

# A konjugált gradiens-módszer konvergenciája nemszimmetrikus operátorokra

Diplomamunka

Írta: Kurics Tamás

Alkalmazott matematikus szak

Témavezető:

Karátson János, egyetemi docens

Alkalmazott Analízis Tanszék

Eötvös Loránd Tudományegyetem, Természettudományi Kar



Eötvös Loránd Tudományegyetem

Természettudományi Kar

2005

## TARTALOMJEGYZÉK

<i>Előszó</i> . . . . .	ii
1. <i>Bevezetés</i> . . . . .	1
1.1. Motiváció . . . . .	1
1.2. Felhasznált definíciók, tételek . . . . .	2
2. <i>A konjugált gradiens-módszer</i> . . . . .	5
2.1. A konjugált gradiens-módszer algoritmus . . . . .	5
2.2. Szuperlineáris konvergencia az identitás kompakt perturbációjára . . . . .	7
3. <i>Prekondicionálás az operátor szimmetrikus részével</i> . . . . .	9
3.1. Korlátos operátor esete . . . . .	9
3.2. Nemkorlátos operátor esete . . . . .	10
4. <i>Prekondicionálás és szuperlineáris konvergencia</i> . . . . .	13
5. <i>Konvekció-diffúzió egyenletek</i> . . . . .	18
5.1. Egy speciális egyenlet . . . . .	18
5.2. Az általános egyenlet . . . . .	20
5.3. Egyéb prekondicionáló eljárások . . . . .	23
6. <i>Numerikus kísérletek</i> . . . . .	25
6.1. A diszkrét probléma rácsfüggetlen szuperlineáris konvergenciája . . . . .	25
6.2. A vizsgált tesztfeladatok . . . . .	26
6.3. A szuperlineáris konvergencia numerikus ellenőrzése . . . . .	28
6.4. Konjugált gradiens-módszer vs. egy gradiens-típusú iteráció . . . . .	37
<i>Irodalomjegyzék</i> . . . . .	40

## ELŐSZÓ

A dolgozatban a konjugált gradiens-módszer egy speciális tulajdonságát vizsgáljuk. Ha az adott  $Lu = g$  egyenlet operátora felírható az identitás és egy kompakt, normális operátor összegeként, akkor úgynevezett szuperlineáris konvergencia teljesül, ami azt jelenti, hogy létezik egy  $\varepsilon_k \xrightarrow{k \rightarrow \infty} 0$  sorozat, hogy az algoritmus  $e_k$  hibavektorára teljesül az

$$\left( \frac{\|e_k\|}{\|e_0\|} \right)^{1/k} \leq \varepsilon_k$$

becslés. Mivel  $L$  nemkorlátos operátor, a fenti becslésben valamilyen energia-típusú norma szerepel. Ilyen alakra hozható az egyenlet, ha prekondicionálunk az operátor  $S$  szimmetrikus részével, azaz szorzunk az  $S$  inverzével, ahol  $S = (L + L^*)/2$ . Ebben a speciális esetben a fent említett  $\varepsilon_k$  sorozat az  $S^{-1}Q$  kompakt operátor sajátértékeiből elkészített számtani-közepekből álló sorozat, ahol  $Q = (L - L^*)/2$  az  $L$  antiszimmetrikus része. Természetesen az  $L$  és  $L^*$  operátorok értelmezési tartománya különbözhet és ez megnehezíti az  $S$  operátor bevezetését, inverzének elkészítését, ezért ezeket a fogalmakat is elő kell készítenünk speciális és általános esetben is. Megmutatjuk, hogy milyen feltételeket kell tennünk, hogy az algoritmus egyik legegyszerűbb változatát használhassuk és ezt konvekció-diffúzió-típusú egyenletekre ellenőrizzük. Végül konkrét egyenletekre végzünk számításokat, mind a Dirichlet-peremfeltétel, mind a vegyes peremfeltétel esetére. Több numerikus kísérlet fogja igazolni a tételek helyességét, amelyek során a  $h$  rácsállandótól független, konkrétan kiszámolt  $\varepsilon_k$  nullsorozatra ellenőrizzük a fenti becslés teljesülését, sőt kiderül, hogy az elmélet által szolgáltatott becslés jóval pesszimistább a gyakorlati eredmények értékeinél. Ezenkívül egyéb prekondicionáló eljárásokat is megvizsgálunk és összehasonlítjuk algoritmusunk gyorsaságát egy egyszerűbb, gradiens-típusú eljárással.

## KÖSZÖNETNYILVÁNÍTÁS

Ezúton is szeretném megköszönni témavezetőmnek, *Karátson János*nak, hogy erre az elmélet és gyakorlat határán elhelyezkedő érdekes területre ráirányította figyelmemet. Állandó segítőkészsége és lényeges részeket érintő tanácsai rendkívül hasznosnak bizonyultak a dolgozat írása során.

# 1. BEVEZETÉS

## 1.1. Motiváció

A másodrendű lineáris elliptikus parciális differenciálegyenletek numerikus megoldásának sok hatékony módszere ismert, olyanok is, amelyek kivitelezése számítástechnikailag is egyszerű és futási ideje, valamint tárigénye is elfogadható a kis és közepes méretű feladatok esetén. Akár véges-differencia módszert alkalmazunk, akár végelemek eljárásokat, az egyenlet visszavezethető egy  $L_h u_h = f_h$  egyenletre, ami már egy véges dimenziós térben van, azaz  $L_h$  egy mátrix, amelynek mérete a  $h$  rácsállandótól függ. Nagyméretű mátrixok esetén egy ilyen tisztán lineáris algebrai feladat megoldása nem egyszerű, főleg ha a mátrixnak nincs olyan tulajdonsága, ami egyszerűsíthetné a megoldás keresését; a mi esetünkben  $L$  egy nonszimmetrikus mátrix. Prekondicionálás során viszont a feladatot visszavezetjük egy iterációra és minden iterációs lépésben egy szimmetrikus egyenletet oldunk meg - most már egzaktul - ugyanazzal a mátrixszal, valamint segédvektorokat és ezek normáit kell számolnunk. Tehát egyetlen egyenlet megoldását több - de könnyebb - egyenlet megoldására cseréltük fel, emellett több skalárszorzatot is számolni kell az egyes lépésekben. Érezhető, hogy ez akkor éri meg, ha a lépések száma nem túl nagy, azaz gyors konvergenciát tudunk biztosítani. További előnyös tulajdonság lenne, hogy előírt pontosság esetén az iterációk száma ne függjön a rácsállandótól, azaz a felosztás finomításakor a lépések száma ne változzon nagymértékben, vagy inkább egyáltalán ne változzon.

A dolgozat témája a konjugált gradiens-módszer konvergencia-vizsgálata, az  $Lu = f$  egyenlet prekondicionálása az operátor szimmetrikus részével, csak a főrésszel, illetve a főrészt és egy nulladrendű tag hozzávételével kapott - továbbra is szimmetrikus - operátor segítségével, valamint a fentiek illusztrálása egy konvekció-diffúzió egyenleten keresztül. Az elméleti eredményeket követően egy példafeladatra numerikus eredményekkel is illusztrálom az algoritmus működését és egyben hatékonyságát is.

A konjugált gradiens-módszer jól használható olyan nagyméretű lineáris rendszerek megoldására, amelyek elliptikus parciális differenciálegyenletek diszkrétizációjából származnak. Hibabecslések és konvergenciarend is ismert, ám minden ilyen jellegű eredmény természetesen olyan, hogy legrosszabb esetet vesz figyelembe, ezért érdekes kérdés, hogy az elméleti eredmény mennyire túlzó a gyakorlati számítási eredményekhez képest. Ha egy nonszimmetrikus operátor által adott egyenletet prekondicionálunk a szimmetrikus részével, vajon az elméleti eredmények mennyire realizálhatóak a gyakorlati számítások során.

A kiindulási egyenletünk olyan lesz, hogy a prekondicionált alak operátora az iden-

titás és egy kompakt operátor összegére bontható, így ún. szuperlineáris konvergencia is bizonyítható. Bár a módszer igazi ereje a 3 dimenziós feladatoknál mutatkozik meg, kétdimenziós példafeladatokra ellenőriztük az elméleti becsléseket, noha ilyenkor - főleg ha nem óriási méretű a kétdimenziós probléma - még a feladat mérete nem indokolja ezt a fajta iterációs módszert, mert az egyenlet direkt megoldása is versenyképes marad, ami elég sűrű felosztás esetén, vagy méginkább 3 dimenzióban már nincs így.

## 1.2. Felhasznált definíciók, tételek

**1.1. Tétel.** *Legyen  $H$  Hilbert-tér,  $A \in L(H)$  önadjungált. Ekkor*

$$\|A\| = \sup_{\|x\|=1} |\langle Ax, x \rangle|. \quad (1.1)$$

**1.2. Tétel.** *Legyen  $A \in L(H)$  egyenletesen pozitív, azaz  $\exists m > 0$ , hogy  $\langle Ax, x \rangle \geq m \|x\|^2$ . Ekkor  $\forall y \in H$  esetén az  $Ax = y$  egyenletnek  $\exists!$  megoldása.*

**1.3. Tétel.** *Legyen  $H$  végtelen dimenziós Hilbert-tér,  $A \in L(H)$  kompakt önadjungált operátor. Ekkor*

i) *A spektruma megszámlálható,  $\sigma(A) = (\bigcup_{k \in \mathbb{N}} \{\lambda_k\}) \cup \{0\}$ .*

ii) *Ezen  $\lambda_k$  sajátértékek csak a 0-ban torlódhatnak.*

iii) *Az egyes sajátértékekehez véges dimenziós sajátaltér tartozik (ha  $\lambda \neq 0$ ).*

iv) *A sajátvektorok megválaszthatóak úgy, hogy teljes ortonormált rendszert alkossanak.*

**1.4. Tétel (Gauss–Ostrogradszkij).** *Legyen  $\Omega \subset \mathbb{R}^n$  lokálisan téglaszerű korlátos tartomány,  $f : \overline{\Omega} \rightarrow \mathbb{C}^n$  folytonosan differenciálható. Ekkor*

$$\int_{\Omega} \operatorname{div} f = \int_{\partial\Omega} \langle f, \nu \rangle \, d\sigma. \quad (1.2)$$

**1.5. Tétel (Green 1.).** *Legyen  $\Omega \subset \mathbb{R}^n$  lokálisan téglaszerű korlátos tartomány,  $p \in C^1(\overline{\Omega})$ ,  $q \in C(\overline{\Omega})$ . Legyen  $L$  a következő differenciáloperátor:  $Lu = \operatorname{div}(p \cdot \nabla u) + qu$ . Ekkor minden  $u \in C^2(\overline{\Omega})$ ,  $v \in C^1(\overline{\Omega})$  függvényre érvényes:*

$$\int_{\Omega} \bar{v} Lu = \int_{\Omega} (-p \langle \nabla u, \nabla v \rangle + qu \bar{v}) + \int_{\partial\Omega} p \bar{v} \partial_{\nu} u \, d\sigma. \quad (1.3)$$

**1.6. Tétel (Green 2.).** *A fenti feltételek mellett, ha most  $u, v \in C^2(\overline{\Omega})$ , akkor*

$$\int_{\Omega} (\bar{v} Lu - u L\bar{v}) = \int_{\partial\Omega} p (\bar{v} \partial_{\nu} u - u \partial_{\nu} \bar{v}) \, d\sigma. \quad (1.4)$$

*Itt és az előző tételnél  $\nu$  a  $\partial\Omega$  peremfelületnek a tartományból kifelé mutató egységnormálisa.*

Szükségünk lesz a divergenciatételt és a többdimenziós parciális integrálás szabályát kombináló tételre, melyekkel könnyen tudjuk majd kezelni a konvekció-diffúzió típusú elliptikus parciális differenciál-egyenleteinket. Most is és a dolgozat hátralevő részében felteesszük, hogy a szóbanforgó  $\Omega$  tartomány kellően szép, azaz olyan korlátos tartomány, amelynek a pereme elég sima, például legyen a perem Lipschitz-folytonos, vagy szakaszonként folytonosan differenciálható.

A parciális integráláshoz (lásd [13]) legyenek  $\mathbf{v} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ ,  $u : \mathbb{R}^n \rightarrow \mathbb{R}$  függvények, ekkor

$$\int_{\Omega} u \frac{\partial \mathbf{v}_i}{\partial x_i} = \int_{\partial\Omega} u \mathbf{v}_i n_i - \int_{\Omega} \frac{\partial u}{\partial x_i} \mathbf{v}_i.$$

Ezt összegezve  $i = 1$ -től  $n$ -ig:

$$\int_{\Omega} u \operatorname{div} \mathbf{v} = \int_{\partial\Omega} u \langle \mathbf{v}, \mathbf{n} \rangle - \int_{\Omega} \langle \operatorname{grad} u, \mathbf{v} \rangle. \quad (1.5)$$

Végül a szorzatfüggvény gradienseinek kiszámolása:

### 1.7. Állítás.

$$\int_{\Omega} (\langle \mathbf{b}, \nabla u \rangle v + u \langle \mathbf{b}, \nabla v \rangle) = \int_{\Omega} \langle \nabla (uv), \mathbf{b} \rangle. \quad (1.6)$$

Szükségünk lesz a szimmetrikus és az önadjungált operátor fogalmára. Amikor eredményeinket alkalmazni szeretnénk parciális differenciálegyenletekre, akkor a megfelelő differenciál-operátor nem lesz korlátos, továbbá a peremfeltételek miatt általában az adjungált sem értelmezhető mindenhol.

A prekondicionálás során az  $L$  operátor  $S$  szimmetrikus részét képezzük valamilyen értelemben, majd ennek inverzével szorozzuk az egyenletet. Természetesen ahhoz, hogy ez értelmes legyen, az inverz létezése mellett szükséges lesz, hogy az inverz értelmezési tartománya a teljes tér legyen, azaz  $R(S) = H$ . Erre vonatkozó eredményt készítenek elő az alábbi definíciók és állítások.

**1.8. Definíció.** *Legyen  $A : H \rightarrow H$  lineáris operátor, amely sűrűn definiált, azaz  $D(A) \subset H$  sűrű halmaz  $H$ -ban. Az  $A$  operátor szimmetrikus, ha*

$$\langle Ax, y \rangle = \langle x, Ay \rangle \quad \forall x, y \in D(A).$$

**1.9. Definíció.** *Legyen  $A : H \rightarrow H$  sűrűn definiált operátor. Az  $A$  adjungáltja az az  $A^*$  operátor, amelyre  $D(A^*) = \{y \in H : \exists y^* \in H, \text{ hogy } \langle Ax, y \rangle = \langle x, y^* \rangle \forall x \in D(A)\}$ . Az így definiált értelmezési tartomány elemeire  $A^*y := y^*$ . Az  $A$  operátor önadjungált, ha  $A = A^*$ .*

**1.10. Állítás.** *Az  $A : H \rightarrow H$  sűrűn definiált operátor szimmetrikus  $\iff A \subset A^*$ .*

Egy operátorról általában triviálisan látható, hogy szimmetrikus, az önadjungáltság kérdése annál nehezebb, hiszen az adjungált értelmezési tartománya lehet akár diszjunkt<sup>1</sup> is az eredeti operátorétól. A következő állítás egy jó feltételt ad, amit később ki is használunk:

<sup>1</sup> Azaz  $D(A) \cap D(A^*) = \{0\}$ .

**1.11. Állítás.** *Legyen  $A : H \rightarrow H$  sűrűn definiált szimmetrikus operátor, amely egyenletesen pozitív. Ekkor  $A$  pontosan akkor szürjektív, ha önadjungált.*

**1.12. Definíció.** *Legyen  $A : H \rightarrow H$  szigorúan pozitív operátor. Ekkor az  $\langle u, v \rangle_A := \langle Au, v \rangle$  összefüggés skalárszorzatot definiál. Az  $A$  operátor által meghatározott energiatér  $H_A := [D(A), \langle \cdot, \cdot \rangle_A]$  teljessé tétele.*

Speciális eset, ha  $A$  egyenletesen pozitív, azaz  $\exists p > 0$ , hogy  $\langle Ax, x \rangle \geq p \|x\|^2$ . Ekkor

$$\|u\|_A^2 = \langle u, u \rangle_A = \langle Au, u \rangle \geq p \|u\|^2 \quad \forall u \in D(A). \quad (1.7)$$

**1.13. Állítás.** *Ha  $A$  egyenletesen pozitív, akkor  $H_A \subset H$  azonosítás erejéig.*

## 2. A KONJUGÁLT GRADIENS-MÓDSZER

A következőkben feltesszük, hogy  $H$  egy komplex, szeparábilis Hilbert-tér. Legyen a vizsgált operátoregyenlet az

$$Au = b, \tag{2.1}$$

ahol  $b \in H$  egy adott vektor,  $A$  egy nemszimmetrikus operátor, amelynek létezik korlátos inverze.

### 2.1. A konjugált gradiens-módszer algoritmusa

Egy ilyen alakú egyenlet megoldására a következő GCG-LS (*generalized conjugate gradient - least square*) módszer alkalmazható. Az  $u_k$   $k$ -ik közelítést az előző közelítés és az előző  $k$  darab keresési irány lineáris kombinációjaként állítjuk elő. Ez az algoritmus egyszerre állítja elő az  $u_k$  közelítő megoldások sorozatát és a  $d_k$  keresési irányokat oly módon, hogy az  $Ad_k$  vektorok lineárisan függetlenek és az első  $k$  keresési irány által kifeszített altérben éppen az  $u_k$  vektor lesz megoldása a  $\min\{\|Av - b\| : v \in \text{span}\{d_1, d_2, \dots, d_k\}\}$  minimumfeladatnak. Legyen  $s_k = \min\{k, s\}$ , azaz ha elértük az  $s$  küszöböt, akkor az új keresési irány számításánál nem vesszük figyelembe az összes korábbi, de az  $u_{k+1}$  új közelítő megoldásnál továbbra is minden korábbi irányt figyelembe veszünk.

$$\left\{ \begin{array}{l} (1) \quad \text{Legyen } u_0 \in H \text{ tetszőleges és legyenek } r_0 = Au_0 - b, d_0 = -r_0; \\ \text{Tetszőleges } k \in \mathbb{N} \text{ esetén ha } u_k, d_k, r_k \text{ már megvan, akkor legyenek} \\ (2a) \quad \text{az } \alpha_{k-j}^{(k)} \quad (j = 0, \dots, k) \quad \text{számok az alábbi egyenletrendszer megoldásai} \\ \quad \sum_{j=0}^k \alpha_{k-j}^{(k)} \langle Ad_{k-j}, Ad_{k-l} \rangle = -\langle r_k, Ad_{k-l} \rangle \quad (0 \leq l \leq k); \\ (2b) \quad u_{k+1} = u_k + \sum_{j=0}^k \alpha_{k-j}^{(k)} d_{k-j}; \\ (2c) \quad r_{k+1} = r_k + \sum_{j=0}^k \alpha_{k-j}^{(k)} Ad_{k-j}; \\ (2d) \quad \beta_{k-j}^{(k)} = \langle Ar_{k+1}, Ad_{k-j} \rangle / \|Ad_{k-j}\|^2 \quad (j = 0, \dots, s_k); \\ (2e) \quad d_{k+1} = -r_{k+1} + \sum_{j=0}^{s_k} \beta_{k-j}^{(k)} d_{k-j}. \end{array} \right. \tag{2.2}$$

Létezik egy ún. csonkított verzió is, amely csak az előző  $t_k$  keresési irányt használja, ahol adott egy  $t \in \mathbb{N}^+$  természetes szám és  $1 \leq t_k \leq \min\{t_{k-1} + 1, t\}$  minden  $k$ -ra.



Megmutatható, hogy ekkor  $\alpha_{k-j}^{(k)} = 0$  ( $1 \leq j \leq t_{k-1}$ ) (lásd [1, 2]). Ekkor az  $\alpha_{k-t_k}^{(k)} d_{k-t_k}$  tag (control term) nagysága mutatja, hogy további tagok elhagyása elhanyagolható hatású-e vagy sem. Ha ez a control term nulla, akkor  $u_{k+1}$  és  $r_{k+1}$  kiszámításához mindössze az utolsó,  $d_k$  keresési irány szükséges.

$$\left\{ \begin{array}{l} (1) \quad \text{Legyen } u_0 \in H \text{ tetszőleges és legyenek } r_0 = Au_0 - b, d_0 = -r_0; \\ \quad \text{Tetszőleges } k \in \mathbb{N} \text{ esetén ha } u_k, d_k, r_k \text{ már megvan, akkor legyenek} \\ (2a) \quad \gamma_k = \|Ad_k\|^2, \quad \alpha_k = -\frac{1}{\gamma_k} \langle Ad_k, r_k \rangle; \\ (2b) \quad u_{k+1} = u_k + \alpha_k d_k; \\ (2c) \quad r_{k+1} = r_k + \alpha_k Ad_k; \\ (2d) \quad \beta_{k-j}^{(k)} = \langle Ar_{k+1}, Ad_{k-j} \rangle / \|Ad_{k-j}\|^2 \quad (j = 0, \dots, s_k); \\ (2e) \quad d_{k+1} = -r_{k+1} + \sum_{j=0}^{s_k} \beta_{k-j}^{(k)} d_{k-j}. \end{array} \right. \quad (2.3)$$

Ez a csonkított algoritmus az új keresési irány kiszámításához  $s_k$  - ami a  $k$ -ik lépéstől kezdődően  $s + 1$ -gyel egyenlő - korábbi irányt használ fel a (2e) sor szerint. Ennek neve GCG-LS( $s$ ). Mint az később, a konkrét feladat vizsgálatokor kiderül, a nekünk szükséges algoritmus a GCG-LS(0), ami egyetlen keresési irány tárolását igényli. Ekkor a fenti algoritmus (2d) és (2e) sora egyszerűsödik.

$$\left\{ \begin{array}{l} (1) \quad \text{Legyen } u_0 \in H \text{ tetszőleges és legyenek } r_0 = Au_0 - b, d_0 = -r_0; \\ \quad \text{Tetszőleges } k \in \mathbb{N} \text{ esetén ha } u_k, d_k, r_k \text{ már megvan, akkor legyenek} \\ (2a) \quad \gamma_k = \|Ad_k\|^2, \quad \alpha_k = -\frac{1}{\gamma_k} \langle Ad_k, r_k \rangle; \\ (2b) \quad u_{k+1} = u_k + \alpha_k d_k; \\ (2c) \quad r_{k+1} = r_k + \alpha_k Ad_k; \\ (2d) \quad \beta_k = \frac{1}{\gamma_k} \langle Ar_{k+1}, Ad_k \rangle; \\ (2e) \quad d_{k+1} = -r_{k+1} + \beta_k d_k. \end{array} \right. \quad (2.4)$$

Kérdés, hogy vajon mikor tudjuk az egyszerűbb, kevesebb számolást igénylő csonkított algoritmust használni. Erre vonatkozik az alábbi tétel.

**2.1. Tétel.** *Tekintsük a (2.1) egyenlet  $A$  korlátos lineáris operátorát és tegyük fel, hogy  $A + A^* > 0$ , továbbá  $A^*$  polinomja  $A$ -nak, azaz létezik egy valós együtthatós  $p_n$   $n$ -edfokú polinom, hogy  $A^* = p_n(A)$ . Ekkor  $s \geq n - 1$  esetén a (2.3) csonkított GCG-LS( $s$ ) egybeesik a teljes (2.2) algoritmussal.*

**2.2. Megjegyzés.** *Véges dimenziós komplex euklideszi térben bizonyítható (lásd [7]), hogy egy  $A$  lineáris leképezés adjungáltja pontosan akkor polinomja  $A$ -nak, ha normális.*

Végtelen dimenzióban ez nincs így. Kompakt operátor esetén - mivel előáll véges rangú operátorok limeszeként - elég nagy  $s$ -re a csonkított algoritmus már közelítőleg megegyezik a teljes verzióval. Hasonló várható az identitás kompakt perturbációjára. A mi esetünkben hasonlóról lesz szó, de szerencsére majd több feltevés is biztosítani fogja, hogy létezzen a 2.1.tételbeli  $p_n$  polinom, méghozzá elsőfokú és így a már említett GCG-LS(0) algoritmust tudjuk használni.

Természetesen prekondicionálás után már nem a (2.1) egyenletre fogjuk alkalmazni a (2.4) algoritmust, hanem egy  $Bu \equiv S^{-1}Au = S^{-1}b \equiv c$  alakú egyenletre. Arra a kérdésre, hogy mi ez az  $S$  operátor és a bal oldalon szereplő operátor hogyan értelmezhető általában, a 3. fejezetben adunk választ.

## 2.2. Szuperlineáris konvergencia az identitás kompakt perturbációjára

Ha az egyenletünk operátora szimmetrikus, pozitív definit, akkor a konjugált gradiens-módszer konvergenciáját széles körben vizsgálták, könnyen lehet különböző ortogonalitási tulajdonságokat, konvergenciát bizonyítani és az algoritmus is egyszerűbb. A nemszimmetrikus esetet is behatóan megvizsgálták (pl. [1, 2]). A jelen dolgozat szempontjából legfontosabb eredmény a szuperlineáris konvergencia.

Ha tekintjük a (2.1) alakú véges dimenziós egyenletet, ahol  $A$  egy  $n \times n$ -es normális mátrix, amelynek a szimmetrikus része pozitív definit, akkor érvényes az alábbi becslés. Írjuk fel  $A$ -t az identitás perturbációjaként, azaz  $C = A - I$  jelöléssel  $A = I + C$ . Legyenek  $\lambda_i \in \mathbb{C}$  ( $i = 1, \dots, n$ ) a  $C$  mátrix sajátértékei úgy indexelve, hogy  $|\lambda_i| \geq |\lambda_{i+1}|$ . Ekkor

$$\left( \frac{\|e_k\|_A}{\|e_0\|_A} \right)^{1/k} \leq \frac{2}{k} \sum_{i=1}^k \left| \frac{\lambda_i}{1 + \lambda_i} \right| \quad (k = 1, 2, \dots, n), \quad (2.5)$$

ahol az  $\|e_k\|_A = \langle Ae_k, e_k \rangle^{1/2}$  jelölést használjuk, komplex esetben a gyökvonás előtt még valós részt kell venni.

**2.3. Megjegyzés.** *A fenti felbontást úgy is tekinthetjük, hogy az  $A$  operátor előáll az identitás és egy kompakt operátor összegeként, hiszen véges dimenzióban minden lineáris leképezés kompakt.*

Itt és a továbbiakban is  $e_k = u_k - u^*$  jelöli a hibavektort, azaz  $u^*$  az egyenlet megoldása, ami a korlátos inverz létezésének feltételezése miatt egyértelmű.

Az előző, véges dimenziós eset kiterjeszthető végtelen dimenzióba. Legyen szokásos módon  $H$  szeparábilis Hilbert-tér,  $C$  egy kompakt lineáris operátor a  $\lambda_i \in \mathbb{C}$  ( $i = 1, \dots, n$ ), nagyság szerint rendezett  $|\lambda_i| \geq |\lambda_{i+1}|$  sajátértékekkel. Legyen  $A = I + C$  az identitás kompakt perturbációja. Ekkor a konjugált gradiens-módszer hibavektorára fennáll (2.5), sőt, mivel  $C$  kompakt, ezért  $\lambda_i \rightarrow 0$  és így a szummajel mögött álló tört nullához tart, emiatt ezen sorozat számtani-közép sorozata is. További észrevétel, hogy  $A$  korlátos inverzének

létezése - amit mindig felteszünk - egyenértékű azzal, hogy  $\inf_i |1 + \lambda_i| > 0$ . Felhasználva az ebből következő  $\|A^{-1}\| = \sup_i (1/|1 + \lambda_i|)$  egyenlőséget, a szuperlineáris konvergenciát mutató (2.5) becslés jobb oldala felülről becsülhető:

$$\frac{2}{k} \sum_{i=1}^k \left| \frac{\lambda_i}{1 + \lambda_i} \right| \leq 2 \|A^{-1}\| \left( \frac{1}{k} \sum_{i=1}^k |\lambda_i| \right). \quad (2.6)$$

Ha a  $C$  operátor kompakt és önadjungált, akkor a sajátvektorai teljes ortonormált rendszert alkotnak  $H$ -ban, ezért a (2.5) becslés érvényes minden  $n \in \mathbb{N}$ -re. Még általánosabban is érvényes a becslés minden olyan nem feltétlenül szimmetrikus operátorra is, amelyek sajátvektorai teljes ortonormált rendszert alkotnak. Ezen operátorok éppen a kompakt, normális operátorok osztálya  $H$ -ban, ezért tétel formájában a következő mondható:

**2.4. Tétel.** *Legyen  $H$  szeparábilis Hilbert-tér,  $C$  egy kompakt, normális lineáris operátor a  $\lambda_i \in \mathbb{C}$ ,  $|\lambda_i| \geq |\lambda_{i+1}|$  sajátértékekkel. Legyen  $A = I + C$  az identitás kompakt perturbációja, amelynek létezik korlátos inverze. Ekkor a KGM hibavektorára teljesül a*

$$\left( \frac{\|e_k\|_A}{\|e_0\|_A} \right)^{1/k} \leq 2 \|A^{-1}\| \left( \frac{1}{k} \sum_{i=1}^k |\lambda_i| \right) \xrightarrow{k \rightarrow \infty} 0. \quad (2.7)$$

*becslés minden  $k \in \mathbb{N}$ -re.*

### 3. PREKONDICIONÁLÁS AZ OPERÁTOR SZIMMETRIKUS RÉSZÉVEL

Ezen fejezetben megmutatjuk, hogy a 2.1. tétel alkalmazható, ha az operátort a szimmetrikus részével prekondicionáljuk, sőt, a 2.2. megjegyzésben szereplő előzetes állítás bizonyítása miatt a teljes algoritmus helyett elég lesz a GCG-LS(0) használata. A téma részletes tárgyalásához lásd [3]-at.

Legyen most  $H$  valós Hilbert-tér, és tekintsük a

$$Bu = c \quad (3.1)$$

egyenletet, ahol  $B : H \rightarrow H$  lineáris operátor, többnyire nem szimmetrikus, sőt a korlátosságot sem célszerű feltenni, hiszen egy differenciál-operátorra szeretnénk alkalmazni az eredményt. Tegyük fel viszont, hogy létezik  $p > 0$ , hogy

$$\langle Bu, u \rangle \geq p \|u\|^2 \quad (u \in D(B)). \quad (3.2)$$

#### 3.1. Korlátos operátor esete

Tegyük fel elsőként, hogy  $B$  korlátos operátor. Legyen  $S$  és  $Q$  az operátor szimmetrikus, illetve antiszimmetrikus része, azaz  $S = (B + B^*)/2$ ,  $Q = (B - B^*)/2$ . Vegyük észre, hogy a definíció értelmes, mert az adjungált operátor mindenhol értelmezett.

**3.1. Állítás.** *Az  $S$  operátorra teljesülnek az alábbi tulajdonságok:*

i)  $S$  önadjungált és  $p \|u\|^2 \leq \langle Su, u \rangle \leq \|B\| \|u\|^2 \quad (u \in H)$ ,

ii)  $R(S) = H$ ,

iii) Létezik az  $S^{-1}B : H \rightarrow H$  operátor.

**Bizonyítás.** Az  $S$  önadjungáltsága a definíciójának következménye, valamint  $\langle Bu, u \rangle = (\langle Bu, u \rangle + \langle Bu, u \rangle)/2 = (\langle Bu, u \rangle + \langle B^*u, u \rangle)/2 = \langle Su, u \rangle$  miatt (3.2)-ből következik i), az operátoregyenletek megoldhatóságáról szóló 1.2. tétel miatt ii) is teljesül, amiből már iii) triviális.  $\square$

Az i)-es állítás miatt  $H$  teljes marad, ha az  $\|\cdot\|_S$  normát használjuk az eredeti helyett. A prekondicionált egyenlet ekkor a következő alakot ölti: ha  $A = S^{-1}B$ , akkor  $Au = b = S^{-1}c$ .

**3.2. Állítás.** *A korlátos operátor az  $\|\cdot\|_S$  normában.*

**Bizonyítás.** Tetszőleges  $u \in H$  esetén

$$\begin{aligned} \|S^{-1}Bu\|_S &= \sup_{\|v\|_S=1} \langle S^{-1}Bu, v \rangle_S = \sup_{\|v\|_S=1} \langle Bu, v \rangle \leq \sup_{\|v\|_S=1} \|B\| \|u\| \|v\| \leq \\ &\leq \sup_{\|v\|_S=1} (\|B\|/p) \|u\|_S \|v\|_S = (\|B\|/p) \|u\|_S. \end{aligned}$$

□

Most belátjuk, hogy korlátos  $B$  operátor esetén az adjungált polinomja  $B$ -nek, így a csonkított algoritmus egybeesik a teljes verzióval.

**3.3. Tétel.** *Legyen  $B : H \rightarrow H$  lineáris, korlátos operátor, melyre teljesül (3.2), és legyen  $A = S^{-1}B$ . Ekkor  $GCG-LS(0)$  az (2.1) egyenletre alkalmazva egybeesik a teljes verzióval.*

**Bizonyítás.** Legyen  $A_S^*$  az  $A$  adjungáltja az  $\|\cdot\|_S$  normára nézve. Mivel tetszőleges  $u, v \in H$ -ra

$$\langle Au, v \rangle_S = \langle Bu, v \rangle = \langle u, B^*v \rangle = \langle u, S^{-1}B^*v \rangle_S,$$

ezért  $A_S^* = S^{-1}B^*$ . Továbbá

$$A = S^{-1}B = S^{-1}(S + Q) = I + S^{-1}Q,$$

emiatt

$$A_S^* = S^{-1}B^* = S^{-1}(S - Q) = I - S^{-1}Q = 2I - A = p_1(A),$$

ahol  $p_1(t) = -t + 2$  elsőfokú polinom. □

**3.4. Megjegyzés.** *A  $B$  operátor korlátosságát erősen kihasználtuk. A 3.1. állításból adódott, hogy  $S$  korlátos és lineáris és az  $i)$  becslés miatt alkalmazható megoldhatósági tétel szerint szűrjektív. Ez biztosítja, hogy az  $S^{-1}B$  operátor értelmes legyen. Ugyanígy a 3.2. állítás sem igaz a korlátosság nélkül. Ez utóbbit egy feltétel fogja helyettesíteni, ezért elég az első problémán gondolkozni.*

Két esetet különböztetünk meg aszerint, hogy  $S$  önadjungált, illetve csak szimmetrikus. Az első esetben az  $R(H) = H$  feltétel biztosított, a másodikban a 1.10. állítás szerint  $S^* \supset S$  (feltéve, hogy az értelmezési tartomány sűrű), azaz  $D(S^*) \supset D(S)$ , így az  $A$  operátort gyenge értelemben definiáljuk.

### 3.2. Nemkorlátos operátor esete

Most tehát legyen  $B : D(B) \rightarrow H$  lineáris, *nemkorlátos* operátor, mely  $D(B)$  értelmezési tartománya sűrű  $H$ -ban. Az adjungáltja a bevezető fejezetben szereplő 1.9. definícióval értelmezhető. Ekkor az  $S$  és  $Q$  operátorok ugyanolyan alakúak, mint korlátos esetben, de értelmezési tartományuk most nem az egész  $H$ , hanem csak  $D \equiv D(B) \cap D(B^*)$ . Tegyük fel továbbá, hogy  $Q$   $S$ -korlátos. Ez nem feltétlenül teljesül, lásd [3].

Elsőnek tegyük fel, hogy  $S$  önadjungált, továbbá teljesüljön

$$\langle Su, u \rangle \geq p \|u\|^2 \quad (u \in D(B)). \quad (3.3)$$

Ez utóbbi automatikusan teljesül (3.2) miatt, lásd a 3.1. állítás i) részének első egyenlőtlenségének bizonyítását, ami itt is megismételhető  $S$  önadjungáltsága miatt. Felhasználva az 1.11. tételt, kapjuk, hogy  $R(S) = H$ , azaz az  $A = S^{-1}B$  operátor ugyanúgy létezik, mint a korlátos esetben, most az értelmezési tartománya megegyezik a  $D(B)$  sűrű halmazzal.

**3.5. Állítás.** *Tegyük fel, hogy  $Q$   $S$ -korlátos, azaz az  $(u, v) \mapsto \langle Qu, v \rangle$  bilineáris forma korlátos az  $\|\cdot\|_S$ -norma szerint minden  $u, v \in D$ -re. Ekkor  $A = S^{-1}B$  korlátos az  $\|\cdot\|_S$  normában  $D$ -n.*

**3.6. Tétel.** *Legyen  $B : D(B) \rightarrow H$  sűrűn definiált lineáris operátor, melyre teljesül (3.2), továbbá  $S$  önadjungált és  $Q$   $S$ -korlátos. Legyen  $A = S^{-1}B$ . Ekkor  $GCG$ - $LS(0)$  az (2.1) egyenletre alkalmazva egybeesik a teljes verzióval.*

**Bizonyítás.** Ugyanúgy belátható, hogy  $A^* = 2I - A$ , pontosabban az egyenlőség csak az értelmezési tartomány közös részén,  $D$ -n érvényes. Ugyanis tetszőleges  $u \in D(B), v \in D$ -re

$$\langle Au, v \rangle_S = \langle Bu, v \rangle = \langle u, B^*v \rangle = \langle u, 2Sv - Bv \rangle = \langle u, 2Sv - SAV \rangle = \langle u, (2I - A)v \rangle_S.$$

Most viszont csak azt tudjuk, hogy  $A|_D^* = 2I - A$ . Belátható viszont, hogy a 2.1. tétel igaz akkor is, ha csak  $A|_{R(A)}^* = p_n(A)$  teljesül és mivel most  $R(A) \subset D$ , ezért a tétel alkalmazható.  $\square$

Végül foglalkozunk azzal az esettel, amikor  $S$  csak szimmetrikus, de nem önadjungált. Ekkor  $A$  nem értelmezhető az  $S^{-1}B$  képlettel, hanem gyenge értelemben kell definiálni. Ennek a létezését Riesz reprezentációs tétele biztosítja.

**3.7. Állítás.** *Tegyük fel, hogy  $Q$   $S$ -korlátos, azaz  $|\langle Qu, v \rangle| \leq M \langle Su, u \rangle^{1/2} \langle Sv, v \rangle^{1/2}$  minden  $u, v \in D$ -re valamilyen pozitív  $M$ -re. Ekkor egyértelműen létezik egy olyan  $A : H_S \rightarrow H_S$  lineáris korlátos operátor, amelyre*

$$\langle Au, v \rangle_S = \langle Bu, v \rangle \quad (u, v \in D).$$

**3.8. Tétel.** *Legyen  $B$  sűrűn definiált lineáris operátor  $H$ -n, amelyre teljesül (3.2) és legyen  $A : H_S \rightarrow H_S$  az előző állításbeli operátor. Tegyük fel, hogy  $Q$   $S$ -korlátos. Ekkor  $GCG$ - $LS(0)$  az (2.1) egyenletre alkalmazva egybeesik a teljes verzióval.*

**Bizonyítás.** Az előző bizonyítások megismételhetők, tetszőleges  $u, v \in D$ -re

$$\langle Au, v \rangle_S = \langle u, (2I - A)v \rangle_S.$$

Ez minden  $u \in H_S$ -re is igaz a sűrűség miatt, ezért  $A^*v = (2I - A)v$  ( $v \in D$ )-re. Ebből következik, hogy  $A^* = 2I - A$  a  $H_S$ -en. Azaz a 2.1. tétel alkalmazható az  $A$  operátorra a  $H_S$  térben.  $\square$

**3.9. Megjegyzés.** Ahhoz, hogy a 2.1. tételben szereplő polinom létezését biztosítsuk, elég sok feltevést kellett tennünk, de itt még nem esett szó az eredeti célunkról, a szuperlineáris konvergenciáról. Ha ránézünk az erről szóló tételre, akkor láthatóan még nem biztosítottuk az  $S^{-1}Q$  operátor normalitását és kompaktságát. Ez szükséges lesz, mert a  $Bu = c$  egyenletet átalakítva  $S^{-1}Bu \equiv Au = (I + S^{-1}Q)u = b \equiv S^{-1}c$  alakot kapjuk, ami megfelel a 2.4. tételben szereplő egyenlet alakjának  $C = S^{-1}Q$  szereposztással. Erre a következő fejezetben visszatérünk.

## 4. PREKONDICIONÁLÁS ÉS SZUPERLINEÁRIS KONVERGENCIA

Ennek a fejezetnek a célja, hogy olyan feltételeket adjunk meg a vizsgálandó operátor-egyenletre, amelyek egyrészt biztosítják, hogy a 3. fejezet eredményei érvényben maradjanak<sup>1</sup>, valamint a szuperlineáris konvergenciát biztosító 2.4. tétel feltételei is teljesüljenek.

Tekintsük az

$$Lu = g \tag{4.1}$$

egyenletet, ahol  $L : D \rightarrow H$  sűrűn definiált - általában nemkorlátos - operátor a  $H$  komplex Hilbert-téren,  $g \in H$  tetszőleges adott vektor.

### 4.1. Feltevések. Tegyük fel, hogy

- i)  $L$  felbontható  $L = S + Q$  alakban, ahol  $S$  önadjungált  $H$ -ban,
- ii) Létezik  $p > 0$ , hogy  $\langle Su, u \rangle \geq p \|u\|^2 \quad (u \in D)$ ,
- iii) Létezik  $\rho > 0$ , hogy  $\Re \langle Lu, u \rangle \geq \rho \langle Su, u \rangle \quad (u \in D)$ ,
- iv) A  $Q$  operátor kiterjeszhető a  $H_S$  térbe és  $S^{-1}Q$  kompakt normális operátor  $H_S$ -en.

**4.2. Következmény.** Mivel ii) miatt  $S$  egyenletesen pozitív, a 1.13. tétel szerint  $H_S \subset H$ , valamint  $S$  önadjungáltsága és egyenletes pozitivitása együtt  $S$  szürjektivitását jelenti az 1.11. tétel szerint, így  $S^{-1}Q$  értelmes.

Helyettesítsük az (4.1) egyenletet a prekondicionált alakkal:

$$S^{-1}Lu \equiv \underbrace{(I + S^{-1}Q)}_A u = f \equiv S^{-1}g. \tag{4.2}$$

Ha felhasználjuk, hogy

$$\langle Au, u \rangle_S = \langle (I + S^{-1}Q)u, u \rangle_S = \langle S(I + S^{-1}Q)u, u \rangle = \langle (S + Q)u, u \rangle = \langle Lu, u \rangle, \tag{4.3}$$

és a iii) feltevést, akkor kapjuk, hogy  $A$  korlátos lineáris operátor a  $H_S$  téren, melynek létezik korlátos inverze. Ezek szerint a prekondicionált egyenletnek egyértelműen létezik megoldása a  $H_S$  Hilbert-térben. Teljesül továbbá a 2.4. tétel, amit újból kimondunk. Figyelembevételre, hogy  $\|A^{-1}\|_S \leq 1/\rho$ , valamint  $\|e_k\|_A^2 = \Re \langle Ae_k, e_k \rangle_S = \Re \langle Le_k, e_k \rangle = \|e_k\|_L^2$ , a tétel következő alakját kapjuk:

---

<sup>1</sup> Ez kell ahhoz, hogy a legegyszerűbb, legkevesebb számolást igénylő (2.4) algoritmust tudjuk használni.



**4.3. Tétel.** *Tegyük fel, hogy teljesülnek a 4.1. feltevések. Ekkor a konjugált gradiens-módszert a (4.2) egyenletre alkalmazva fennáll az*

$$\left( \frac{\|e_k\|_L}{\|e_0\|_L} \right)^{1/k} \leq \frac{2}{\rho} \left( \frac{1}{k} \sum_{i=1}^k |\lambda_i(S^{-1}Q)| \right) \xrightarrow{k \rightarrow \infty} 0 \quad (4.4)$$

becslés minden  $k \in \mathbb{N}$ -re, ahol  $\lambda_k(S^{-1}Q)$  az  $S^{-1}Q$  operátor rendezett sajátértékei.

Minket az a speciális eset érdekel, amikor az  $S$  prekondicionáló operátor az  $L$  szimmetrikus része, mert erről szólt a 3. fejezet. Tekintsük a (4.1) egyenletet és tegyük fel, hogy létezik  $p > 0$ , hogy

$$\Re \langle Lu, u \rangle \geq p \|u\|^2 \quad (u \in D := D(L) \cap D(L^*)), \quad (4.5)$$

továbbá legyen  $S$  és  $Q$  az  $L$  szimmetrikus, illetve antiszimmetrikus része, azaz vegyük  $L = S + Q$ -nak azt a felbontását, ahol  $S = (L + L^*)/2$ , illetve  $Q = (L - L^*)/2$ . Tegyük fel még, hogy  $D \subset H$  sűrű  $H$ -ban,  $R(S) = H$ ,  $Q$  kiterjeszthető  $H_S$ -re és  $S^{-1}Q$  kompakt operátor.

**4.4. Állítás.** *Ekkor a 4.1. feltevések teljesülnek.*

Ezen feltevések tehát biztosítják a szuperlineáris konvergenciát és azt is, hogy a kevesebb számolást igénylő GCG-LS(0) algoritmust használjuk. Ha az egyenletünknek homogén Dirichlet-peremfeltétele van, akkor nincs probléma  $D$  sűrűségével, ugyanis  $D(L) = D(L^*)$ . Vegyes peremfeltétel esetén  $D$  nem lesz sűrű az  $L^2(\Omega)$  térben. Általánosítsuk előző eredményeinket arra az esetre, ha nem tudunk minden feltevést teljesíteni. Ekkor nem mindig lehetséges erős értelemben definiálni az operátorokat.

Vizsgáljuk meg ismét az (4.1) egyenletet. Most  $S$ -ről csak szimmetriát teszünk fel az önadjungáltság helyett. Ekkor nincs szürjektivitás, azaz  $R(S) \neq H$ . Ekkor nem tudjuk közvetlenül értelmezni  $S^{-1}Q$ -t, hanem gyenge értelemben kell. Legyen  $Q_S$  a  $H_S$ -en értelmezett operátor, amelyet minden  $u \in H_S$ -re a

$$\langle Q_S u, v \rangle_S = \langle Qu, v \rangle \quad (\forall v \in H_S) \quad (4.6)$$

egyenlőség definiál<sup>2</sup>. Legyenek  $S$  és  $Q$  ugyanazok, mint eddig.

**4.5. Feltevések.** *Tegyük fel, hogy*

i) *Létezik  $p > 0$ , hogy  $\Re \langle Lu, u \rangle \geq p \|u\|^2$  ( $u \in D$ ),*

ii)  *$D = D(L) \cap D(L^*)$  sűrű  $H$ -ban,*

iii) *A  $Q$  operátor kiterjeszthető a  $H_S$  térbe és  $Q_S$  kompakt operátor  $H_S$ -en.*

<sup>2</sup> Ha  $R(S) = H$ , akkor  $Q_S = S^{-1}Q$ , azaz visszkapjuk a korábbi definíciót.

Mivel  $S$  szimmetrikus,  $Q$  pedig antiszimmetrikus operátor, ezért  $\Re \langle Qu, u \rangle = 0$ , tehát  $\langle Su, u \rangle = \Re \langle Su, u \rangle = \Re \langle Su, u \rangle + \Re \langle Qu, u \rangle = \Re \langle Lu, u \rangle \geq p \|u\|^2$ , azaz a  $H_S$  energiatér értelmezhető és a  $H$ -beli norma felülről becsülhető az energia-normával. Ennek következménye, hogy a  $Q_S$  operátor valóban létezik, ugyanis tetszőleges rögzített  $u \in H$ -ra a  $\psi_u : H_S \rightarrow \mathbb{C}$ ,  $\psi_u v := \langle Qu, v \rangle$  lineáris és korlátos operátort definiál  $H_S$ -ben, így a Riesz reprezentációs tételből adódik egy  $Q_S u \in H_S$  elem létezése, amelyre  $\psi_u v = \langle Q_S u, v \rangle_S$ .

Hasonlóan a korábbiakhoz, itt is megmutatható a 2.4. tétel megfelelője. Mivel minden  $u \in D$ -re

$$\Re \langle Au, u \rangle_S = \Re \langle (I + Q_S)u, u \rangle_S = \Re \langle (S + Q)u, u \rangle = \Re \langle Lu, u \rangle = \langle Su, u \rangle = \|u\|_S^2,$$

ez a sűrűség miatt  $H_S$ -en is teljesül. Emiatt  $A$  korlátos lineáris operátor  $H_S$ -en, amelynek van korlátos inverze. Mivel  $Q$  antiszimmetrikus  $H$ -ban, ezért minden  $u, v \in H_S$ -re  $\langle Q_S u, v \rangle_S = \langle Qu, v \rangle = -\langle u, Qv \rangle = -\langle u, Q_S v \rangle_S$ , tehát  $Q_S$  antiszimmetrikus lesz  $H_S$ -ben. Ezenkívül teljesül még, hogy  $\|u\|_A^2 = \Re \langle Au, u \rangle_S = \Re \langle Lu, u \rangle = \|u\|_S^2$ , valamint  $\|A^{-1}\|_S = 1$ , ezért a 4.3. tétel alakja az alábbi lesz.

**4.6. Tétel.** *Tegyük fel, hogy teljesülnek a 4.5. feltevések. Ekkor a konjugált gradiens-módszert a prekondicionált egyenletre alkalmazva fennáll az*

$$\left( \frac{\|e_k\|_S}{\|e_0\|_S} \right)^{1/k} \leq \frac{2}{k} \sum_{i=1}^k |\lambda_i(Q_S)| \xrightarrow{k \rightarrow \infty} 0 \quad (4.7)$$

becslés minden  $k \in \mathbb{N}$ -re, ahol  $\lambda_k(Q_S)$  a  $Q_S$  operátor rendezett sajátértékei.

Utoljára maradt az az eset, amikor semmit sem teszünk fel  $D = D(L) \cap D(L^*)$ -ről. Így az  $S$  operátor nem definiálható a szokott módon, hanem már ezt is gyengén tudjuk csak értelmezni. Ilyenkor elég lesz a gyengébb

$$\Re \langle Lu, u \rangle > 0 \quad (u \in D(L) \setminus \{0\}) \quad (4.8)$$

feltétel.

**4.7. Állítás.** *Az*

$$\langle u, v \rangle_S := \frac{1}{2} (\langle Lu, v \rangle + \langle u, Lv \rangle) \quad (u, v \in D(L)) \quad (4.9)$$

formula skalárszorzatot definiál  $D(L)$ -en.

Ez nyilvánvaló, hiszen az első és második változó szerinti linearitás és konjugált linearitás a definíció következménye, ugyanis két skalárszorzat összege. Hasonlóan a konjugált szimmetria is a definícióból jön, az egyetlen dolog, amit meg kell gondolni, a pozitivitás. Felhasználva (4.8)-et,  $\langle u, u \rangle_S = \Re \langle Lu, u \rangle > 0$ , ha  $u \neq 0$ . Ezzel a skalárszorzattal definiálhatjuk a  $H_S$  energiateret: legyen  $H_S$  az  $[D(L), \langle \cdot, \cdot \rangle_S]$  teljessé tétele. Most definiálnunk kell az  $L$ -nek megfelelő bilineáris formát, hiszen már a  $H_S$ -teret is külön kellett értelmeznünk az  $S$ -skalárszorzattal. Ehhez tegyük fel, hogy az  $(u, v) \mapsto \langle Lu, v \rangle$  leképezés  $S$ -korlátos, azaz létezik  $M > 0$ , hogy minden  $u, v \in D(L)$ -re  $|\langle Lu, v \rangle| \leq M \|u\|_S \|v\|_S$ .

**4.8. Állítás.** *Egyértelműen létezik egy olyan korlátos bilineáris forma  $H_S$ -en, melyre*

$$\langle u, v \rangle_L = \langle Lu, v \rangle \quad (u, v \in D(L)), \quad (4.10)$$

*és az előző becslés kiterjeszhető  $H_S$ -re, továbbá létezik egyértelműen egy  $Q_S : H_S \rightarrow H_S$  operátor, amelyre*

$$\langle Q_S u, v \rangle_S = \frac{1}{2} (\langle u, v \rangle_L - \overline{\langle v, u \rangle_L}) \quad (\forall v \in H_S). \quad (4.11)$$

Ezen operátor létezése hasonlóan az eddigiekhez a reprezentációs tétel következménye a  $\psi_u v = \frac{1}{2} (\langle u, v \rangle_L - \overline{\langle v, u \rangle_L})$  lineáris és a  $H_S$ -normában folytonos funkcionálra alkalmazva. Így az egyenletünk új - gyenge - alakja a következő lesz: keressük azt az  $u \in H_S$ -et, amelyre

$$\langle u, v \rangle_L = \langle u, v \rangle_S + \langle Q_S u, v \rangle_S = \langle f, v \rangle_S = \langle g, v \rangle \quad (\forall v \in H_S), \quad (4.12)$$

azaz

$$(I + Q_S)u = f. \quad (4.13)$$

**4.9. Feltevések.** *Tegyük fel, hogy*

*i) Létezik  $p > 0$ , hogy  $\Re \langle Lu, u \rangle \geq p \|u\|^2 \quad (u \in D(L))$ ,*

*ii) Létezik  $M > 0$ , hogy  $|\langle Lu, v \rangle|^2 \leq M^2 \Re \langle Lu, u \rangle \Re \langle Lv, v \rangle \quad (u, v \in D(L))$ ,*

*iii) A  $Q_S$  operátor kompakt  $H_S$ -en.*

Most is az előzőekhez teljesen hasonlóan látható be, hogy ezen feltételekből következik, hogy  $A = I + Q_S$ -nek létezik korlátos inverze, létezik továbbá olyan  $f \in H_S$  elem, amelyre  $\langle f, v \rangle_S = \langle g, v \rangle$ , a  $Q_S$  antiszimmetrikus - és így egyben  $S$ -normális - valamint kompakt  $H_S$ -ben. Az algoritmus alakját érdemes felírni, hiszen a prekondicionált egyenlet iterációs megoldása már a  $H_S$  térben zajlik. Figyelembevéve az  $\langle \cdot, \cdot \rangle_S$  és  $\langle \cdot, \cdot \rangle_L$  jelentését az (4.9) és (4.10) egyenletekből, GCG-LS(0) alakja:

$$\left\{ \begin{array}{l} (1) \quad \text{Legyen } u_0 \in H_S \text{ tetszőleges és legyen } r_0 \in H_S \text{ az} \\ \quad \langle r_0, v \rangle_S = \langle u_0, v \rangle_L - \langle g, v \rangle \quad (\forall v \in H_S) \\ \quad \text{egyenlet megoldása, } d_0 = -r_0; \\ \quad \text{Tetszőleges } k \in \mathbb{N} \text{ esetén ha } u_k, d_k, r_k \text{ már megvan, akkor legyen} \\ (2a) \quad p_k \in H_S \text{ a } \langle p_k, v \rangle_S = \langle d_k, v \rangle_L \quad (\forall v \in H_S) \text{ egyenlet megoldása,} \\ \quad \gamma_k = \|p_k\|_S^2, \quad \alpha_k = -\frac{1}{\gamma_k} \langle p_k, r_k \rangle_S; \\ (2b) \quad u_{k+1} = u_k + \alpha_k d_k; \\ (2c) \quad r_{k+1} = r_k + \alpha_k p_k; \\ (2d) \quad \beta_k = \frac{1}{\gamma_k} \langle r_{k+1}, p_k \rangle_L; \\ (2e) \quad d_{k+1} = -r_{k+1} + \beta_k d_k. \end{array} \right. \quad (4.14)$$

Ismét fennáll a 2.4. tétel megfelelője, azaz

**4.10. Tétel.** *Tegyük fel, hogy teljesülnek a 4.9. feltevések. Ekkor a konjugált gradiens-módszert a prekondicionált (4.13) egyenletre alkalmazva fennáll a*

$$\left( \frac{\|e_k\|_S}{\|e_0\|_S} \right)^{1/k} \leq \frac{2}{k} \sum_{i=1}^k |\lambda_i(Q_S)| \xrightarrow{k \rightarrow \infty} 0 \quad (4.15)$$

*becslés minden  $k \in \mathbb{N}$ -re, ahol  $\lambda_k(Q_S)$  a  $Q_S$  operátor rendezett sajátértékei.*

## 5. KONVEKCIÓ-DIFFÚZIÓ EGYENLETEK

Ebben a fejezetben az eddig elhangzott elméleti eredményeket alkalmazzuk konvekció-diffúzió egyenletekre, ahol tehát a  $H$  Hilbert-tér egy Szoboljev-tér lesz. Homogén Dirichlet-peremfeltétel esetén az egyenlet szimmetrikus részével történő prekondicionálást vizsgálták [4]-ben, a vegyes peremfeltétel esetét illetően lásd [10]. Most ez utóbbit kövejük, mert speciális esetként megkapjuk a Dirichlet-peremfeltétel esetét is. Később a numerikus kísérletek tárgyalásánál mindkét típusú peremfeltételt megvizsgáljuk. Először a konstansgyűtthetős eset következik, majd utána a függvénygyűtthetős. Az elvégzett numerikus kísérletek nagyrésze a speciális feladathoz kapcsolódik. Az általános problémához tartozó algoritmus programozása sem okoz nehézséget, de fő célunk a szuperlineáris konvergencia megmutatása és az  $e_k = u_k - u^*$  hibavektor számítása bonyodalmakat okozhat, amit csak a konstansgyűtthetős esetben sikerült kikerülni. Erre később visszatérünk.

### 5.1. Egy speciális egyenlet

Tekintsük tehát a következő egyenletet:

$$\begin{cases} Lu \equiv -\Delta u + \mathbf{b} \cdot \nabla u + cu = g \\ u|_{\Gamma_D} = 0, \quad \frac{\partial u}{\partial \nu} + \phi u|_{\Gamma_N} = 0. \end{cases} \quad (5.1)$$

Itt az egyűtthetőkra egyszerű feltételeket teszünk, később azonban általánosabb esetet is vizsgálunk. A speciális feladat esetén a következő feltevéseket tesszük:

#### 5.1. Feltevések (Speciális probléma). *Tegyük fel, hogy*

*i)  $\Omega \subset \mathbb{R}^d$  korlátos tartomány, amelynek pereme szakaszonként folytonosan differenciálható,  $\Gamma_D$  és  $\Gamma_N$  olyan nyílt, mérhető részhalmazai a tartomány peremének, hogy  $\partial\Omega = \bar{\Gamma}_D \cup^* \bar{\Gamma}_N$ ,*

*ii)  $\mathbf{b} \in C^1(\bar{\Omega})^d$  divergenciamentes, továbbá  $\mathbf{b} \cdot \nu = 0$   $\Gamma_N$ -en,*

*iii)  $c, \phi \geq 0$  konstansok,*

*iv)  $g \in L^2(\Omega)$ ,*

*v)  $\Gamma_D \neq \emptyset$  vagy  $c > 0$  vagy  $\phi > 0$ .*

**5.2. Megjegyzés.** *A feltételek közül valójában a ii) fontos, ez biztosítja, hogy  $D(L) = D(L^*)$ ;  $\alpha$  és  $\phi$  lehetnének akár függvények is. A (6.4) tesztyenletek azonban ilyen alakúak lesznek, ezért érdemes lehet a fenti 5.1. feltevéseknél maradni.*

Legyen a  $H$  Hilbert-tér az  $L^2(\Omega)$  a szokásos

$$\langle u, v \rangle_{L^2} = \int_{\Omega} u \bar{v} \, dx$$

skalárszorozattal, valamint legyen

$$H_D^1(\Omega) := \{u \in H^1(\Omega) : u|_{\Gamma_D} = 0\} \quad (5.2)$$

ellátva az

$$\langle u, v \rangle_{H_D^1} := \int_{\Omega} (\nabla u \cdot \nabla \bar{v} + cu \bar{v}) + \int_{\Gamma_N} \phi u \bar{v} \, d\sigma, \quad (5.3)$$

skalárszorozattal.

Az  $L$  operátor legyen

$$Lu := -\Delta u + \mathbf{b} \cdot \nabla u + cu,$$

amelynek

$$D(L) := \left\{ u \in H^2(\Omega) : u|_{\Gamma_D} = 0, \frac{\partial u}{\partial \nu} + \phi u|_{\Gamma_N} = 0 \right\}$$

értelmezési tartománya sűrű  $H_D^1$ -ben.

Az  $L$  operátor szimmetrikus részének konstruálásához használjuk fel az (1.5) összefüggést, igaz, kissé más szereposztásban; elvégezve az  $u \rightsquigarrow uv$  és  $v \rightsquigarrow \mathbf{b}$  helyettesítéseket, továbbá figyelembevéve azt, hogy komplex Hilbert-térben vagyunk, az

$$\int_{\Omega} (\operatorname{div} \mathbf{b}) u \bar{v} = \int_{\partial \Omega} (\mathbf{b} \cdot \nu) u \bar{v} \, d\sigma - \int_{\Omega} \nabla(u \bar{v}) \cdot \mathbf{b} \quad (u, v \in H_D^1(\Omega)) \quad (5.4)$$

azonosságot kapjuk. Vegyük észre, hogy a peremintegrál valójában csak  $\Gamma_N$ -en van, a függvények ugyanis eltűnnek  $\Gamma_D$ -n. Az egyenlet utolsó tagját az (1.6) formula alapján kibontva az

$$\int_{\Omega} (\mathbf{b} \cdot \nabla u) \bar{v} = - \int_{\Omega} u (\mathbf{b} \cdot \nabla \bar{v}) - \int_{\Omega} (\operatorname{div} \mathbf{b}) u \bar{v} + \int_{\Gamma_N} (\mathbf{b} \cdot \nu) u \bar{v} \, d\sigma \quad (u, v \in H_D^1(\Omega)) \quad (5.5)$$

azonosságot kapjuk, amelyben a két utolsó tag a feltételek szerint 0, azaz a 5.1. feltevések mellett az (5.5) egyenlet

$$\int_{\Omega} (\mathbf{b} \cdot \nabla u) \bar{v} = - \int_{\Omega} u (\mathbf{b} \cdot \nabla \bar{v}) \quad (u, v \in D(L)) \quad (5.6)$$

alakú, amiből következik, hogy  $L$  szimmetrikus és antiszimmetrikus része

$$Su \equiv -\Delta u + cu, \quad Qu \equiv \mathbf{b} \cdot \nabla u \quad (u, v \in D(L)).$$

A (4.6) egyenlőség értelmében definiáljuk a  $Q_S : H_D^1(\Omega) \rightarrow H_D^1(\Omega)$  operátort a

$$\langle Q_S u, v \rangle_{H_D^1} = \int_{\Omega} (\mathbf{b} \cdot \nabla u) \bar{v} \quad (u, v \in H_D^1(\Omega)) \quad (5.7)$$

azonossággal. Most vegyük elő a (4.14) algoritmust és ezen feltételek mellett alkalmazzuk.

$$\left\{ \begin{array}{l} (1) \quad \text{Legyen } u_0 \in H_D^1(\Omega) \text{ tetszőleges és legyen } r_0 \in H_D^1(\Omega) \text{ az} \\ \quad \quad \quad \begin{cases} -\Delta r_0 + cr_0 = Lu_0 - g \\ r_0|_{\Gamma_D} = 0, \quad \frac{\partial r_0}{\partial \nu} + \phi r_0|_{\Gamma_N} = 0 \end{cases} \\ \quad \quad \quad \text{egyenlet gyenge megoldása, } d_0 := -r_0; \\ \quad \quad \quad \text{Tetszőleges } k \in \mathbb{N} \text{ esetén ha } u_k, d_k, r_k \text{ már megvan, akkor legyen} \\ (2a) \quad p_k \in H_D^1(\Omega) \text{ a következő egyenlet gyenge megoldása} \\ \quad \quad \quad \begin{cases} -\Delta p_k + cp_k = Ld_k \\ p_k|_{\Gamma_D} = 0, \quad \frac{\partial p_k}{\partial \nu} + \phi p_k|_{\Gamma_N} = 0 \end{cases} \\ \quad \quad \quad \gamma_k = \|p_k\|_{H_D^1}^2, \quad \alpha_k = -\frac{1}{\gamma_k} \langle p_k, r_k \rangle_{H_D^1}; \\ (2b) \quad u_{k+1} = u_k + \alpha_k d_k; \\ (2c) \quad r_{k+1} = r_k + \alpha_k p_k; \\ (2d) \quad \beta_k = \frac{1}{\gamma_k} \langle r_{k+1}, p_k \rangle_L; \\ (2e) \quad d_{k+1} = -r_{k+1} + \beta_k d_k. \end{array} \right. \quad (5.8)$$

Az egyenletre tett 5.1. feltevések biztosítják, hogy teljesülnek az általános Hilbert-térbeli 4.5. kikötések (lásd [10]), így fennáll az

**5.3. Tétel.** *Tegyük fel, hogy teljesülnek az 5.1. feltevések. Ekkor a prekondicionált konjugált gradiens-módszert az (5.1) egyenletre alkalmazva fennáll az*

$$\left( \frac{\|e_k\|_{H_D^1}}{\|e_0\|_{H_D^1}} \right)^{1/k} \leq \frac{2}{k} \sum_{i=1}^k |\lambda_i(Q_S)| \xrightarrow{k \rightarrow \infty} 0 \quad (5.9)$$

becslés minden  $k \in \mathbb{N}$ -re, ahol  $\lambda_k(Q_S)$  a  $Q_S$  operátor rendezett sajátértékei.

## 5.2. Az általános egyenlet

Ebben a részben megengedünk függvényegyütthatókat is, valamint nem követeljük meg a  $\mathbf{b}$  vektormező divergencia-mentességét, sem pedig a merőlegességet a Neumann-perem egységnormálisára. Legyen tehát a vizsgált egyenlet általános esetben

$$\begin{cases} Lu \equiv -\Delta u + \mathbf{b} \cdot \nabla u + cu = g \\ u|_{\Gamma_D} = 0, \quad Pu \equiv \frac{\partial u}{\partial \nu} + \phi u|_{\Gamma_N} = 0. \end{cases} \quad (5.10)$$

alakú, amelyre teljesülnek az alábbi

**5.4. Feltevések (Általános probléma).** Tegyük fel, hogy

i)  $\Omega \subset \mathbb{R}^d$  korlátos tartomány, amelynek pereme szakaszonként folytonosan differenciálható,  $\Gamma_D$  és  $\Gamma_N$  olyan nyílt, mérhető részhalmazai a tartomány peremének, hogy  $\partial\Omega = \overline{\Gamma_D} \cup^* \overline{\Gamma_N}$ ,

ii)  $\mathbf{b} \in C^1(\overline{\Omega})^d$ ,  $c \in L^\infty(\Omega)$ ,  $\phi \in L^\infty(\Gamma_N)$ , és  $c, \phi \geq 0$ ,

iii)  $\hat{c} := c - \frac{1}{2} \operatorname{div} \mathbf{b} \geq 0$   $\Omega$ -n és  $\hat{\phi} := \phi + \frac{1}{2}(\mathbf{b} \cdot \nu) \geq 0$   $\Gamma_N$ -en,

iv)  $g \in L^2(\Omega)$ ,

v)  $\Gamma_D \neq \emptyset$ , vagy  $\hat{c}$  vagy  $\hat{\phi}$  nem azonosan nulla.

Most a (4.9) és (5.5) egyenleteket figyelembevéve adódik, hogy

$$\begin{aligned} \langle u, v \rangle_S &= \frac{1}{2} (\langle Lu, v \rangle_{L^2} + \langle u, Lv \rangle_{L^2}) = \\ &= \int_{\Omega} \left( \nabla u \cdot \nabla \bar{v} + \left( c - \frac{1}{2} \operatorname{div} \mathbf{b} \right) u \bar{v} \right) dx + \int_{\Gamma_N} \left( \phi + \frac{1}{2} (\mathbf{b} \cdot \nu) \right) u \bar{v} d\sigma = \quad (5.11) \\ &= \int_{\Omega} (\nabla u \cdot \nabla \bar{v} + \hat{c} u \bar{v}) dx + \int_{\Gamma_N} \hat{\phi} u \bar{v} d\sigma \quad (u, v \in D(L)). \end{aligned}$$

A korábbiakhoz hasonlóan

$$\begin{aligned} \Re \langle Lu, u \rangle &= \langle u, u \rangle_S = \frac{1}{2} (\langle Lu, u \rangle_{L^2} + \langle u, Lu \rangle_{L^2}) = \\ &= \int_{\Omega} (|\nabla u|^2 + \hat{c} |u|^2) dx + \int_{\Gamma_N} \hat{\phi} |u|^2 d\sigma > 0. \quad (5.12) \end{aligned}$$

Ugyanúgy, mint prekondicionálásnál az általános esetben, a most bevezetett  $S$ -skalárszorzzattal ellátott  $D(L)$ -et teljessé tesszük, ez lesz a  $H_S$ , a  $Q_S$  operátort pedig a (4.11) azonosság értelmében a következő módon kell definiálnunk:

$$\langle Q_S u, v \rangle_S = \frac{1}{2} \left( \int_{\Omega} (\mathbf{b} \cdot \nabla u) \bar{v} dx - \int_{\Omega} u (\mathbf{b} \cdot \nabla \bar{v}) dx \right) \quad (5.13)$$

amit (5.5) miatt átalakíthatunk

$$\begin{aligned} &\frac{1}{2} \left( \int_{\Omega} (\mathbf{b} \cdot \nabla u) \bar{v} dx - \int_{\Omega} u (\mathbf{b} \cdot \nabla \bar{v}) dx \right) = \\ &= \int_{\Omega} (\mathbf{b} \cdot \nabla u) \bar{v} dx - \frac{1}{2} \int_{\Gamma_N} (\mathbf{b} \cdot \nu) u \bar{v} d\sigma + \frac{1}{2} \int_{\Omega} (\operatorname{div} \mathbf{b}) u \bar{v} dx = \quad (5.14) \\ &= - \int_{\Omega} u (\mathbf{b} \cdot \nabla \bar{v}) dx + \frac{1}{2} \int_{\Gamma_N} (\mathbf{b} \cdot \nu) u \bar{v} d\sigma - \frac{1}{2} \int_{\Omega} (\operatorname{div} \mathbf{b}) u \bar{v} dx. \end{aligned}$$

Ismét a (4.14) algoritmust szeretnénk alkalmazni. Ebben a helyzetben az  $\langle r_0, v \rangle_S = \langle u_0, v \rangle_L - \langle g, v \rangle$  ( $\forall v \in H_S$ ) részfeladat - felhasználva  $\langle \cdot, \cdot \rangle_S$  és  $\langle \cdot, \cdot \rangle_L$  integrálos alakját - a következő feladat gyenge megoldásával egyenértékű:



$$\begin{cases} -\Delta r_0 + \hat{c}r_0 = Lu_0 - g \\ r_0|_{\Gamma_D} = 0, \quad \frac{\partial r_0}{\partial \nu} + \hat{\phi}r_0|_{\Gamma_N} = \frac{\partial u_0}{\partial \nu} + \phi u_0, \end{cases} \quad (5.15)$$

azaz nem az eredeti  $c$  és  $\phi$  együtthatókkal dolgozik az (1) lépés, sőt az iterációs lépés sem, hanem a korábban bevezetett  $\hat{c}$  és  $\hat{\phi}$  együtthatókkal<sup>1</sup>.

Ezeket a változásokat figyelembevéve, a (4.14) algoritmus alakja az általános konvekció-diffúzió egyenletre a következő alakot ölti.

$$\left\{ \begin{array}{l} (1) \quad \text{Legyen } u_0 \in H_D^1(\Omega) \text{ tetszőleges és legyen } r_0 \in H_D^1(\Omega) \text{ az} \\ \quad \begin{cases} -\Delta r_0 + \hat{c}r_0 = Lu_0 - g \\ r_0|_{\Gamma_D} = 0, \quad \frac{\partial r_0}{\partial \nu} + \hat{\phi}r_0|_{\Gamma_N} = Pu_0 \end{cases} \\ \quad \text{egyenlet gyenge megoldása, } d_0 := -r_0; \\ \quad \text{Tetszőleges } k \in \mathbb{N} \text{ esetén ha } u_k, d_k, r_k \text{ már megvan, akkor legyen} \\ (2a) \quad p_k \in H_D^1(\Omega) \text{ a következő egyenlet gyenge megoldása} \\ \quad \begin{cases} -\Delta p_k + \hat{c}p_k = Ld_k \\ p_k|_{\Gamma_D} = 0, \quad \frac{\partial p_k}{\partial \nu} + \hat{\phi}p_k|_{\Gamma_N} = Pd_k \end{cases} \\ \quad \gamma_k = \|p_k\|_S^2, \quad \alpha_k = -\frac{1}{\gamma_k} \langle p_k, r_k \rangle_S; \\ (2b) \quad u_{k+1} = u_k + \alpha_k d_k; \\ (2c) \quad r_{k+1} = r_k + \alpha_k p_k; \\ (2d) \quad \beta_k = \frac{1}{\gamma_k} \langle r_{k+1}, p_k \rangle_L; \\ (2e) \quad d_{k+1} = -r_{k+1} + \beta_k d_k. \end{array} \right. \quad (5.16)$$

Ismét érvényes a szuperlineáris konvergenciáról szóló tétel, azzal a különbséggel, hogy az  $e_k$  vektoroknak nem a  $H_D^1$ -normája, hanem a „kalapos” együtthatókat tartalmazó  $S$ -normája<sup>2</sup> szerepel.

**5.5. Tétel.** *Tegyük fel, hogy teljesülnek a 5.4. feltevések. Ekkor a prekondicionált konjugált gradiens-módszert az (5.10) egyenletre alkalmazva fennáll az*

$$\left( \frac{\|e_k\|_S}{\|e_0\|_S} \right)^{1/k} \leq \frac{2}{k} \sum_{i=1}^k |\lambda_i(Q_S)| \xrightarrow{k \rightarrow \infty} 0 \quad (5.17)$$

becslés minden  $k \in \mathbb{N}$ -re, ahol  $\lambda_k(Q_S)$  a  $Q_S$  operátor rendezett sajátértékei.

A bizonyításhoz - ami az 5.4. feltételeket használja ahhoz, hogy ellenőrizze a 4.9 tulajdonságok teljesülését - lásd ismét [10]-et, az egyszerűbb, homogén Dirichlet-peremfeltételes feladat részletes tárgyalásához [3, 4]-et.

<sup>1</sup> A speciális feladat esetén a pluszfeltételek - divergenciamentesség és egy ortogonalitási tulajdonság - miatt  $\hat{c} = c$  és  $\hat{\phi} = \phi$ .

<sup>2</sup> Hiszen most az  $S$  operátor máshogy néz ki, mint a speciális eset  $S$ -operátora.

### 5.3. Egyéb prekondicionáló eljárások

Természetesen sok egyéb módszer is elképzelhető, azonban ezek nagyrésze a lineáris konvergenciát tűzi ki célul, azaz a diszkrétizált egyenlet hibájára valamilyen

$$\|e_k\| \leq r^k \|e_0\|$$

egyenlőtlenség teljesülését szeretnénk elérni, ahol a  $\|\cdot\|$  norma egy energianorma diszkrét megfelelője,  $0 \leq r < 1$  egy a  $h$  rácsállandótól független konstans. Ilyen kérdéseket segít megválaszolni az ekvivalens operátorok elmélete (lásd [6, 8, 9]).

Ismeretes, hogy szimmetrikus esetben a konjugált gradiens-módszer hibájára érvényes a következő becslés:

$$\frac{\|e_k\|_A}{\|e_0\|_A} \leq 2 \left( \frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1} \right)^k, \quad (5.18)$$

ahol  $\kappa$  az  $A$  szimmetrikus operátor kondíciószáma. A prekondicionált egyenlet esetén ekkor  $\kappa$  helyett más, de szintén egy mátrix kondíciószáma fog állni. Ahhoz, hogy ez korlátos maradjon, ha finomítjuk a felosztást, szükséges, hogy a végtelen dimenziós térben ekvivalens operátorral<sup>3</sup> prekondicionáljunk. Az alábbi tételek megtalálhatóak [12]-ben.

**5.6. Tétel.** *Legyenek  $L$  és  $S$  invertálható másodrendű lineáris elliptikus operátorok. Legyen  $\{L_h\}$  és  $\{S_h\}$  az operátorok végeselem diszkrétizációjából származó mátrixok családja és tegyük fel, hogy  $L_h^{-1} \rightarrow L^{-1}$ , valamint  $S_h^{-1} \rightarrow S^{-1}$  pontonként  $L^2$ -ben. Tegyük fel, hogy létezik  $K > 0$   $h$ -tól független konstans, hogy*

$$\|S_h^{-1}L_h\|_{L^2} \leq K.$$

*Ekkor létezik  $K'$ , hogy*

$$\|S^{-1}L\|_{L^2} \leq K'. \quad (5.19)$$

**5.7. Tétel.** *Az előző tétel feltételei mellett, ha most  $(S^*)_h^{-1} \rightarrow (S^*)^{-1}$  és  $(L^*)_h^{-1} \rightarrow (L^*)^{-1}$ , továbbá létezik  $M > 0$ , hogy*

$$\|L_h S_h^{-1}\|_{L^2} \leq M,$$

*akkor létezik  $M' > 0$ , hogy*

$$\|L S^{-1}\|_{L^2} \leq M'. \quad (5.20)$$

Feltételezve a feladat  $H^2$ -regularitását, azaz hogy létezik  $K_L > 0$ , hogy minden  $g \in L^2(\Omega)$ -ra létezik a problémának egyértelmű  $u \in H^2(\Omega)$  disztribúciós megoldása, amelyre

$$\|u\|_{H^2(\Omega)} \leq K_L \|g\|_{L^2(\Omega)}, \quad (5.21)$$

a következő tételt specializálhatjuk az általunk vizsgált (5.10) egyenletre:

<sup>3</sup> Ez azonban nem elég és a gyakorlatban nehezen kezelhető. [6]-ben bevezetnek egy félmetrikát is az ekvivalenciaosztályokon belül, amellyel az operátorhoz egy vele ekvivalens, hozzá „közeli” operátor választható prekondicionernek, ám ez gyakorlati szempontból használhatatlan.

**5.8. Tétel.**  $L$  és  $S$ -re pontosan akkor teljesül (5.19), ha  $L^*$ -nak és  $S^*$ -nak ugyanolyan peremfeltételei vannak, ahol az adjungáltat  $L^2$ -értelemben vesszük. Továbbá (5.20) pontosan akkor igaz, ha  $L$  és  $S$ -nek ugyanazok a peremfeltételei.

A vizsgált egyenletnél azonban energianorma - illetve a programban ennek egy diszkrét megfelelője - szerepel, ami  $c = 1$  esetben a  $H^1$ -norma lesz, más  $c \geq 0$ -ra pedig ekvivalens vele. Ilyen esetben az  $L$  és az  $S$  prekondicioner viszonya a következő kell, hogy legyen:

**5.9. Tétel.** Az  $S^{-1}L$  operátor  $H^1$ -kondíciószáma a  $h$  rácsállandótól független számmal felülről korlátozható pontosan akkor, ha  $L$ -nek és  $S$ -nek a  $\partial\Omega$  perem ugyanazon részén van Dirichlet-peremfeltétel megadva.

Egyéb tipikus prekondicioner választás lehet, ha az  $L$  operátort nem az  $S$  szimmetrikus részével, hanem  $S + \sigma I$ -vel prekondicionáljuk, ahol  $\sigma \geq 0$ . Ez gyakran előnyösebb lehet, mint csupán  $S$  használata; ha  $L$ -ben az elsőrendű tag együtthatója nagy, akkor érdemes  $\sigma \gg 0$  értéket választani. Noha ezt az eljárást elsősorban a lineáris konvergencia javításánál alkalmazták (pl. [11]), a mi példaegyenletünk bizonyos egyszerű eseteiben a szuperlineáris konvergencia is teljesül. Erre térünk ki itt röviden, részletesen lásd [4]-et.

Tekintsük a speciális (5.1) egyenlet speciális esetét, azaz legyen a  $\mathbf{b}$  vektor konstans-elemű,  $c \in \mathbb{R}$  és  $\Gamma_N = \emptyset$ .

$$\begin{cases} Lu \equiv -\Delta u + \mathbf{b} \cdot \nabla u + cu = g \\ u|_{\partial\Omega} = 0. \end{cases} \quad (5.22)$$

Így az 5.1 feltételek teljesüléséhez mindössze a  $c \geq 0$ -t kell megkövetelnünk. Legyen  $D(L) = H^2(\Omega) \cap H_0^1(\Omega)$ , a prekondicionáló operátor pedig

$$Su = -\Delta u + \sigma u \quad (u \in D(L)) \quad (5.23)$$

valamilyen  $\sigma \geq 0$  számra. Legyen  $d = \sigma - c$  és

$$Qu = \mathbf{b} \cdot \nabla u - du \quad (u \in D(L)). \quad (5.24)$$

**5.10. Állítás.** Az  $L$  és  $S$  operátorokra teljesülnek a 4.1. feltevések.

Ezek ellenőrzése egyszerű, az első két tulajdonság a definícióból adódik, a ii)-hoz  $\rho = \min\{\frac{c}{\sigma}, 1\}$  és végül  $S^{-1}Q$  kompaktsága is teljesül. A normalitásért viszont meg kell küzdeni és itt bonyolódik a dolog, ami miatt a függvényegyütthatós esetre ez nem megy. Ebben az esetben viszont ez igaz és fennáll a szuperlineáris konvergenciát biztosító 4.3. tétel megfelelője: a  $2/\rho$  konstanszt kell kicserélni  $2 \cdot \max\{\frac{\sigma}{c}, 1\}$ -re.

## 6. NUMERIKUS KÍSÉRLETEK

Ebben a fejezetben alkalmazunk minden korábbi eredményt. Az (5.1) és (5.10) egyenlet-típusok mindegyikét teszteljük a (5.8) és (5.16) algoritmusok futtatásával konkrét egyenletekre. Kipróbálunk a 5.3. fejezetben szereplő (5.23) alakú prekondicionereket és numerikusan is megmutatjuk a 5.9. tétel fontosságát. Előtte viszont egy lényeges dolog hátravan még. Szuperlineáris konvergenciát szeretnénk megmutatni a numerikus kísérletek során is, ekkor viszont a (5.9) és (5.17) becslések bal oldalán nem folytonos, hanem diszkrét normák állnak, amelyek függhetnek a diszkretizációs paramétertől és persze az iterációs lépés sem a  $H_D^1(\Omega)$  térben, hanem a  $V_h$  végelem-altérben történik.

### 6.1. A diszkrét probléma rácsfüggetlen szuperlineáris konvergenciája

Legyen  $V_h = \text{span}\{w_1, w_2, \dots, w_n\} \subset H_S$  egy végelem altér és tekintsük az (5.10) egyenlet végelem diszkretizációját. Keressük ebben az altérben az  $u_h$  megoldást  $u_h = \sum_{i=1}^n u_i w_i$  alakban. Ekkor a feladat diszkrét megoldását az

$$\mathbf{L}_h \mathbf{u} = \mathbf{g} \tag{6.1}$$

véges dimenziós lineáris egyenletrendszer adja, ahol

$$(\mathbf{L}_h)_{i,j} = \int_{\Omega} (\nabla w_i \nabla \overline{w_j} + (\mathbf{b} \cdot \nabla w_j) \overline{w_i} + c w_i w_j) \, dx + \int_{\Gamma_N} \phi w_i w_j \, d\sigma$$

és  $\mathbf{g}_j = \int_{\Omega} g \overline{w_j}$ . Ekkor (5.12) miatt (6.1)-nek egyértelműen létezik megoldása. Itt is a természetes módon definiálhatjuk a mátrix szimmetrikus és antiszimmetrikus részét:

$$\mathbf{S}_h = \frac{1}{2} (\mathbf{L}_h + \mathbf{L}_h^*), \quad \mathbf{Q}_h = \mathbf{L}_h - \mathbf{S}_h.$$

A prekondicionált egyenlet alakja

$$\mathbf{S}_h^{-1} \mathbf{L}_h \mathbf{u} = (\mathbf{I}_h + \mathbf{S}_h^{-1} \mathbf{Q}_h) \mathbf{u} = \mathbf{S}_h^{-1} \mathbf{g}.$$

A következő tétel biztosítja, hogy fennáll a rácsfüggetlenség<sup>1</sup>.

**6.1. Tétel.** *Ha teljesülnek az 5.4 feltételek, akkor*

$$\sum_{m=1}^k |\lambda_m(\mathbf{S}_h^{-1} \mathbf{Q}_h)| \leq \sum_{m=1}^k |\lambda_m(Q_S)| \quad (k = 1, 2, \dots, n), \tag{6.2}$$

---

<sup>1</sup> A bizonyításhoz lásd [4, 10].

ami - figyelembevéve a véges dimenzióban is érvényes becslést - azt jelenti, hogy

$$\left( \frac{\|e_k\|_{\mathbf{S}_h}}{\|e_0\|_{\mathbf{S}_h}} \right)^{1/k} \leq \frac{2}{k} \sum_{m=1}^k |\lambda_m(Q_S)| \xrightarrow{k \rightarrow \infty} 0, \quad (6.3)$$

ahol a jobb oldalon álló nullsorozat független a  $V_h$  altértől.

Dirichlet-peremfeltétel esetén az is belátható, hogy ha nem a szimmetrikus résszel prekondicionálunk, hanem  $\mathbf{S}_h + \sigma \mathbf{I}_h$ -val, akkor ha  $\mathbf{S}_h^{-1} \mathbf{Q}_h \mathbf{S}_h$ -normális és a konstansgyűthetős (5.22) feladatról van szó, szintén teljesül a rácsfüggetlenség.

**6.2. Megjegyzés.** A becslés érvényességéhez hallgatólagosan feltettük, hogy az  $\mathbf{S}_h$  mátrixú segédfeladatokat egzaktul oldjuk meg, de ez nem szükséges. Ha az iterációs lépés egyenleteit  $\eta$  hibával oldjuk meg, akkor  $\|\tilde{e}_k\| = (1 + \eta)^k \|e_k\|$ -t kell tekintenünk, de erre is érvényes a

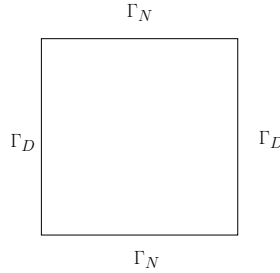
$$\left( \frac{\|\tilde{e}_k\|_{\mathbf{S}_h}}{\|\tilde{e}_0\|_{\mathbf{S}_h}} \right)^{1/k} \leq (1 + \eta) \cdot \epsilon_k \xrightarrow{k \rightarrow \infty} 0$$

szuperlineáris konvergenciát adó becslés, ahol  $\epsilon_k$  a (6.3) jobb oldalán álló nullához tartó sorozat.

## 6.2. A vizsgált tesztfeladatok

Végül élesben is kipróbáljuk az elméletet. Két tesztfeladatra futtatuk le az algoritmust, egyik feladat homogén Dirichlet peremfeltétellel rendelkezik, a másik vegyes peremfeltételű, de szintén homogén, mindkettő az egységnyízetten. A vegyes peremfeltétel esetén a Dirichlet-, illetve Neumann-peremek az ábrának megfelelően helyezkednek el.

6.1. ábra. A peremek elhelyezkedése vegyes peremfeltétel esetén



A cél nemcsak az, hogy az algoritmust ellenőrizzük és szuperlineáris konvergenciát mutassunk, hanem a kapott értékeket összehasonlítsuk a becslések jobb oldalán álló számtani-közép sorozat értékeivel. Ehhez pedig a  $Q_S$  operátor sajátértékeire van szükségünk. Ennek kiszámítására viszont csak egyszerű egyűthetők esetén van esélyünk.

A következő két feladatra készültek részletes futási eredmények:

$$\begin{cases} Lu \equiv -\Delta u + \frac{\partial u}{\partial x} + cu = g \\ (a) \quad u|_{\partial\Omega} = 0. \\ (b) \quad u|_{\Gamma_D} = 0, \quad \frac{\partial u}{\partial \nu}|_{\Gamma_N} = 0. \end{cases} \quad (6.4)$$

Mindkét feladat az (5.1) egyenlet speciális esete és teljesülnek az (5.1) feltevések, ugyanis  $\Gamma_D \neq \emptyset$ ,  $\mathbf{b} = (1, 0)$ , ami konstans lévén divergenciamentes és teljesül a merőlegességi feltétel a Neumann-peremek  $(0, \pm 1)$  normálvektoraira is. Jelölje  $u^*$  a feladat pontos megoldását, ekkor az (a) feladatra az  $u^*$  és ebből adódóan a  $g$  jobb oldali függvény értéke

$$\begin{aligned} u^*(x, y) &= (x - x^2)(y - y^2) \\ g(x, y) &= 2(y - y^2 + x - x^2) + (1 - 2x)(y - y^2) + c(x - x^2)(y - y^2), \end{aligned}$$

míg a (b) feladat estén

$$\begin{aligned} u^*(x, y) &= (x - x^2)(3y^2 - 2y^3) \\ g(x, y) &= 2y^2(3 - 2y) + (12y - 6)(x - x^2) + (1 - 2x)(3y^2 - 2y^3) + \\ &\quad + cx(1 - x)(3y^2 - 2y^3). \end{aligned}$$

Az (5.16) algoritmust fogjuk alkalmazni, hogy az iterációval egy  $u_1, u_2, \dots$  sorozatot nyerjük, hogy az 5.5. tételben szereplő hibavektorok normáit kiszámoljuk. Valójában azonban a 6.2. tételt alkalmazzuk, hiszen az algoritmus egy  $V_h$  végelem altérben fut, így adott felosztás esetén egy  $u_{h,1}, u_{h,2}, \dots$  sorozatot kapunk, ami az eredeti  $Lu = g$  egyenlet folytonos megoldása helyett az  $L_h u_h = g_h$ , a végelem altérben fennálló egyenlet  $u_h$  megoldásához tart. A feladat végelem diszkretizációjánál a szokásos háromszög-felosztást vettük és Courant-elemeket használtunk, azaz  $V_h$  az elemenként lineáris függvények által kifeszített tér.

Emiatt persze vigyázni kell az  $e_k = u_k - u^*$  hibavektor számításánál, hiszen az  $u_k = u_{h,k}$  közelítés a  $V_h$  altérben van, míg  $u^* \in C(\Omega)$  folytonos függvény, azaz valamilyen vetítő eljárással le kellene vinni a pontos megoldást is  $V_h$ -ra. Azaz  $e_{k,h} = u_{k,h} - u^*$  helyett az  $u_{k,h} - P_h(u^*)$ -t számítanánk. A projekció hibája viszont megakadályozza, hogy a rácsfüggetlen szuperlineáris konvergenciát meg tudjuk mutatni.

Ezért a hibavektor számításánál nem az  $u^*$  pontos megoldást vesszük figyelembe, hanem az  $L_h u_h = g_h$   $V_h$ -beli egyenlet  $u_h$  megoldását. Az algoritmus által szolgáltatott közelítő sorozat is ehhez tart. Ezzel a számítási móddal kiderül, hogy igazából nem kell olyan feladatot választani eredményeink teszteléséhez, aminek tudnánk a pontos megoldását, hanem tetszőleges  $g$  jobb oldali függvényt beadhatunk inputnak. Ami miatt mégis jó lesz a mi szempontunkból, hogy egy rögzített  $g$  függvényt használunk, az az, hogy ki kell számolnunk az algoritmus első lépésében a  $\langle g, w_i \rangle_{L^2}$  mennyiségeket<sup>2</sup>, amiket egzaktnak számoltunk a nagyobb pontosság eléréséhez. Később megnézzük, hogy közelítő integrálások mennyire befolyásolják eredményeinket.

Az (5.16) algoritmust adaptáljuk a konkrét feladatokra. Jelölje  $c_L$  az  $L$  operátorban szereplő  $c$  konstans,  $c_S$  pedig az  $S$  operátorban szereplő nulladrendű tag együtthatóját.

<sup>2</sup> A  $V_h$  altér bázisfüggvényeit jelöljük  $w_i$ -vel.

**6.3. Megjegyzés.** Ha  $S$  az  $L$  szimmetrikus része, akkor  $c_L = c_S = c$ , de kipróbáljuk majd azt az esetet is, ha  $c_L \neq c_S$ . Ennek szuperlineáris konvergenciájáról a 5. fejezet végén volt szó, illetve bizonyítottan teljesül a rácsfüggetlenség a homogén Dirichlet peremfeltételű feladatra, bizonyos pluszfeltételek esetén, lásd a 6.2. megjegyzés előtti sorokat.

Ekkor az algoritmus a következő, már ténylegesen programozható alakot ölti:

$$\left\{ \begin{array}{l}
 (1) \quad \text{Legyen } u_0 \equiv 0 \text{ és legyen } r_0 \in V_h \text{ az} \\
 \qquad \qquad \int_{\Omega} (\nabla r_0 \cdot \nabla v + c_S r_0 v) = - \int_{\Omega} g v \quad (v \in V_h) \\
 \qquad \qquad \text{egyenlet megoldása, } d_0 := -r_0; \\
 \qquad \qquad \text{Tetszőleges } k \in \mathbb{N} \text{ esetén ha } u_k, d_k, r_k \text{ már megvan, akkor legyen} \\
 (2a) \quad p_k \in V_h \text{ a következő egyenlet megoldása} \\
 \qquad \qquad \int_{\Omega} (\nabla p_k \cdot \nabla v + c_S p_k v) = \\
 \qquad \qquad \qquad = \int_{\Omega} \left( \nabla d_k \cdot \nabla v + \frac{\partial d_k}{\partial x} v + c_L d_k v \right) \quad (v \in V_h) \quad (6.5) \\
 \qquad \qquad \gamma_k = \int_{\Omega} \left( |\nabla p_k|^2 + c_S |p_k|^2 \right); \\
 \qquad \qquad \alpha_k = -\frac{1}{\gamma_k} \int_{\Omega} (\nabla p_k \cdot \nabla r_k + c_S p_k r_k); \\
 (2b) \quad u_{k+1} = u_k + \alpha_k d_k; \\
 (2c) \quad r_{k+1} = r_k + \alpha_k p_k; \\
 (2d) \quad \beta_k = \frac{1}{\gamma_k} \int_{\Omega} \left( \nabla r_{k+1} \cdot \nabla p_k + \frac{\partial r_{k+1}}{\partial x} p_k + c_L r_{k+1} p_k \right); \\
 (2e) \quad d_{k+1} = -r_{k+1} + \beta_k d_k.
 \end{array} \right.$$

Mivel a két tesztfeladatra kapott eredmények viselkedése nem különböz, ezért együtt tárgyaljuk őket. Elsőként a szuperlineáris konvergencia megmutatása a cél.

### 6.3. A szuperlineáris konvergencia numerikus ellenőrzése

Legyen  $c = 1$  mindkét esetben, a prekondicionálás az operátor szimmetrikus részével történik. Az egységnyezetet mindkét irányban ugyanazzal a  $h$  osztásközzel felosztjuk és a kapott rácson oldjuk meg a feladatot, elsőként a korábban megadott  $g$  jobb oldali függvényekkel. A (6.5) algoritmus (1) lépésében a jobb oldali integrált egzaktul számoljuk. A táblázat első oszlopa az iterációk számát jelenti, a táblázat értékei pedig az  $\|e_k\|_S / \|e_0\|_S$  hányados értékei az adott finomságú felosztás esetén.

Látható, hogy adott pontosság eléréséhez ugyanannyit kell iterálni, kivételt képez talán a  $h = 1/4$  eset, nyilván a kevés számú ismeretlen miatt jobbak kissé az eredmények. Mindenestre az első oszlop utolsó két elemét összehasonlítva látható, hogy a használt matematikai szoftver pontossági határa korlátot szab a szuperlineáris konvergencia ellenőrzésénél is.

6.1. táblázat.  $\frac{\|e_k\|_S}{\|e_0\|_S}$  értékei az (a) feladat esetén, egzakt integrálás

Itr.	1/h			
	4	8	16	32
1	6.12650e-02	7.44796e-02	7.80156e-02	7.89163e-02
2	2.47770e-03	4.23736e-03	4.76624e-03	4.90557e-03
3	5.52553e-05	1.97167e-04	2.48989e-04	2.63357e-04
4	1.23354e-06	7.28446e-06	1.10184e-05	1.21435e-05
5	2.06502e-08	2.09954e-07	4.09078e-07	4.77815e-07
6	2.78039e-10	5.15051e-09	1.29021e-08	1.62756e-08
7	8.55155e-13	1.51943e-10	3.51329e-10	4.85233e-10
8	9.97818e-15	5.60147e-12	8.75952e-12	1.28606e-11
9	4.23692e-16	1.59044e-13	2.50982e-13	3.08620e-13
10	4.23588e-16	4.63531e-15	1.09194e-14	1.06872e-14

Tehát  $10^{-15}$  hiba alatt már a kerekítési hibák miatt nem várható további csökkenés, azaz 8-9 iterációnál többet nem érdemes figyelni. Nézzük meg a még finomabb felosztásokat is a 6.2. táblázatban.

6.2. táblázat.  $\frac{\|e_k\|_S}{\|e_0\|_S}$  értékei az (a) feladat esetén, egzakt integrálás

Itr.	1/h		
	64	128	256
1	7.91425e-02	7.91991e-02	7.92133e-02
2	4.94087e-03	4.94973e-03	4.95194e-03
3	2.67046e-04	2.67974e-04	2.68206e-04
4	1.24383e-05	1.25128e-05	1.25315e-05
5	4.96406e-07	5.01145e-07	5.02336e-07
6	1.72334e-08	1.74805e-08	1.75427e-08
7	5.26365e-10	5.37169e-10	5.39904e-10
8	1.43548e-11	1.47602e-11	1.48704e-11
9	3.53420e-13	3.84043e-13	5.90133e-13
10	3.16509e-14	1.17767e-13	4.61047e-13

Itt már nagyon sok ismeretlenről van szó, konkrétan  $h = 1/256$  esetén  $255^4 \approx 4.2$  milliárd eleme van az  $\mathbf{S}_h$  mátrixnak, ezért a 9-10.-ik iteráció eredményeibe már erősen beleszólnak a kerekítési hibák. Egyébként ha  $u_{10}$  az utolsó közelítés, akkor ebben az esetben  $L_h u_{10} - g_h = 3.09 \cdot 10^{-16}$ , azaz a 10.-ik közelítés ilyen pontosan elégíti ki az eredeti egyenletet gyenge értelemben. Mivel a hibacsökkenés gyors, ezért remélhető a rácsfüggetlen szuperlineáris konvergencia is.

Láthatóan az utolsó iteráció a fordulópont, a  $10^{-15}$  nagyságrendű hibák elérésétől



6.3. táblázat.  $\left(\frac{\|e_k\|_S}{\|e_0\|_S}\right)^{1/k}$  értékei az (a) feladat esetén, egzakt integrálás

Itr.	1/h						
	4	8	16	32	64	128	256
1	0.06127	0.07448	0.07802	0.07892	0.07914	0.07920	0.07921
2	0.04978	0.06510	0.06904	0.07004	0.07029	0.07035	0.07037
3	0.03809	0.05820	0.06291	0.06410	0.06440	0.06447	0.06449
4	0.03332	0.05195	0.05761	0.05903	0.05939	0.05948	0.05950
5	0.02904	0.04618	0.05277	0.05443	0.05485	0.05495	0.05498
6	0.02555	0.04156	0.04843	0.05034	0.05082	0.05094	0.05097
7	0.01888	0.03957	0.04461	0.04671	0.04726	0.04739	0.04743
8	0.01778	0.03922	0.04148	0.04352	0.04412	0.04427	0.04431
9	0.01958	0.03784	0.03981	0.04073	0.04135	0.04173	0.04377
10	0.02902	0.03686	0.04016	0.04008	0.04467	0.05095	0.05840

kezdve az  $\|e_k\|_S$  normák már nem tudnak csökkenni, ezért itt megfordul a tendencia és ha növelnénk az iterációs számot, akkor a hibahányadosok  $k$ -ik gyökei nem nullához, hanem 1-hez kezdenének tartani. Emiatt csak az első 9 iterációt vegyük figyelembe és ezekre valóban látható a rácsfüggetlen szuperlineáris konvergencia. A rácsfüggetlenség olyan értelemben látható, hogy egy sort tekintve ugyan nőnek a számok, de egyre kevésbé és várhatóan a rács további finomítása esetén egy ilyen sor értékei felülről korlátosak maradnak.

Az eddigiekben mindent pontosan számoltunk a programban, a gyakorlatban azonban a jobb oldali  $g$  függvény bonyolult lehet, ezért nem tudunk pontos integrált számítani, vagy pedig az is előfordulhat, hogy a függvényt esetleg analitikus formában nem is ismerjük, csak bizonyos pontokban, mérések révén. Ekkor csak valamilyen közelítő integrálás jöhet szóba. Nézzük meg, hogy mi történik, ha az  $\int_{\Omega} g w_i$  integrálokat trapézszabállyal számoljuk.

6.4. táblázat.  $\frac{\|e_k\|_S}{\|e_0\|_S}$  értékei az (a) feladat esetén, trapézszabály

Itr.	1/h			
	32	64	128	256
1	7.89170e-02	7.91418e-02	7.91989e-02	7.92132e-02
2	4.91135e-03	4.92806e-03	4.94565e-03	4.95087e-03
3	8.97559e-04	3.41631e-04	2.72926e-04	2.68444e-04
4	8.59985e-04	2.15329e-04	5.51357e-05	1.83294e-05
5	8.59962e-04	2.15034e-04	5.37638e-05	1.34504e-05
6	8.59961e-04	2.15033e-04	5.37607e-05	1.34402e-05

Ekkor nyilván nem várhatunk szuperlineáris konvergenciát, hiszen a numerikus séma rendje nem engedi a hiba rácsfüggetlen módon történő nullához tartását. Jól láthatóan

rögzített felosztás esetén a hibahányados konvergál egy  $a_h$  pozitív értékhez és ha mindkét irányban megduplázzuk a felosztások számát, akkor negyedelődik, azaz  $4a_h \approx a_{2h}$ , ami másodrendűséget mutat. Ez a numerikus séma gyakran használt, mert csak az egyes háromszögek csúcspontjaiban felvett értékekre van szükségünk, a Courant-elemek formafüggvényei pedig pontosan egy csúcspan vesznek fel az 1 értéket, a maradék kettőben 0-t és a háromszögön pedig lineáris függvény. Azaz nem kell túl sokat számolni, mégis elég jó eredményt kapunk. Hasonlítsuk össze a most használt,

$$\int_{\Delta} h \, dx = \frac{|\Delta|}{3} (h(x_1) + h(x_2) + h(x_3)) \quad (6.6)$$

képlettel számolt közelítő integrált - ahol  $x_i$  a háromszög csúcsai - egy pontosabb, de 7 pontot, nevezetesen a csúcspontokat, oldalfelezőket és a súlypontot használó

$$\int_{\Delta} h \, dx = \frac{|\Delta|}{60} \left( 3 \cdot \sum_{i=1}^3 h(x_i) + 8 \cdot \sum_{1 \leq i < j \leq 3} f(x_{ij}) + 27 \cdot h(x_s) \right) \quad (6.7)$$

képlettel. Most az egyszerű, lineáris formafüggvények miatt -  $h$  helyére  $gw_i$ -t gondolva -  $w_i$  értékei az adott pontokban előre megmondható, de a  $g$  függvényt sok helyen kell kiértékelni. Érdekes módon kevésbé lettek pontosak az eredmények és a számítási idő is jelentősen növekedett, a 256-os eset 30 percen belül sem futott le.

6.5. táblázat.  $\frac{\|e_k\|_S}{\|e_0\|_S}$  értékei az (a) feladat esetén, 7 pontos séma

Itr.	1/h		
	32	64	128
1	7.96110e-02	7.94916e-02	7.93741e-02
2	5.06820e-03	5.01084e-03	4.98160e-03
3	8.03859e-04	4.73259e-04	3.34796e-04
4	7.52652e-04	3.83880e-04	1.94098e-04
5	7.52606e-04	3.83728e-04	1.93743e-04
6	7.52598e-04	3.83720e-04	1.93735e-04

Eddig a homogén Dirichlet-peremfeltételű esettel foglalkoztunk, de látni fogjuk, hogy a vegyes peremfeltétel esete sem ront el semmit. Tekintsük tehát a (6.4) teszategyenletünket a (b) vegyes peremfeltétellel. Mivel hasonló jelenséget fogunk tapasztalni, ezért nem ismétlem mindhárom táblázatot az összes felosztásra, hanem csak egy részüket, a kellően finom felosztásokra vonatkozókat.

A 6.6. táblázatban látható viselkedése a hibahányadosnak teljesen hasonló az 6.1 és 6.2 táblázatokban szereplő értékek viselkedéséhez és a 5.3. tétel szerint itt is érvényes a szuperlineáris konvergencia, ami látható a 6.7. táblázatból is, amely a 6.3. táblázat megfelelője a (b) feladat esetén.

Összehasonlítva a 6.3. és 6.7. táblázatokat, oszloponként láthatóan nullához tartanak és ugyan a táblázatok egy sorát nézve kissé emelkednek az értékek, a rácsfüggetlenség

6.6. táblázat.  $\frac{\|e_k\|_S}{\|e_0\|_S}$  értékei a (b) feladat esetén, egzakt integrálás

Itr.	1/h			
	32	64	128	256
1	1.02890e-01	1.03065e-01	1.03109e-01	1.03120e-01
2	8.81758e-03	8.85220e-03	8.86087e-03	8.86303e-03
3	6.89621e-04	6.94412e-04	6.95613e-04	6.95913e-04
4	4.92038e-05	4.96977e-05	4.98219e-05	4.98530e-05
5	3.27167e-06	3.31630e-06	3.32754e-06	3.33036e-06
6	1.92308e-07	1.96300e-07	1.97308e-07	1.97560e-07
7	9.32505e-09	9.63548e-09	9.71431e-09	9.73410e-09
8	3.77370e-10	3.96025e-10	4.00806e-10	4.01987e-10
9	1.39880e-11	1.48677e-11	1.50963e-11	1.51669e-11
10	5.53039e-13	5.92687e-13	6.33148e-13	9.56383e-13

6.7. táblázat.  $\left(\frac{\|e_k\|_S}{\|e_0\|_S}\right)^{1/k}$  értékei a (b) feladat esetén, egzakt integrálás

Itr.	1/h						
	4	8	16	32	64	128	256
1	0.08893	0.09945	0.10219	0.10289	0.10306	0.10311	0.10312
2	0.07836	0.09024	0.09317	0.09390	0.09409	0.09413	0.09414
3	0.07105	0.08428	0.08753	0.08835	0.08855	0.08860	0.08862
4	0.06726	0.07962	0.08292	0.08375	0.08397	0.08401	0.08403
5	0.06047	0.07567	0.07911	0.07997	0.08019	0.08025	0.08026
6	0.04935	0.07062	0.07493	0.07597	0.07623	0.07630	0.07632
7	0.04367	0.06431	0.06990	0.07125	0.07159	0.07167	0.07170
8	0.03924	0.05828	0.06478	0.06639	0.06679	0.06690	0.06692
9	0.03441	0.05436	0.06057	0.06223	0.06265	0.06276	0.06279
10	0.03183	0.05228	0.05803	0.05947	0.05988	0.06028	0.06281

azonban a kerekítési hibáktól eltekintve elfogadhatónak tűnik. Természetesen az nem baj, hogy a (b) feladat táblázatában nagyobb számok vannak, hiszen a konvergencia-becslés jobb oldalán más-más  $Q_S$  operátor áll, így a sajátértékek különbözőek<sup>3</sup>. Hogy ezek a numerikus eredmények milyen viszonyban állnak az elmélet által adott 6.1. tételben szereplő (6.3) egyenlőtlenség jobb oldalán álló kifejezéssel, arra akkor tudunk választ adni, ha meghatározzuk ezeket a sajátértékeket.

Ehhez vissza kell térnünk az (5.1) probléma (5.7) szerinti  $Q_S : H_D^1(\Omega) \rightarrow H_D^1(\Omega)$

<sup>3</sup> Pontosabban ugyanazt a képletet fogjuk kapni mindkét esetben, de máshonnan indul az indexelés, így valóban azonos index esetén különbözőek lesznek a sajátértékek.

operátorához. A  $H_D^1(\Omega)$  tér jelentése szerint a sajátértékegyenlet

$$\left. \begin{array}{l} Q_S u = \lambda u \\ u|_{\Gamma_D} = 0 \end{array} \right\} \iff \left\{ \begin{array}{l} \langle Q_S u, v \rangle_{H_D^1} = \lambda \langle u, v \rangle_{H_D^1} \quad \forall v \in H_D^1(\Omega) \\ u|_{\Gamma_D} = 0 \end{array} \right. \quad (6.8)$$

alakú lesz. Figyelembe véve a skalárszorzatok jelentését (5.3) szerint, valamint bevezetve a  $\mu = 1/\lambda$  jelölést, a következő két feladatot kapjuk a (6.4) egyenlet (a) és (b) esetében:

$$\left\{ \begin{array}{l} -\Delta u - \mu \frac{\partial u}{\partial x} + cu = 0, \\ u|_{\partial\Omega} = 0; \end{array} \right. \quad (6.9)$$

illetve

$$\left\{ \begin{array}{l} -\Delta u - \mu \frac{\partial u}{\partial x} + cu = 0, \\ u|_{\Gamma_D} = 0, \\ \frac{\partial u}{\partial \nu}|_{\Gamma_N} = 0. \end{array} \right. \quad (6.10)$$

Tekintsük mondjuk a második egyenletet. Keresünk tehát olyan nemnulla  $u$  függvényt és  $\mu$  sajátértéket, amelyre teljesül a (6.10) egyenlet, a peremfeltételekkel együtt. Első lépésben ehelyett inkább oldjuk meg a

$$-\Delta v - \mu \frac{\partial v}{\partial x} + cv = \delta(\mu)v \quad (6.11)$$

egyenletet - a peremfeltételek  $v$ -re ugyanazok, mint  $u$ -ra -  $v$ -re és  $\delta$ -ra, majd keressük meg azon  $\mu$ -ket, amelyekre  $\delta(\mu) = 0$ . A

$$v_{jk}(x, y) = \exp\left(-\frac{\mu}{2}x\right) \sin(j\pi x) \cos(k\pi y) \quad (j \neq 0)$$

függvényekről könnyen ellenőrizhető, hogy teljesítik a peremfeltételeket és a

$$\delta(\mu) = (j^2 + k^2) \pi^2 + \frac{\mu^2}{4} + c$$

számokkal az (6.11) egyenlet is teljesül. A (6.9) sajátérték-egyenlet segédfeladatának megoldása

$$v_{jk}(x, y) = \exp\left(-\frac{\mu}{2}x\right) \sin(j\pi x) \sin(k\pi y) \quad (j, k \neq 0)$$

és a megfelelő  $\delta$ -k ismét ugyanazok:

$$\delta(\mu) = (j^2 + k^2) \pi^2 + \frac{\mu^2}{4} + c.$$

A (6.9) és (6.10) sajátérték-feladatot megoldását kapjuk, ha a  $\delta(\mu) = 0$  egyenletet megoldjuk  $\mu$ -re. Figyelembevéve, hogy a valójában keresett  $\lambda$  sajátértékekre  $\lambda = 1/\mu$ , ezért a tisztán képzetes

$$\lambda_{jk} = \pm \frac{i}{2\sqrt{(k^2 + j^2) \pi^2 + c}},$$

értékek adódnak, ahol az (a) esetben  $j, k \neq 0$ , a (b) esetben  $j \neq 0$ . A  $Q_S$  operátor sajátértékeinek birtokában összehasonlíthatjuk a 6.3. és 6.7. táblázatok eredményeit a (6.3)-beli

6.8. táblázat.  $\left(\frac{\|e_k\|_S}{\|e_0\|_S}\right)^{1/k}$  értékeinek és a (6.3) becslés összehasonlítása

Itr.	(a) feladat			(b) feladat			$\frac{2}{k} \sum_{i=1}^k  \lambda_i(Q_S) $	
	64	128	256	64	128	256	(a)	(b)
1	0.0791	0.0792	0.0792	0.1031	0.1031	0.1031	0.2196	0.3033
2	0.0703	0.0704	0.0704	0.0941	0.0941	0.0941	0.2196	0.3033
3	0.0644	0.0645	0.0645	0.0886	0.0886	0.0886	0.1934	0.2754
4	0.0594	0.0595	0.0595	0.0840	0.0840	0.0840	0.1803	0.2615
5	0.0549	0.0550	0.0550	0.0802	0.0803	0.0803	0.1724	0.2406
6	0.0508	0.0509	0.0510	0.0762	0.0763	0.0763	0.1671	0.2267
7	0.0473	0.0474	0.0474	0.0716	0.0717	0.0717	0.1592	0.2144
8	0.0441	0.0443	0.0443	0.0668	0.0669	0.0669	0.1533	0.2053
9	0.0414	0.0417	0.0438	0.0627	0.0628	0.0628	0.1474	0.1981

becsléssel. Ehhez készítsünk egy táblázatot, amelyben szerepelnek a hibahányados hatványainak értékei, illetve a  $k$ -ik sorban az első  $k$  sajátérték abszolútértékének számtani közepének kétszerese.

Az utolsó két oszlop tartalmazza a számtani-közepes becslést, míg a táblázat elején mindkét feladatra a sűrűbb felosztásokhoz tartozó hibahányadosok  $1/k$ -ik hatványa található. Összehasonlítva az eredményeket, láthatóan a kiszámolt értékek sokkal jobbák, mindkét esetben nagyjából a harmada az elméleti becslés által szolgáltatott értéknek. Az is észrevehető, hogy ez az arány végig megmarad az iteráció során, azaz nem kezd el a numerikusan kiszámolt érték az elméleti felső határhoz tartani.

Eddig a szokásos szimmetrikus résszel történő prekondicionálást ellenőriztük numerikusan, most térjük át arra az esetre, ha  $c_L \neq c_S$ , lásd a 6.3. megjegyzést. Ilyen prekondicionálás szempontjából az (a) feladatot vizsgálták korábban a lineáris konvergencia szempontjából [11]-ben. Rögzítsük a  $c_L = 1$  konstans és  $c_S$ -t változtassuk.

**6.4. Megjegyzés.** *A 3. fejezet egésze arra kellett, hogy a GCG-LS(0) algoritmus használatának jogosságát ellenőrizzük. Ha nem a szimmetrikus résszel prekondicionálunk, akkor ebben az esetben elvileg nem használható ez az algoritmus.*

Az elvégzett számítások azt mutatják, hogy ekkor is igaz a szuperlineáris konvergencia, de nem tetszőleges  $c_S$  konstansra. Van a  $c_L$ -nek egy nem túl nagy környezete, amelyből  $c_S$ -t választva igaz marad a szuperlineáris konvergencia, de nem annyi lépésen keresztül, mintha  $c_L = c_S$  lenne, nem innen választva viszont csak néhány iteráción keresztül fog csökkenni az  $(\|e_k\|_S / \|e_0\|_S)^{1/k}$  értéke, utána viszont előfordulhat, hogy néhány nagyobb érték után újra csökkenni kezd. Ez azt is jelentheti, hogy szuperlineáris konvergenciát lehet, hogy az elmélet sem tudna biztosítani a különböző konstansok esetére, de mindenesetre „majdnem” igaz ez is. A konvergencia viszont igaz marad tetszőleges  $c_S$  konstansra. Nézzük meg ezt

számokban is a (b) feladatunkra; ezt az esetet jelenleg az elmélet nem tudja kezelni, bár azt várjuk, hogy igaz marad ebben az esetben is a tétel, így ezek a numerikus kísérletek alátámaszthatják a sejtés igazságát. Nézzük meg 2 felosztás esetén az eredményeket, ha  $c_S$ -t  $c_L = 1$  körül mozgatjuk.

6.9. táblázat.  $\left(\frac{\|e_k\|_S}{\|e_0\|_S}\right)^{1/k}$  értékei a (b) feladat esetén,  $c_S \neq 1 = c_L$ .

Itr.	1/h=32				1/h=128			
	$c_S = 0$	$c_S = 0.5$	$c_S = 1.5$	$c_S = 5$	$c_S = 0$	$c_S = 0.5$	$c_S = 1.5$	$c_S = 5$
1	0.1032	0.1023	0.1047	0.1331	0.1034	0.1025	0.1049	0.1332
2	0.0940	0.0939	0.0939	0.1160	0.0943	0.0942	0.0941	0.1162
3	0.0924	0.0897	0.0888	0.1005	0.0926	0.0900	0.0890	0.1008
4	0.0911	0.0856	0.0847	0.0923	0.0914	0.0858	0.0849	0.0923
5	0.0897	0.0845	0.0833	0.0988	0.0899	0.0847	0.0835	0.0985
6	0.0937	0.0857	0.0846	0.0989	0.0939	0.0860	0.0848	0.0990
7	0.0945	0.0869	0.0860	0.0967	0.0947	0.0871	0.0862	0.0968
8	0.0926	0.0879	0.0866	0.0902	0.0929	0.0881	0.0868	0.0906
9	0.0945	0.0872	0.0865	0.0896	0.0948	0.0875	0.0868	0.0896
10	0.0951	0.0870	0.0869	0.0886	0.0954	0.0872	0.0871	0.0890

Ha a 6.9. táblázat eredményeit összehasonlítjuk a 6.7. táblázat megfelelő oszlopaival, akkor látható, hogy ugyanazt a monoton csökkenő típusú eredményt csak  $c_L$ -hez közeli  $c_S$  választás esetén tudunk mutatni. Más konstansok esetén a szuperlineáris konvergencia követelménye túl erősnek tűnik, nem tudni, hogy a kerekítési hibák játszanak-e döntő szerepet, vagy pedig a megfigyeléssel összhangban csak a  $c_L$  egy környezetében teljesül ez a konvergenciasebesség. Azt is észre lehet venni, hogy rögzített  $c_S$  érték esetén a felosztás finomságának változtatása alig változtat a számokon (ez nem csak ezen két választott felosztás esetén teljesül, hanem az összesre, durvább rácson kisebbek az értékek, de a finomítás során rögzített  $c_S$ -re a számok növekedése lassul és várhatóan korlátos marad), ami ismét rácsfüggetlenséget mutat, bár ez sem bizonyított.

**6.5. Megjegyzés.** *Ennek szerepe lehet a gyakorlati számítások során, ha mondjuk  $c_S = 0$ -t választunk. Ekkor ugyanis az  $S_h$  mátrix egyszerűbben számolható, a számítási idő csökkenthető, főleg ha nem Courant-elemekkel dolgozunk, hanem magasabbfokú bázisfüggvényekkel. A merevségi mátrix összeállítása során ebben az esetben sok integrál kiszámításától megszabadulhatunk és a fentiek szerint a konvergenciasebességre ez csekély hatást gyakorol.*<sup>4</sup>

Ennek fordítottja, ha az  $L$  operátorban nem szerepel nulladrendű tag, de a prekondi-

<sup>4</sup> Ne felejtjük el, hogy már 4-5 iteráció után elérjük a  $10^{-7}$  nagyságú hibát, ami a gyakorlatban már sokszor elég lehet. Emiatt lényegtelen, hogy esetleg csak az első néhány lépésben sikerült szuperlineáris konvergenciát mutatni, mert ez éppen elég nekünk, a konvergencia amúgy sem áll le, hanem csak kissé lassul.

cionerbe beleteszünk egyet. Most is csak a (b) feladatot nézzük meg.

6.10. táblázat.  $\left(\frac{\|e_k\|_S}{\|e_0\|_S}\right)^{1/k}$  értékei a (b) feladat esetén,  $c_S \neq 0 = c_L$ .

Itr.	1/h=32				1/h=128			
	$c_S = 0.2$	$c_S = 0.5$	$c_S = 1.5$	$c_S = 5$	$c_S = 0.2$	$c_S = 0.5$	$c_S = 1.5$	$c_S = 5$
1	0.1085	0.1100	0.1174	0.1551	0.1088	0.1102	0.1175	0.1552
2	0.0985	0.0985	0.1000	0.1383	0.0988	0.0987	0.1002	0.1384
3	0.0928	0.0933	0.0946	0.1224	0.0931	0.0936	0.0948	0.1226
4	0.0879	0.0891	0.0988	0.1286	0.0881	0.0893	0.0990	0.1286
5	0.0843	0.0876	0.0975	0.1393	0.0846	0.0878	0.0978	0.1396
6	0.0820	0.0894	0.0974	0.1302	0.0823	0.0897	0.0977	0.1304
7	0.0830	0.0910	0.1019	0.1226	0.0832	0.0912	0.1022	0.1229
8	0.0860	0.0910	0.1021	0.1256	0.0862	0.0912	0.1024	0.1258
9	0.0849	0.0909	0.0992	0.1275	0.0853	0.0912	0.0995	0.1280
10	0.0817	0.0920	0.1020	0.1202	0.0821	0.0923	0.1023	0.1206

A 6.10. táblázatban sem monoton csökkenő sorozatokat látunk. Mivel  $c_S \neq c_L$  esetén kissé lassabb a konvergencia, ezért ezekben az esetekben 10 lépés alatt a  $\|e_k\|_{S_h}$  hiba nem feltétlenül lesz olyan kicsi, hogy tovább már nem tud javulni, ezért ilyen esetekben érdemesebb akár 20-at is iterálni, hogy jobban lássuk a hibahányados hatványainak - immár nem monoton csökkenő - viselkedést. A fentiekből adódik, hogy elég közel maradva az  $L$  operátor szimmetrikus részéhez, továbbra is érvényes a szuperlineáris konvergencia, legalábbis az elején. Átmeneti növekedés után pedig újra csökkenhet az érték. Az, hogy ez is működik, némileg meglepő, hiszen eddig mindent a szimmetrikus résszel történő prekondicionálásra láttunk. Most nem a szimmetrikus részt használtuk, így az is meglepő, hogy az ebben az esetben elvileg nem is használható algoritmus is konvergál, sőt ha közel vagyunk a jó prekondicionerhez, akkor szuperlineárisan is. Érdekes módon még azt is meg lehet engedni a program szempontjából, hogy negatív  $c_S$ -t vegyünk! Ráadásul az eredmények közel azonosak, ha  $c_L = 0$  és  $c_S = \pm 0.1$ . Mindkét esetben a közelség miatt érvényes a szuperlineáris konvergencia 7 lépésen keresztül, majd átmeneti növekedés után a minimumát 0.0765, illetve 0.0767-ben veszik fel ( $N=64$ ), ekkor  $\|e_{11}\|_S$  értéke  $10^{-13}$  nagyságrendű.

Végezetül azt az esetet nézzük meg, amikor olyan prekondicionert választunk, amelynek a 5.9. tétel értelmében rosszak a peremfeltételei, azaz a (b) feladatot prekondicionáljuk az (a) feladathoz tartozó  $S$  operátorral. Ezek azonban nem teljesítik a peremfeltételkövetelményeket, azaz a peremnek nem ugyanazon a részén vannak a Dirichlet-peremfeltételek, hiszen az (a)-ban csupa ilyen van, (b)-ben pedig az egységnyezet két oldalán is Neumann-peremfeltétel adott.

Az eredmények szerint ez valóban hibás eljárást ad, mert még konvergencia sincsen, nemhogy szuperlineáris, továbbá a rács finomítása esetén sem jutunk közelebb a megoldás-

6.11. táblázat.  $\frac{\|e_k\|_S}{\|e_0\|_S}$  értékei az (b) feladatra, az (a) feladat  $\mathbf{S}_h$  mátrixát használva.

Itr.	1/h					
	4	8	16	32	64	128
1	0.8338	0.8064	0.7989	0.7970	0.7965	0.7964
2	0.8321	0.8038	0.7961	0.7941	0.7936	0.7934
3	0.8321	0.8038	0.7960	0.7940	0.7935	0.7934
4	0.8321	0.8038	0.7960	0.7940	0.7935	0.7934
5	0.8321	0.8038	0.7960	0.7940	0.7935	0.7934

hoz, mint ez kiolvasható a 6.11. táblázatból.

Összefoglalva elmondható, hogy a szimmetrikus résszel prekondicionálva teljesül a szuperlineáris konvergencia és a numerikus eredmények lényegesen jobbák - nagyjából a harmada - az elméleti felső korlátnál. Ha  $c_S \neq c_L$ , akkor valószínűleg nem teljesül végig a szuperlineáris konvergencia, esetleg egy nagyon kis környezetében  $c_L$ -nek, hanem csak néhány lépésen keresztül, esetleg hol lelassulva, hol ismét felgyorsulva. Ez persze azt jelenti, hogy a konvergencia minden esetben teljesül, méghozzá elég gyors, csak kevés hiányzik a szuperlineáris gyorsasághoz. Bár az erősebb tulajdonság nem teljesül, ez még így is meglepő, hiszen a „nem megfelelő” prekondicionerrel és a valóban nem megfelelő algoritmussal is igaz egy gyors konvergencia.

#### 6.4. Konjugált gradiens-módszer vs. egy gradiens-típusú iteráció

Bár a 6.1. táblázat értékei mutatják, hogy a konvergencia gyors, felmerülhet a kérdés, hogy vajon a számítási idő megfelelő, azaz az iterációs lépésben egzaktul kiszámított - szimmetrikus mátrixú - egyenlet megoldásának és a segédkonstansok kiszámításának ideje talán túlságosan nagy. Esetleg egy lassabban konvergáló, de kevesebb műveletet végrehajtó módszer adott pontosság eléréséhez hamarabb jut el, mint a mi gyors, de nagyobb számítási igényű algoritmusunk. Ezért most egy sokkal egyszerűbb, gradiens-típusú iterációt alkalmazunk tesztfeladatainkra és megvizsgáljuk, hogy adott pontosság eléréséhez mekkora a futási idő mindkét algoritmusra, illetve mennyit iterál az egyik és mennyit a másik.

$$\left\{ \begin{array}{l}
 (1) \quad \text{Legyen } u_0 \in H_D^1(\Omega) \text{ tetszőleges.} \\
 \text{Tetszőleges } k \in \mathbb{N} \text{ esetén ha } u_k, \text{ már megvan, akkor legyen} \\
 (2a) \quad r_k \in H_D^1(\Omega) \text{ a következő egyenlet gyenge megoldása} \\
 \quad \quad \quad \left\{ \begin{array}{l}
 -\Delta r_k + cr_k = Lu_k - g \\
 r_k|_{\Gamma_D} = 0, \quad \frac{\partial r_k}{\partial \nu} = 0
 \end{array} \right. \\
 (2b) \quad u_{k+1} = u_k - \frac{m}{M^2} r_k, \text{ ahol } m = 1, M = \sqrt{1 + \frac{1}{\pi^2}};
 \end{array} \right. \quad (6.12)$$



Természetesen most a szuperlineáris konvergencia helyett az  $\|e_k\|_S$  normát figyeljük és hasonlítjuk össze a konjugált gradiens-módszerrel. Ennek az algoritmusnak a konkrét alakja most a következő:

$$\left\{ \begin{array}{l} (1) \quad \text{Legyen } u_0 \equiv 0 \\ \text{Tetszőleges } k \in \mathbb{N} \text{ esetén ha } u_k, \text{ már megvan, akkor legyen} \\ (2a) \quad r_k \in V_h \text{ a következő egyenlet megoldása} \\ \qquad \int_{\Omega} (\nabla r_k \cdot \nabla v + c_S r_k v) = \\ \qquad \qquad \qquad = \int_{\Omega} \left( \nabla u_k \cdot \nabla v + \frac{\partial u_k}{\partial x} v + c_L u_k v - g v \right) \quad (v \in V_h) \\ (2b) \quad u_{k+1} = u_k - \frac{m}{M^2} r_k, \text{ ahol } m = 1, M = \sqrt{1 + \frac{1}{\pi^2}}; \end{array} \right. \quad (6.13)$$

Itt minden lépésben meg kell oldani egy szimmetrikus mátrixú - amely ugyanaz, mint a konjugált gradiens-módszer mátrixa - lineáris egyenletrendszert, de nincs szükségünk a segédvektorokra és konstansokra, azaz nem kell normákat és skalárszorokat számolnunk. A következő táblázatok szerint ezek műveletigénye csekély, azaz ez a nyereség nem kompenzálja a konvergencia lassúságát.

6.12. táblázat. Módszerek összehasonlítása,  $n = 16$  esetén a (b) feladatra

	Gradiens-módszer		KGM	
pontosság	iteráció	futásidő	iteráció	futásidő
$10^{-2}$	2	0.25	2	0.06
$10^{-4}$	5	0.32	4	0.14
$10^{-6}$	7	0.34	6	0.19
$10^{-8}$	10	0.41	7	0.21
$10^{-10}$	12	0.42	9	0.23
$10^{-12}$	15	0.47	10	0.24

Érdekes módon itt még az elején is jobb a konjugált gradiens-módszer, ennek oka az lehet, hogy kevés ismeretlen esetén a végeelem program jobb oldalán álló kifejezés utolsó tagjának számítása elviszi azt az előnyt, hogy segédvektorokat nem kell számolni, de lehet pusztán véletlen is. A táblázatok első oszlopában rögzítünk egy  $\varepsilon$  küszöbszámot és az figyeljük, hogy  $\|e_k\|_S < \varepsilon$  hányadik iteráció után következik be és ez mennyi ideig tart másodpercben mérve.

Ebből a szempontból a 6.13. és a 6.14. táblázatokon jobban látszik, hogy a gradiens-típusú algoritmus azonos iterációs szám esetén gyorsabb, de adott pontossághoz egyre többet kell iterálnia és így jelentősen nő a számítás ideje.

Végezetül elmondható, hogy egy rendkívül hatékony módszerről van szó, amelynek gyors konvergenciája kompenzálja azt, hogy az algoritmusa kissé bonyolultabb. Az általunk

6.13. táblázat. Módszerek összehasonlítása,  $n = 64$  esetén a (b) feladatra

pontosság	Gradiens-módszer		KGM	
	iteráció	futásidő	iteráció	futásidő
$10^{-2}$	2	1.26	2	1.66
$10^{-4}$	5	2.78	4	2.76
$10^{-6}$	7	3.77	6	3.89
$10^{-8}$	10	5.35	7	4.44
$10^{-10}$	12	6.34	9	5.58
$10^{-12}$	15	7.49	10	6.19

6.14. táblázat. Módszerek összehasonlítása,  $n = 256$  esetén a (b) feladatra

pontosság	Gradiens-módszer		KGM	
	iteráció	futásidő	iteráció	futásidő
$10^{-2}$	2	69.17	2	100.36
$10^{-4}$	5	162.81	4	164.86
$10^{-6}$	7	222.89	6	228.54
$10^{-8}$	10	313.45	7	262.39
$10^{-10}$	12	382.35	9	325.22
$10^{-12}$	15	470.54	10	358.41

vizsgált konstansegyütthatós esetben mindkét peremfeltétel esetén jó működött. Sikerült ellenőrizni az elméleti eredmények teljesülését, megmutattuk, hogy még jobb is a helyzet, mint amit a tételek állítanak és olyan eseteket is vizsgáltunk, amelyekre az elmélet nem terjed ki.

## IRODALOMJEGYZÉK

- [1] O. Axelsson, *A generalized conjugate gradient least square method*, Numer. Math. 51 (1987), 209-227.
- [2] O. Axelsson, *Iterative Solution Methods*, Cambridge University Press, 1994.
- [3] O. Axelsson, Karátson J., *Symmetric part preconditioning for the conjugate gradient method in Hilbert space*, Numer. Funct. Anal. 24 (2003), No. 5-6, 455-474.
- [4] O. Axelsson, Karátson J., *Superlinearly convergent CG methods via equivalent preconditioning for nonsymmetric elliptic operators*, Numer. Math. 99 (2004), 197-223.
- [5] Ph. Ciarlet, *The Finite Element Method For Elliptic Problems*, North-Holland, 1978.
- [6] V. Faber, Thomas A. Manteuffel and Seymour V. Parter, *On the Theory of Equivalent Operators and Application to the Numerical Solution of Uniformly Elliptic Partial Differential Equations*, Advances In Applied Mathematics 11 (1990), 109-163.
- [7] Fried Ervin, *Algebra I.*, Nemzeti Tankönyvkiadó, 2000.
- [8] C. I. Goldstein, Thomas A. Manteuffel and Seymour V. Parter, *Preconditioning and Boundary Conditions Without  $H^2$  Estimates:  $L^2$  Condition Numbers and the Distribution of the Singular Values*, SIAM J. Numer. Anal. 30 (1993), No. 2, 343-376.
- [9] Wayne Joubert, Thomas A. Manteuffel, Seymour V. Parter and Sze-Ping Wong, *Preconditioning Second-Order Elliptic Operators: Experiment and Theory*, SIAM J. Sci. Stat. Comput. 13 (1992), No. 1, 259-288.
- [10] Karátson J., *Superlinear PCG algorithms: symmetric part preconditioning and boundary conditions*, submitted
- [11] Thomas A. Manteuffel and James Otto, *Optimal Equivalent Preconditioners*, SIAM J. Numer. Anal. 30 (1993), No. 3, 790-812.
- [12] Thomas A. Manteuffel and Seymour V. Parter, *Preconditioning and boundary conditions*, SIAM J. Numer. Anal. 27 (1989), 656-694.
- [13] Stoyan Gisbert, Takó Galina, *Numerikus módszerek III*, ELTE - TypoT<sub>E</sub>X, 1997.