P}'\mathbf{P} = \mathbf{I}.$$

Jelölje $\mathbf{R}=\mathbf{T}^{-1}\mathbf{P}$. Ekkor $\mathbf{R}'\mathbf{A}\mathbf{R}=\mathbf{\Lambda}$, hiszen

$$\mathbf{R}'\mathbf{A}\mathbf{R} = \mathbf{P}'\mathbf{T}^{-1\prime}\mathbf{A}\mathbf{T}^{-1}\mathbf{P} = \mathbf{P}'\mathbf{C}\mathbf{P} = \mathbf{\Lambda}.$$

Ugyancsak igaz, hogy

$$\mathbf{R}'\mathbf{B}\mathbf{R} = \mathbf{R}'\mathbf{T}'\mathbf{T}\mathbf{R} = \mathbf{P}'\mathbf{P} = \mathbf{I}.$$

Most már csak az $\mathbf{A}\mathbf{R}_i=\lambda_i\mathbf{B}\mathbf{R}_i$ összefüggést kell igazolni. Jelölje $\mathbf{P}_i$ a $\mathbf{P}$ mátrix i-edik oszlopát. Mivel

$$\mathbf{C}\mathbf{P}_i = \lambda_i\mathbf{P}_i,$$

így

$$\mathbf{A}\mathbf{R}_i = \mathbf{T}'\mathbf{C}\mathbf{T}\mathbf{R}_i = \mathbf{T}'\mathbf{C}\mathbf{P}_i = \mathbf{T}'\lambda_i\mathbf{P}_i = \lambda_i\mathbf{T}'\mathbf{P}_i = \lambda_i\mathbf{T}'\mathbf{T}\mathbf{R}_i = \lambda_i\mathbf{B}\mathbf{R}_i.$$

Ezzel az állítást bebizonyítottuk.

Speciálisan a (4.1) feladatban $\mathbf{A}=\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}$, $\mathbf{B}=\mathbf{\Sigma}_{22}$, $m=\operatorname{rang}(\mathbf{\Sigma}_{22})=p-q$. Ekkor a

$$|\mathbf{C}-\lambda\mathbf{I}| = 0$$

sajátfeladatot kell megoldani, ahol

$$\mathbf{C} = \mathbf{T}'^{-1}\mathbf{A}\mathbf{T}^{-1} = \mathbf{\Sigma}_{22}^{-1/2\prime}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1/2},$$

ti.: $\mathbf{B}=\mathbf{\Sigma}_{22}=(\mathbf{\Sigma}_{22}^{1/2})'(\mathbf{\Sigma}_{22})^{1/2}$ és

$$\mathbf{R}_i = \mathbf{T}^{-1}\mathbf{P}_i = \mathbf{\Sigma}_{22}^{-1/2}\mathbf{P}_i,$$

ahol $\mathbf{P}_i$ a $\mathbf{C}$ szimmetrikus mátrix i-edik sajátvektora.

5. Kovarianciamagyarázat

A faktoranalízis témájában ismertek a főkomponensek (főfaktorok) optimizáló tulajdonságai. Emlékeztetőül ezek közül kettőt itt is közlünk:

1. Az eredeti változók lineáris kombinációiból álló összes $k\leqq q=\operatorname{rang}(\mathbf{\Sigma}_{11})$ elemű rendszer közül az első k főkomponens az, melyekre vonatkozóan az eredeti változók lineáris regressziós becsléseinek a kovarianciamátrixa a legjobban közelíti az eredeti változók kovarianciamátrixát. Ebben a részben 2 azonos méretű mátrix távolságát elemeik különbségének euklideszi normájával mérjük.
2. Az első k főkomponens alkotja azt a k elemű rendszert, melyre az eredeti változókból, ezeknek a k elemű rendszerre vonatkozó lineáris regressziós becslései által nem magyarázott rész kovariancia mátrixa minimális (a legkevésbé tér el az azonosan nulla mátrixtól).

Érdemes megvizsgálni, hogy milyen analóg állítások mondhatók ki a kanonikus korrelációanalízissel kapcsolatban. Tegyük fel, hogy az eredeti $\mathbf{U}_1$ és $\mathbf{U}_2$ változók már faktorizálva vannak, azaz $\mathbf{\Sigma}_{11}=\mathbf{I}_{q\times q}$ és $\mathbf{\Sigma}_{22}=\mathbf{I}_{(p-q)\times(p-q)}$. Itt is igaz, hogy a $k\leqq m=\operatorname{rang}(\mathbf{\Sigma}_{12})$ legnagyobb kanonikus korrelációs együtthatóhoz tartozó (röviden: legnagyobb) jobb és bal oldali kanonikus faktor alkotja azt a rendszert, melyre vonatkozóan az eredeti változók lineáris regressziós becsléseinek a kovarianciamátrixa a legjobban közelíti a $\mathbf{\Sigma}_{12}$ mátrixot. Ezt könnyen beláthatjuk, felhasználva a polárfelbontási tételt [6]:

$$\min_{r(\mathbf{A})=k}\|\mathbf{\Sigma}_{12}-\mathbf{A}\| = \left\|\mathbf{\Sigma}_{12}-\sum_{i=1}^{k}\varrho_i\mathbf{L}_i\mathbf{M}_i'\right\| = \left\{\sum_{i=k+1}^{m}\varrho_i^2\right\}^{1/2}, \tag{5.1}$$

ahol $\|\cdot\|$ a fent említett normát jelöli, $\mathbf{L}_i$, illetve $\mathbf{M}_i$ az i-edik bal, illetve jobb oldali kanonikus együtthatók, ϱ_i az i-edik kanonikus korreláció és az egyszerűség kedvéért most az $\mathbf{A}$ mátrix rangját $r(\mathbf{A})$ jelöli. Ebben a pontban $\mathbf{A}$, $\mathbf{B}$ és $\mathbf{C}$ egy-egy $q\times(p-q)$ típusú mátrixot jelölnek.

Ugyancsak igaz, hogy ezek azok a lineáris kombinációk, amelyekre az eredeti változókból, ezeknek a megfelelő k-asra vonatkozó legkisebb négyzetes lineáris becslései által nem magyarázott részek kovarianciamátrixa a legkisebb.

Bizonyítás. Jelölje $\mathbf{f}_x(\mathbf{Y})$ az $\mathbf{X}$ vektorváltozó komponenseinek az $\mathbf{Y}$-ra vonatkozó lineáris regressziós becsléseiből, mint oszlopvektorból alkotott vektort. Ekkor

$$\min_{r(\mathbf{A})=r(\mathbf{B})=k} \left\|\operatorname{cov}\left(\mathbf{U}_1-\mathbf{f}_{\mathbf{U}_1}(\mathbf{A}'\mathbf{U}_1),\ \mathbf{U}_2-\mathbf{f}_{\mathbf{U}_2}(\mathbf{B}'\mathbf{U}_2)\right)\right\| = M.$$

Nyilván teljesül, hogy:

$$\min_{r(\mathbf{C})=k} \|\boldsymbol{\Sigma}_{12}-\mathbf{C}\| \leqq M \leqq \|\boldsymbol{\Sigma}_{12}-\tilde{\mathbf{L}}_k\tilde{\mathbf{L}}_k'\boldsymbol{\Sigma}_{12}-\boldsymbol{\Sigma}_{12}\tilde{\mathbf{M}}_k\tilde{\mathbf{M}}_k'+$$

$$+\tilde{\mathbf{L}}_k\tilde{\mathbf{L}}_k'\boldsymbol{\Sigma}_{12}\tilde{\mathbf{M}}_k\tilde{\mathbf{M}}_k'\| = \|\boldsymbol{\Sigma}_{12}-\tilde{\mathbf{L}}_k\mathbf{Q}_k\tilde{\mathbf{M}}_k'\|,$$

ahol $\tilde{\mathbf{L}}_k=(\mathbf{L}_1|\mathbf{L}_2|\ldots|\mathbf{L}_k)_{q\times k}$, $\tilde{\mathbf{M}}_k=(\mathbf{M}_1|\mathbf{M}_2|\ldots|\mathbf{M}_k)_{(p-q)\times k}$ és $\mathbf{Q}_k = \begin{pmatrix} \varrho_1 & & 0 \\ & \varrho_2 & \\ & \ddots & \\ 0 & & \varrho_k \end{pmatrix}_{k\times k}$.

Felhasználva az (5.1) összefüggést adódik, hogy

$$M = \|\boldsymbol{\Sigma}_{12}-\tilde{\mathbf{L}}_k\mathbf{Q}_k\tilde{\mathbf{M}}_k'\| = \left\{\sum_{i=k+1}^{m}\varrho_i^2\right\}^{1/2}.$$

Ha az eredeti változókat az ellenkező oldali változókból képzett k darab lineáris kombinációra vonatkozóan becsüljük, akkor a k darab legnagyobb kanonikus faktor általában kevesebbet magyaráz az eredeti változókból (ti.: a fenti 2 értelemben), mintha a saját oldali változók segítségével becsültük volna. Ugyanis könnyen látható, hogy

$$\mathbf{f}_{\mathbf{U}_1}(\tilde{\mathbf{M}}_k'\mathbf{U}_2) = \boldsymbol{\Sigma}_{12}\tilde{\mathbf{M}}_k\tilde{\mathbf{M}}_k'\mathbf{U}_2,$$

$$\mathbf{f}_{\mathbf{U}_2}(\tilde{\mathbf{L}}_k'\mathbf{U}_1) = \boldsymbol{\Sigma}_{21}\tilde{\mathbf{L}}_k\tilde{\mathbf{L}}_k'\mathbf{U}_1$$

és

$$\operatorname{cov}\left(\boldsymbol{\Sigma}_{12}\tilde{\mathbf{M}}_k\tilde{\mathbf{M}}_k'\mathbf{U}_2,\ \boldsymbol{\Sigma}_{21}\tilde{\mathbf{L}}_k\tilde{\mathbf{L}}_k'\mathbf{U}_1\right) =$$

$$= \boldsymbol{\Sigma}_{12}\tilde{\mathbf{M}}_k\tilde{\mathbf{M}}_k'\boldsymbol{\Sigma}_{21}\tilde{\mathbf{L}}_k\tilde{\mathbf{L}}_k'\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{12}\tilde{\mathbf{M}}_k\mathbf{Q}_k\tilde{\mathbf{L}}_k'\boldsymbol{\Sigma}_{12} =$$

$$= \tilde{\mathbf{L}}_k\mathbf{Q}_k\mathbf{Q}_k\mathbf{Q}_k\tilde{\mathbf{M}}_k' = \tilde{\mathbf{L}}_k\mathbf{Q}_k^3\tilde{\mathbf{M}}_k';$$

illetve

$$\operatorname{cov}\left(\mathbf{U}_1-\boldsymbol{\Sigma}_{12}\tilde{\mathbf{M}}_k\tilde{\mathbf{M}}_k'\mathbf{U}_2,\ \mathbf{U}_2-\boldsymbol{\Sigma}_{21}\tilde{\mathbf{L}}_k\tilde{\mathbf{L}}_k'\mathbf{U}_1\right) =$$

$$= \boldsymbol{\Sigma}_{12}-\boldsymbol{\Sigma}_{12}\tilde{\mathbf{M}}_k\tilde{\mathbf{M}}_k'-\tilde{\mathbf{L}}_k\tilde{\mathbf{L}}_k'\boldsymbol{\Sigma}_{12}+\boldsymbol{\Sigma}_{12}\tilde{\mathbf{M}}_k\tilde{\mathbf{M}}_k'\boldsymbol{\Sigma}_{21}\tilde{\mathbf{L}}_k\tilde{\mathbf{L}}_k'\boldsymbol{\Sigma}_{12} =$$

$$= \boldsymbol{\Sigma}_{12}-2\tilde{\mathbf{L}}_k\mathbf{Q}_k\tilde{\mathbf{M}}_k'+\tilde{\mathbf{L}}_k\mathbf{Q}_k^3\tilde{\mathbf{M}}_k' =$$

$$= \boldsymbol{\Sigma}_{12}-\tilde{\mathbf{L}}_k(2\mathbf{Q}_k-\mathbf{Q}_k^3)\tilde{\mathbf{M}}_k'.$$

Amennyiben $\varrho_1=\varrho_2=\ldots=\varrho_k=1$ nem teljesül, akkor

$$\mathbf{L}_k\mathbf{Q}_k^3\tilde{\mathbf{M}}_k' \neq \tilde{\mathbf{L}}_k\mathbf{Q}_k\tilde{\mathbf{M}}_k'$$

és

$$\tilde{\mathbf{L}}_k(2\mathbf{Q}_k-\mathbf{Q}_k^3)\tilde{\mathbf{M}}_k' \neq \tilde{\mathbf{L}}_k\mathbf{Q}_k\tilde{\mathbf{M}}_k'.$$

Köszönetnyilvánítás

Köszönetet mondok KRÁMLI ANDRÁSnak és TUSNÁDY GÁBORnak a kovariancia-magyarázat problémájának a felvetéséért és az (5.1) összefüggéssel való megismertetéséért. Az itt közölt cikk az MTA Matematikai Kutató Intézet és az MTA SZTAKI közös szemináriumán elhangzott előadás bővített változata.

IRODALOM

[1] ANDERBERG, M. R., *Cluster Analysis for Applications* (Academic Press, New York—London, 1973).
[2] COOLEY, W. W. and LOHNES, P. R., *Multivariate Data Analysis* (John Wiley and Sons, New York, 1971).
[3] LENGYEL, T., „A kanonikus korrelációanalízis alkalmazása szívkoszorúér-megbetegedések előrejelzésére", Számítástechnikai és kibernetikai módszerek alkalmazása az orvostudományban és a biológiában, 8. Neumann Kollokvium, Szeged, 1977, 11—17.
[4] MILLER, J. K., "The development and application of bimultivariate correlation: a measure of statistical association between multivariate measurement sets", Ed. D. Dissertation, Faculty of Educational Studies, State University of New York at Buffalo, 1969.
[5] STEWART, D. K. and LOVE, W. A., "A general canonical correlation index", *Psychological Bulletin* **70** (1968) 160—163.
[6] TUSNÁDY, G., „Mátrixok szinguláris felbontása", *Alkalmazott Matematikai Lapok* **5** (1979).
[7] RAO, C. R., *Linear Statistical Inference and Its Applications* (John Wiley and Sons, New York, 1965).

(Beérkezett: 1979. október 20.)

LENGYEL TAMÁS
MTA SZTAKI
1132 BUDAPEST, VICTOR HUGO U. 18—22.

THE CANONICAL CORRELATION ANALYSIS AND SOME RELATED PROBLEMS

T. LENGYEL

We deal with some optimizing properties of canonical correlation coefficients and canonical factors.

A külföldi szakirodalomból

MI IS AZ A SZÁMÍTÓGÉPES KÍSÉRLET?[1]

A. A. SZAMARSZKIJ
Moszkva

A „számítógépes kísérlet" kifejezés paradoxnak tűnhet. Sok olvasó joggal gondolhatja, kísérletet akkor kell végezni, ha egy bonyolult fizikai jelenség nem követhető számítással, ha pedig a számolás lehetséges, a kísérlet fölösleges. Tehát vagy számítás — vagy kísérlet.

Bár a gondolatmenet alapjában véve helyes, kiindulási pontja mégis megkérdőjelezhető. Azt tételezi fel, hogy minden fizikai jelenség kísérleti úton tanulmányozható vagy elméletileg számítható. Sajnos, nem minden bennünket érdeklő fizikai jelenség váltja be ezeket a reményeket. Előfordul, hogy a kísérlet nem végezhető el, mert bonyolult, drága és kockázatos, az ismert számítási módszerek viszont nem írják le a jelenséget a szükséges pontossággal.

Épp ez történt, amikor az emberiség az atomenergia meghódítására vállalkozott. A nukleáris fűtőanyaggal végzendő kísérlet katasztrofális robbanás kockázatával járt volna, míg a klasszikus matematika a felmerülő feladatok megoldásához gyengének bizonyult.

Próbáljuk meg alaposabban megérteni, miért elégtelenek a matematika klasszikus módszerei bizonyos fizikai jelenségek leírására, milyen új lehetőségeket adnak az elektronikus számítógépek, mi a számítógépes kísérlet, és segítségével hogyan oldhatók meg egyes aktuális tudományos-műszaki problémák.

A matematikai modell

A materialista világnézet egyik alaptétele az, hogy a természeti jelenségek bonyolultsága kimeríthetetlen, egy vizsgálat során az összes tényező nem vehető figyelembe. Ezért a kutató legelőször megkísérli kiválasztani azokat a tényezőket, amelyek a tanulmányozandó jelenségben a kitűzött feladat szempontjából a leglényegesebbek, a lényegteleneket pedig figyelmen kívül hagyja.

GALILEI óta egy fizikai jelenség leírását akkor tekintjük megbízhatónak, ha összetevőit számszerű mennyiségek fejezik ki. Ezeknek egy része közvetlenül mérhető, míg a többiek meghatározásához a természet törvényeit használjuk fel, melyek e mennyiségek közötti összefüggéseket fejezik ki.

Például a mechanika törvényei lehetővé teszik, hogy két tömegpont kezdeti állapotából meghatározzuk, milyen viszonylagos helyzetben lesznek bármely későbbi

[1] Ez a dolgozat A. A. Самарский: Что такое вычислительный эксперимент, Наука на марше dolgozatának fordítása. A fordítás közléséhez a szerző hozzájárult.

időpontban; vagy a hővezetés törvényei alapján, ismerve a hőmérséklet eloszlását egy test felületén, kiszámíthatjuk bármely belső pont hőmérsékletét.

A természet törvényeinek segítségével a tanulmányozandó jelenségben szereplő mennyiségek összefüggéseit egyenletek alakjában fogalmazzuk meg. Ezek rendszerint differenciál- vagy integrál-, integro-differenciál-, algebrai vagy másfajta egyenletek.

A kapott egyenletrendszert a megoldáshoz szükséges ismert adatokkal együtt (kezdeti és peremfeltételek, az egyenlet együtthatóinak értéke stb.) a jelenség matematikai modelljének nevezzük.

A bonyolultság Szküllája és a megbízhatatlanság Kharübdisze között

A kutató egyik legfőbb gondja a jelenség matematikai modelljének felállításakor az, vajon megoldhatók-e a kapott egyenletek. Nem egyszerűsíthetők-e valahogy, nem hagyható-e el valamelyik tag, hogy ily módon egyszerűbb módszerekkel lehessen megoldani. Minden ilyen egyszerűsítés ekvivalens egy a tanulmányozandó jelenség jellegével kapcsolatos feltételezéssel.

Vizsgáljuk például a földre leeső kő mozgását! A követ a Föld vonzóereje készteti mozgásra. Ahogy a kő gyorsul, mozgását egyre inkább befolyásolja a levegő ellenállása, mely a kő sebességével négyzetesen arányos. Elhanyagolható-e ez az erő, mondhatjuk-e, hogy a kő esését csupán a Föld vonzóereje határozza meg? Igen, megtehetjük, ha az előbbi erő elegendően kicsiny az utóbbihoz képest, vagyis ha a kő elegendően kicsiny magasságból esik, és nincs ideje, hogy nagy sebességre tegyen szert.

Miközben a jelenség matematikai modelljét egyszerűsítjük, egyúttal kijelöljük alkalmazhatóságának határait is. Következtetéseinket e határokon túl nem terjeszthetjük ki, megfeledkezve róluk, értelmetlen eredményeket kapunk.

Vegyük a világmindenség „hőhalálának" hipotézisét! Eszerint idővel a világmindenségben beáll a hőegyensúly (hasonlóan ahhoz, ahogy az bekövetkezik minden korlátos, izolált térfogatú anyagban), és megszűnik minden folyamat, beleértve az életet is. A hipotézis tévedése abban rejlik, hogy azokat a következtetéseket, melyeket a zárt térfogatokra vonatkozóan kaptunk, átvittük a világűr korlátlan terére, és eközben nem vettük figyelembe az anyag kölcsönös nehézkedési erőit, azok pedig a világmindenség evolúciójában igen fontosak.

Az elmondottak megértethetik, hogy a leírt klasszikus úton haladó kutató állandóan a bonyolultság *Szküllája* és a megbízhatatlanság *Kharübdisze* között evez. Felállított modelljének egyrészt matematikai szempontból elég egyszerűnek kell lennie, hogy alaposan tanulmányozható legyen a meglevő eszközökkel. Másrészt viszont, az egyszerűsítések során meg kell őrizni a modell „racionális magvát", a probléma lényegét. A matematikai modell felállítása különleges művészet, melyben összefonódik az elmélet ismerete, a tapasztalat és az intuíció.

A kísérlet lehetetlen, a számítás erőtlen

A fizikai jelenségek matematikai modelljeinek tanulmányozásával foglalkozik a matematikai fizika, a matematika egyik jelentős ága. A matematikai fizika elvégezte nagyon sok természeti folyamat egyenletének mély, analitikus vizsgálatát, olyanokét,

mint a bolygómozgás, a folyadékáramlás, a rugalmas alakváltozások és a hullámterjedés, a hővezetés és a diffúzió stb.

Már említettük, hogy minden elmélet alkalmazhatósági területe korlátozott. A matematikai fizika klasszikus módszereinek, amiket főként a múlt század és a századelő matematikusai dolgoztak ki, szintén megvannak a maguk határai. Éppen ezért a gyakorlatban felvetődő mind bonyolultabb feladatok elegendően pontos megoldására irányuló kísérletek idővel egyre gyakrabban vezettek olyan nehézségekhez, amelyeket a meglevő módszerekkel nem sikerült leküzdeni.

Ha valamely jelenség megértéséhez szükséges mennyiségeket vagy azok összefüggéseit nem lehet kiszámítani, akkor mérésekkel, kísérletileg kell meghatározni őket. Azonban napjainkban az élet olyan problémákat szül, amelyeknek kísérleti tanulmányozása különös nehézségekkel jár, gyakran nem is veszélytelen. A nukleáris energiatermelés feladatai mellett említhetjük a kozmikus tér meghódításának nem kevésbé bonyolult feladatait. Ökológiai szempontból kockázatosak az éghajlatszabályozási kísérletek, indokolt az óvatosság a szociális kísérletekben, és tilos a kísérletezés az emberi egészséggel.

Az ezekhez hasonló helyzetekben egyetlen lehetőség marad: olyan eszközök és módszerek kidolgozása, amelyek lehetővé teszik a tanulmányozandó jelenség kiszámítását bármilyen előírt pontossággal.

Ilyen eszközök az elektronikus számítógépek, ilyen módszerek pedig a numerikus matematika módszerei vagy másképpen a numerikus módszerek.

Meg kell jegyeznünk, hogy numerikus módszerek már jóval az elektronikus számítógépek megjelenése előtt léteztek, és használták is őket fizikai jelenségek kiszámítására, különböző szerkezetek tervezésére. Azonban arra, hogy segítségükkel, kézi számolással vagy mechanikus számológépek felhasználásával kielégítő pontosságot érjenek el, óriási időt kellett fordítani. Elfogadható idő alatt csak olyan eredményeket lehetett kapni, amelyekből csak nagyjából következtethettek a jelenség lefolyására vagy a készülő gép viselkedésére.

Hővezetés. A differenciálegyenlet

Képzeljék el, hogy kezükben egy hosszú, fém kötőtű egyik végét fogják, a másik végét pedig a gáztűzhely lángjába dugják. Rövid idő múlva a kísérletet meg kell szakítaniuk, a fém vezeti a hőt, és az a kötőtűben a melegített végtől terjedve elviselhetetlen hőmérsékletre melegíti a másik véget is, amit éppen kezükben tartanak.

Próbáljuk meg az egyszerű fizikai kísérlet matematikai leírását! Állítsuk fel a hő terjedésének egyenletét a rúdban (így nevezi a matematikai fizika a kötőtűhöz hasonló tárgyakat)! Gondolatban osszuk fel hosszában az egész rudat nagyszámú apró darabkára. A fizika egyik alaptörvénye, az energia megmaradásának törvénye azt sugallja, hogy minden egyes darabkában a hőmennyiség változását az határozza meg, mennyi hő érkezett az adott darabkába a szomszédos darabkából, és mennyi távozott a másik szomszédos darabkába, vagyis (ahogyan a fizikusok mondják) mekkora a hőáramok különbsége. Az egyszerűség kedvéért elhanyagoljuk azt a hőt, ami a tűt körülvevő levegőbe távozik. A darabkában „bennmaradó" hő felmelegíti azt.

Így egy viszonylag egyszerű algebrai egyenletet kapunk, amiben szerepel a hőáramok különbsége a darabka végein, a két időpontnak megfelelő hőmennyiség

különbsége, e két időpont különbsége, továbbá a darabka hossza, vagyis a jobb oldali és bal oldali vég koordinátájának különbsége.

E pontnál ér véget a fizika, és kezdődik a matematika. Az egyenletet oly módon alakítjuk át, hogy az említett különbségek hányadosai jelenjenek meg, mégpedig úgy, hogy a számlálóban álljanak a keresett függvények értékeinek különbségei (a hőmennyiségek és a hőáramokéi), a nevezőben meg az időértékeknek és a rúd koordinátáinak különbségei, vagyis olyan mennyiségeké, melyek a keresett függvények független változói.

Ezután az összes különbséggel zérushoz tartunk az egyenletben. A hányadosok helyett azok határértékei jelennek meg, ezeket deriváltaknak nevezzük. A keresett függvények deriváltjait tartalmazó egyenleteket pedig differenciálegyenleteknek hívjuk.

Példánkban az ismertetett eljárás során az ún. hővezetési egyenletet kapjuk. Ez a hővezető rúd melegedését és hűlését írja le.

Azonban ez a hővezetési egyenlet pusztán első közelítés. Hiszen elhanyagoltuk a hő elfolyását a környezetbe, ez pedig gyakran igen lényeges (ha, mondjuk, nem a kötőtűt, hanem egy radiátort nézünk). Az anyag hővezetési tulajdonságai függhetnek a hőmérséklettől is (ez a körülmény döntő szerephez jut a termonukleáris plazmareaktorban).

A vizsgálat céljától és a kívánt pontosságtól függően bonyolódik a jelenség matematikai modellje, amely már olyan tényezőket is figyelembe vesz, amiket az első közelítésben elhagytunk.

Hővezetés. A differenciaséma

Beszéljünk most a numerikus módszerekről! Tegyük fel, hogy nem a hővezetés általános egyenlete érdekel bennünket, hanem konkrét feladatunk megoldása, a rúd melegedésének leírása, méghozzá számokban kifejezve, ahogyan ezt a gyakorlat is megkívánja.

Álljunk meg előző gondolatmenetünkben annál a lépésnél, amikor a kicsiny rúddarab melegedését vizsgáltuk, és a darabkában levő hőmennyiségre, valamint a végein átáramló hőre vonatkozó algebrai egyenletet felírtuk. Ilyen egyenletet annyit írhatunk fel, ahány darabra a rudat felosztottuk. Ezek az egyenletek „egymáshoz vannak láncolva", ugyanis az egyik darabkából kilépő hő belép a másikba. Tehát nem független egyenleteink vannak, hanem egy algebrai egyenletrendszerünk, amit az algebra hagyományos módszereivel meg is tudunk oldani.

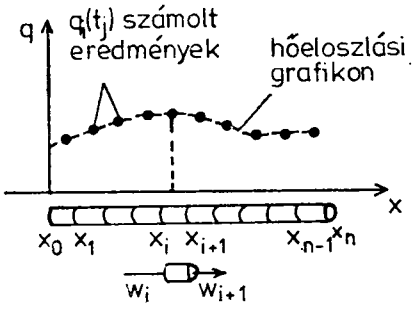
Világos, hogy ha ezt az egyenletrendszert megoldjuk, csak korlátozott számú pontban és csak bizonyos időpontokban kapjuk meg a hőmérsékletet.

Minél több pontunk van, vagyis minél apróbbra szabdaljuk a rudat, rendszerint annál pontosabb a közelítő megoldás. Így azonban nő a megoldandó rendszer egyenleteinek száma is. Itt aztán szükségessé válik az elektronikus számítógép. De még a legnagyobb számítógép lehetőségei is korlátosak. Így a darabkák hossza, amikre a rudat felosztottuk, nem csökkenhet korlátlanul, hanem véges marad. Innen kapta a nevét az ismertetett módszer is: véges differenciák módszere.

Az argumentumok azon értékeinek halmazát, amelyekben a keresett függvény értékeit meghatározzuk, differenciarácsnak nevezzük. (Példánkban a rácsot az határozza meg, hogyan osztjuk fel a rudat, és milyen időpontokban akarjuk kiszámí-

tani az állapotát.) Azt az algebrai egyenletrendszert, ami arra szolgál, hogy a jelenséget a véges differenciamódszerrel számítsuk ki, differenciasémának nevezzük.

Konkrét példánkban a rúddal, amikor a differenciaséma fogalmát magyaráztuk, a szemléletesség kedvéért közvetlenül magából a jelenségből indultunk ki. Lehetséges, hogy ez a bonyolult fogalom így világossabbá vált, de helytelen lenne azt gondolnunk, hogy ténylegesen így állítják fel a differenciasémákat.



1. ábra
A hővezetés differenciálegyenletének levezetése a hővezető rúd egy darabkájára felírt hőegyensúlyból

q — fajlagos hőmennyiség t — idő	w — hőáram x — koordináta (rúd mentén mérve)
hőegyensúly-egyenlet az i-edik darabkában a j-edik időpontban	$[q_i(t_{j+1})-q_i(t_j)](x_{i+1}-x_i) =$ $= -(w_{i+1}-w_i)(t_{j+1}-t_j)$
differenciaegyenlet	$\frac{q_i(t_{j+1})-q_i(t_j)}{t_{j+1}-t_j} = -\frac{w_{i+1}-w_i}{x_{i+1}-x_i} \quad (t_{j+1}-t_j) \to 0 \quad (x_{i+1}-x_i) \to 0$
differenciálegyenlet	$\frac{\partial q}{\partial t} = -\frac{\partial w}{\partial x}$
hővezetési egyenlet	$\frac{\partial}{\partial t}(\varrho c T) = \frac{\partial}{\partial x}\left(\lambda \frac{\partial T}{\partial x}\right)$ sűrűség, fajhő, hővezetési együttható

A felállítás módszereinek leírásánál a hővezetés differenciálegyenletéből kellett volna kiindulnunk, és meg kellett volna mutatnunk, hogyan vezethető le a differenciálegyenlet-rendszer, ami a rúd melegedésének feladatát oldja meg. Ezután azt kellett volna mondanunk, hogy a gyakorlatban, amikor a kutató egy jelenség kiszámítására szolgáló differenciasémát állít fel, nem magából a jelenségből indul ki, hanem annak matematikai modelljéből.

A numerikus és a klasszikus matematika

Fontos észrevételt kell itt tennünk. Amikor a rudat felosztottuk, a figyelmes olvasó bizonyára megjegyezte, hogy a felosztást elég sokféleképpen végezhetjük. Más szóval a differenciarács megválasztása legkevésbé sem egyértelmű. Még inkább érvényes ez a feladat megoldásához használt differenciasémára.

De ha sok séma van, akkor ezek között lenniük kell valamilyen szempontból jobbaknak, például olyanoknak, amelyek a megadott pontosság eléréséhez minimális számolást igényelnek. A differenciasémák elemzése és összehasonlítása, a kitűzött célnak legjobban megfelelő séma kiválasztása — művészet. Ehhez szükségünk van a numerikus módszerek elméletén kívül fizikai meggondolásokra; ha pedig az elmélet cserbenhagy, heurisztikus fogásokat és általánosabb elveket egyaránt használnunk kell.

A mai numerikus matematika bővében van olyan elveknek, amelyekre épülve sikeresen fejlődhet a numerikus módszerek elmélete.

De hamis úton járnánk, ha a numerikus matematikát, sajátosságait hangsúlyozva, a matematika többi ágától elszakítva, mi több, velük szembeállítva vizsgálnánk. A numerikus módszerek elmélete használja a lineáris algebra, a funkcionálanalízis eredményeit, sőt az egész klasszikus, „tiszta" matematikát. Ugyanakkor, a „tiszta" matematika is gazdagodik a numerikussal folyó együttműködés révén. Például a fizikai problémák numerikus megoldása újabb kutatásokra ösztönzött a hiperbolikus egyenletek, a szakadásos együtthatójú parabolikus és elliptikus egyenletek, a nemlineáris egyenletek körében. Az utóbbi években több jelentős dolgozat látott napvilágot e kérdésekről.

A számítógépes kísérlet

Nézzük most meg, hogyan működnek a numerikus módszerek, hogyan számíthatók ki segítségükkel a fizikai jelenségek!

Minden úgy kezdődik, mint a klasszikus változatban, tehát a matematikai modell felállításával, annak tisztázásával, mely tényezőket tükrözze, melyeket hagyja figyelmen kívül a modell. Itt a döntő szó a fizikusoké. A matematikusok arra törekszenek, hogy felállított egyenleteik a lehető legalkalmasabbak legyenek a gépi számolásra.

A matematikai modell felállításával több kérdés jár együtt. Matematikai szempontból értelmes-e a jelenséget leíró egyenletrendszer? Van-e megoldása, s ha igen, egyetlen-e? Van-e a rendszernek pontos analitikus megoldása bizonyos speciális esetekben? (Ez utóbbi igen fontos: Ha elvégezzük a számolást ezekben az esetekben és összehasonlítjuk az eredményt a pontos analitikus megoldással, képet alkothatunk számítási módszerünk pontosságáról.)

Ezután kidolgozzuk a feladat megoldásának algoritmusát. Ez aritmetikai és logikai műveletek sorozata, amit a számítógép számára ún. program valósít meg. E program szerint kerül sor a számításokra.

Ha a vizsgálandó jelenség bizonyos változatai kísérletileg is tanulmányozhatók, ezek számítása különös jelentőségre tesz szert. Ha a kutató összeveti a számítás eredményeit a kísérleti adatokkal, képet kaphat a matematikai modell megbízhatóságáról, megbecsülheti alkalmazhatóságának határait. Kiderülhet, hogy a modell

nem elég pontos vagy nem elég teljes, pontosításra, kiegészítésre szorul, olyan tényezőket is tükröznie kell, amelyeket kezdetben jogtalanul félredobtunk. Előfordulhat, hogy a modell túl bonyolult, és ugyanazok az eredmények egyszerűbb modell segítségével is elérhetők. Mindez a jobb modell felállítását segíti elő.

Végül eljutunk oda, hogy elfogadjuk a jelenség matematikai modelljének egyik változatát. A feladat különböző paramétereit (a perem- és kezdeti feltételeket, az egyenletek együtthatóinak értékét) variálva a modell keretein belül behatóan tanulmányozhatjuk a fizikai folyamatot: kideríthetjük a fő törvényszerűségeket, megbecsülhetjük különféle tényezők hatását, egyszóval összegyűjthetjük mindazokat az információkat, amiket a fizikai kísérlet során is megszerezhetünk.

A dolog lényegét tekintve ez a munka nagyon közel áll a kísérlethez, csak kísérleti berendezés helyett számítógépet használunk, a fizikai jelenség helyett pedig annak matematikai modelljét alkalmazzuk.

Ezért kapta a számítógépes (numerikus vagy matematikai) kísérlet elnevezést a fizikai jelenség számításának fentebb ismertetett módszere.

A számítógépes kísérlet előnyei

A számítógépes kísérlet lehetőségei jóval nagyobbak, mint a fizikai kísérletei.

A fizikai modellkísérlet előkészítéséhez és elvégzéséhez sok időre és eszközre van szükség. Minden egyes kísérlethez külön mérőberendezést és mérési metodikát kell kidolgozni. Ha kiderül, hogy a berendezés nem teszi lehetővé a tanulmányozandó jelenség valamely aspektusának vizsgálatát, akkor újat kell építeni. Egyébként ugyanígy, ha nem tudjuk jó előre figyelembe venni az összes fontos alkatrészt, a tervezési feladat is elbonyolódik: az első tervezettől az utolsó változatig a kísérleti modellek hosszú lánca vezet.

A számítógépes kísérlet olcsóbb, gyorsabb, egyszerűbb és könnyen irányítható. Különösebb nehézség nélkül is beleavatkozhatunk. Olyan körülményeket is modellezhetünk, amelyeket laboratóriumban még nem tudunk megteremteni.

A matematikai vizsgálat klasszikus módszereivel sok fizikai jelenséget csak minőségileg írhatunk le, és csupán egyes feladatokat oldhatunk meg pontosan. Ugyanakkor a számítógépes kísérlet megnyitja az utat nagy, komplex problémák megoldásához, a műszaki konstrukciók optimális számításához, a kutatás tudományosan megalapozott tervezéséhez.

Még egy szempontot kell említenünk, ami miatt a számítógépes kísérlet előnyösebb a természetinél. A mai fizikai és műszaki problémák változatossága ellenére matematikai leírásuk korlátozott számú egyenletre (pontosabban egyenlettípusra) vezet. Például a diffúziós, a hővezetési és a felmágnesezési folyamatok formálisan ugyanazzal az egyenlettel írhatók le. Hasonló egyenletek fejezik ki a torziónak kitett rugalmas rúd feszültségi állapotát, a folyadékáramlást, dielektrikumban az elektromos tér eloszlását. Különbség mindössze az egyenletben szereplő mennyiségek fizikai jelentésében van.

Ezért az egyik feladat megoldásának numerikus módszerei könnyen átvihetők mások megoldására. Ugyanakkor nem állíthatjuk, hogy az egyik fizikai jelenség tanulmányozására készült kísérleti berendezés épp oly könnyedén átépíthető lenne mások vizsgálatára.

Természetesen a számítógépes kísérletnek is vannak hátrányai. A legfőbb

közülük az, hogy a számítások eredményeinek alkalmazását behatárolja a matematikai modell, a modell a már ismert fizikai törvényszerűségekre épül, azokat pedig kísérleti úton állapítják meg. Ezért a számítógépes kísérlet sohasem szoríthatja ki a természetit. A jövő ésszerű összekapcsolásuké.

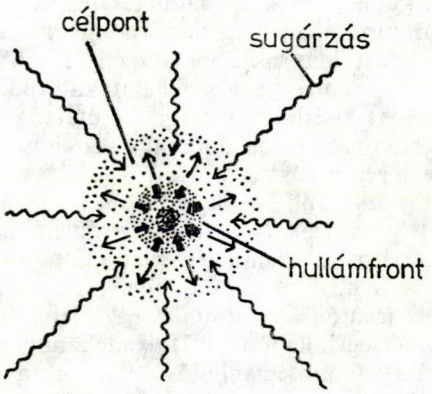
Úton a lézeres termonukleáris reaktor felé

Nézzünk most egy olyan példát, ami illusztrálja a numerikus kísérletek alkalmazásának lehetőségeit aktuális tudományos-technikai feladatok megoldásában.

A manapság annyit emlegetett energiaprobléma radikális megoldási módja a termonukleáris fúzió. A legmegfelelőbb fűtőanyagnak a hidrogén nehézizotópjainak, a deutériumnak és a triciumnak a keverékét tartják. A termonukleáris reakció megindításához ezt a keveréket több száz milliónyi fokra kell felhevíteni. Ilyen körülmények laboratóriumi megteremtése számos technikai nehézségbe ütközik.

Az optikai kvantumgenerátorok, a lézerek megteremtése új utat nyitott a kutatók előtt. A következő ötlet vetődött fel: a lézersugarat kicsiny termonukleáris célpontra kell fókuszálni (a deutériumból és triciumból álló gömböcske sugara 0,1 mm nagyságrendű), és rövid idő ($10^{-9}-10^{-10}$ sec) alatt nagy energiát kell „beléfektetni". Hővé alakulva a lézersugárzás magas hőmérsékletet hoz létre, a tehetetlenség pedig nem engedi, hogy ugyanilyen kicsiny idő alatt az anyag lényegesen kiterjedjen. Ez biztosíthatja a körülményeket a termonukleáris fűtőanyag „begyulladásához".

Sajnos, a meglevő lézerberendezések kapacitása még nem elegendő ahhoz, hogy a lézeres termonukleáris fúziót laboratóriumban megvalósítsuk.



2. ábra

A „lézeres termonukleáris reakciót" magyarázó ábra. A deutériumból és triciumból álló gömbbe több irányból erős lézersugarat bocsátanak. A gömb párologni kezd és a gőzök a térbe sugárzódnak. E sajátos anyagleadás következtében létrejövő hullám a gömb közepéhez tart, magját összenyomja és magas hőmérsékletre hevíti. Ez a hőmérséklet már elegendő a deutérium és a tricium közti termonukleáris reakció beindulásához

És mégis, a tudósok már ma kísérleteznek, lézersugarakat irányítanak a célpontokra, figyelik a termonukleáris mikrorobbanás közben lejátszódó folyamatokat, rögzítik az anyag sűrűsödését és ritkulását, mérik a hőmérsékletet, meghatározzák a rendszer hatásfokát. A kapott adatok alapján választják ki az optimális tulajdonságú (energiájú, impulzushosszúságú, kisugárzási frekvenciájú) lézereket. Igaz, a kísérletet nem valódi anyaggal végzik, hanem számokkal, egyenletekkel, amelyek a lézeres termonukleáris reakciót írják le. Ezek számítógépes kísérletek. Sok tudományos kollektívában, köztük a Szovjetunió Tudományos Akadémiájának M. V. KELDIS nevét viselő Alkalmazott Matematikai Intézetében is foglalkoznak velük.

Az elektronikus számítógép a találmány társszerzője

A Szovjetunió Minisztertanácsa mellett működő Újítási és Találmányi Bizottság állami nyilvántartásában az 55-ös számot egy új fizikai jelenség, a *T*-effektus viseli.

A felfedezés tulajdonjogát a Szovjet Tudományos Akadémia Lenin renddel kitüntetett Keldis Alkalmazott Matematikai Intézetének munkatársai, A. N. TYIHONOV és A. A. SZAMARSZKIJ akadémikusok, P. P. VOLOSZEVICS, L. M. DJEGTARJOV, SZ. P. KURDJUMOV, JU. P. POPOV, A. P. FAVORSZKIJ, a fizikai-matematikai tudományok kandidátusai, továbbá L. A. ZAKLJAZMINSZKIJ a műszaki tudományok doktora és V. SZOKOLOV a fizikai-matematikai tudományok doktora kapta.

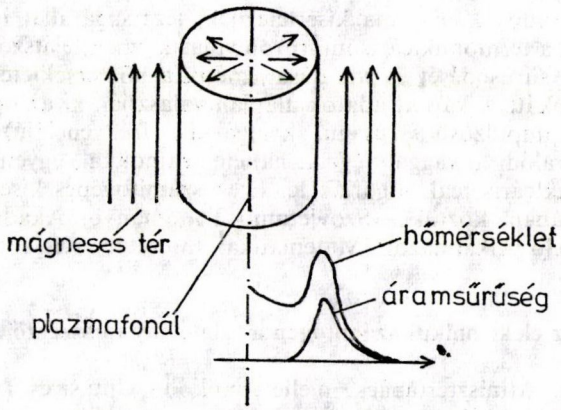
A felfedezett effektus lényege az, hogy a mágneses térrel kölcsönhatásban levő plazmában megfelelő körülmények között viszonylag magas hőmérsékletű zónák jöhetnek létre. Ezekben a zónákban, a *T*-rétegekben (hőrétegekben) elektromos áramok sűrűsödnek, amelyek felhevítik a plazmát és magas hőmérsékletet tartanak fenn.

A *T*-réteg effektus különböző, magneto-hidrodinamikai (MHD) elvek alapján működő berendezésekben használható. Az MHD-energiagenerátorok jól illusztrálják az alkalmazást. A működési elv egyszerűsítve a következő: A plazma „áthalad" a mágneses téren. A forró plazma vezető, és amikor átszeli a mágneses erővonalakat, elektromos áram indukálódik benne. A plazma hőenergiája közvetlenül elektromos energiává alakul, hiányoznak a közbülső állomások, mint például a turbógenerátor turbinája, ami menthetetlenül energiaveszteséggel, a hatásfok csökkenésével járna.

Az MHD-generátorok konstrukciója egyébként komoly nehézségekbe ütközik. Egyrészt a plazma hőmérsékletének minél magasabbnak kell lennie, mert csak így biztosítható intenzív kölcsönhatása a mágneses térrel és így lesz nagy a hatásfok. Másrészt, az MHD-generátorban a plazmának elég hidegnek kell lennie, különben még a legmagasabb olvadáspontú anyag sem maradna meg a környezetében és a berendezés összeroppanna.

Ezen segíthet a *T*-réteg effektus. Képzeljük el, hogy az MHD-generátor csatornájába nem felhevített plazmaáram jut, hanem magas hőmérsékletű *T*-rétegekkel „megspékelt", de viszonylag hideg gáz. A *T*-rétegek, amelyekben az elektromos áram keletkezik, gyorsan áthaladnak a berendezés csatornáján, és nem képesek jelentősen felmelegíteni azt. Ugyanakkor a gáz hideg rétegeinek energiája sem vész el, azok lökik keresztül a *T*-rétegeket a mágneses téren, így hasznos munkát végeznek, ami a továbbiakban elektromossággá alakul.

A *T*-rétegeket számítógépes kísérletekkel fedezték fel. Nagy pontossággal sike-



3. ábra

T-réteg jelenség léphet fel abban a folyamatban is, amelynek diagramja az ábrán látható. A plazmafonal saját nyomásának hatására kiterjed a vákuumban. A mágneses tér a fonal irányába mutat. Mint ismeretes, a plazma ionizált gáz, melyben a molekulák egy része elektronokra és ionokra bomlott. A töltött részecskékre a mágneses térben a *Lorentz-erő* hat, a plazmában áram indukálódik, ami felhevíti azt.

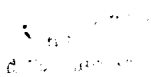
rült leírni azokat a feltételeket, amelyek esetén a korábban ismeretlen effektusnak fel kell lépnie. Akkor még bonyolultságuk miatt nem végeztek olyan fizikai kísérleteket, amelyekben a jelenség bekövetkezhetett volna. Ez kemény vitákhoz vezetett, egyes plazmakutató tudósok hitetlenkedtek; túl szokatlan volt a szituáció: fizikai effektust matematikusok fedeztek fel. Egyébként, miután a matematikusok pontosan leírták az új jelenséget, a fizikusok megpróbálták kísérletileg is előállítani. A próbálkozások sikerre vezettek. Három kutatócsoport Moszkvában, Novoszibirszkben és Szuhumiban egymástól függetlenül, különböző berendezéseken regisztrálta a *T*-réteget.

Ez csupán egy a számítógépes kísérletek sikeres fizikai alkalmazásának számtalan példája közül.

FORDÍTOTTA:

BALLA KATALIN
MTA SZÁMÍTÁSTECHNIKAI ÉS AUTOMATIZÁLÁSI KUTATÓ INTÉZET
1250 BUDAPEST, URI U. 49.

A kiadásért felel az Akadémiai Kiadó igazgatója
Műszaki szerkesztő: Marton Andor
A kézirat nyomdába érkezett:1980. III. 20. Terjedelem: 17,15 (A/5 ív)
80-1341 — Szegedi Nyomda — Felelős vezető: Dobó József igazgató



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 a felelős szerkesztő címére kell beküldeni:

Prékopa András, főszerkesztő, MTA SZTAKI
1502 Budapest XI., Kende u. 13—17.

A kéziratok szerkezeti felépítésének a következő követelményeket kell kielégíteni. A fejlécnek tartalmaznia kell a dolgozat címét, a szerző teljes nevét, valamint annak a városnak a nevét, ahol a szerző dolgozik. A fejléc után egy, képletet nem tartalmazó, legfeljebb 200 szóból álló kivonatot kell minden esetben megadni. A dolgozatot címmel ellátott szakaszokra kell bontani, és az egyes szakaszokat arab sorszámmal kell ellátni. Az esetleges bevezetésnek mindig az első szakaszt kell alkotnia. Az irodalomjegyzék mindig az utolsó szakasz kell hogy legyen, és azt nem kell sorszámmal ellátni. Az irodalomjegyzék után, a kézirat befejezéseképpen fel kell tüntetni a szerző teljes nevét és a munkahelye (illetve lakása) pontos postai címét. A dolgozatban előforduló képleteket szakaszonként újrakezdődően, a képlet előtt két zárójel közé írt kettős számozással kell azonosítani. Természetesen nem szükséges minden képletet számozással ellátni. Az esetleges definíciókat és tételeket (segédtételeket és lemmákat) ugyancsak szakaszonként újrakezdődő, kettős számozással kell ellátni. Kérjük a szerzőket, hogy ezeket, valamint a tételek bizonyítását a szövegben kellő módon emeljék ki. Minden dolgozathoz csatolni kell egy angol, német, francia vagy orosz nyelvű, külön oldalra gépelt összefoglalót. Amennyiben lehetséges, kérjük a nyomtatás számára különösen nehézkes matematikai jelölések használatának az elkerülését.

A dolgozat ábráit és az esetleges lábjegyzeteket a dolgozat végén, különálló lapokon kérjük beküldeni. Mind az ábrákat, mind a lábjegyzeteket a dolgozat szakaszokra bontásától független, folytatólagos arab sorszámozással kell ellátni. Az ábrák elhelyezését a dolgozat megfelelő helyén, széljegyzetként feltüntetett, ábraazonosító sorszámokkal kell megadni. A lábjegyzetekre a dolgozaton belül az azonosító sorszám felső indexkénti használatával lehet hivatkozni.

Az irodalmi hivatkozások formája a következő. Minden hivatkozást fel kell sorolni a dolgozat végén található irodalomjegyzékben, a szerzők, illetve társszerzők esetén az első szerző neve szerinti alfabetikus sorrendben úgy, hogy külön, de folytatólagos sorszámozású listát alkossanak a latin és a cirill betűs nevű szerzők műveire vonatkozó hivatkozások, és mindkét részben a megfelelő alfabetikus sorrend legyen kialakítva. 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 hivatkozni:

[1] Farkas, J., »Über die Theorie der einfachen Ungelichungen«, *Journal für die reine und angewandte Mathematik* **124** (1902) 1—27.

[2] Kéri, G., „DUALSIMP", rutin a CDC 3300-as gépekre (Magyar Tudományos Akadémia Számítástechnikai és Automatizálási Kutató Intézete, CDC 3300 felhasználói ismertetők 2. 1973. május) 19—20.

[3] Prékopa, A., „Sztohasztikus 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 100 darab különlenyomatot kapnak, ezek költsége — nyomott oldalanként 25 forint — a szerzői díjat terheli.

Ára: 42,— Ft

INDEX: 26498
ISSN 0133—3399

TARTALOMJEGYZÉK

Gergely József: A Poisson-egyenlet numerikus megoldásairól 211
Abaffy József: A lineáris egyenletrendszerek általános megoldásának egy direkt módszeroszt ályа 233
Ecsedi István: Egy módszer a hőáram becslésére 241
Kelle Péter: Az alapanyagok keverési arányának és a tárolók nagyságának optimalizálása aszfaltkeverő berendezésekre 249
Aigbe William: Egy kvázi-belső pont eljárás lineáris és nemlineáris feltételeket tartalmazó nemlineáris programozási feladatok megoldására 261
Kálovics Ferenc: Globális minimum meghatározása kizárásos módszerrel 269
Kalmár János: A Pegazus módszerek nemlineáris egyenletek megoldására 277
Juhász Ferenc: Szimmetrikus véletlen (0, 1) mátrix spektrumáról 289
Rétháti László: Ferde eloszlású adatsorok szimmetrikussá tétele hatványozással 295
Pham Ngoc Phuc: Autoregressziós típusú Gauss-folyamatok néhány jellemzési problémájáról 303
Tusnády Gábor: Mátrixok szinguláris felbontása 375
Lengyel Tamás: A kanonikus korrelációanalízis és néhány kapcsolódó probléma 385

A külföldi szakirodalomból

Szamarszkij, A. A.: Mi is az a számítógépes kísérlet? 395

INDEX

Gergely, J., "On numerical solution of the Poisson equation" 211
Abaffy, J., "A direct method class for the general solution of systems of linear equations"..... 233
Ecsedi, I., "A method to estimation of heat-rate" 241
Kelle, P., "Optimization of mixture rate and depot capacities for asphalt mixers" 249
Aigbe, W. F., "A barrier method for solving nonlinear programming problems" 261
Kálovics, F., "Determination of the global minimum by the method of exclusions" 269
Kalmár, J., "The Pegasus methods for the solution of nonlinear equations" 277
Juhász, F., "On the spectrum of a symmetric random (0, 1) matrix" 289
Rétháti, L., "Symmetriesierung von schräg verteilten Datenreihen mit Potenzierung" 295
Pham Ngoc Phuc, "On characterization problems of autoregression type Gaussian processes" 303
Tusnády, G., "On the singular decomposition of matrices" 375
Lengyel, T., "The canonical correlation analysis and some related problems" 385

From the foreign literature

Самарский, А. А., Что такое вычислительный эксперимент? 395