

Iteratív módszerek összehasonlítása diszkretizált nem szimmetrikus elliptikus feladatok esetében

A stacionárius konvekció-diffúziós egyenletek olyan elliptikus parciális differenciálegyenletek, melyek áramlási folyamatokat modelleznek. Amennyiben a diffúziós hatás elhanyagolható, a folyamat konvekció-dominálttá válik. Tekintsünk ilyen egyenleteket az (1)-ben felírt alakban, és a megoldást keressük az $\Omega = (0, 1)^2$ egységnyezeten homogén Dirichlet peremfeltétel mellett.

$$\begin{cases} -\alpha \Delta u + \mathbf{w} \cdot \nabla u = f \\ u|_{\partial\Omega} = 0 \end{cases}$$

Ha α egy pozitív konstans, $\mathbf{w}$ egy folytonosan differenciálható forrásmentes vektormező és $f \in L^2(\Omega)$, akkor a feladatnak egyértelműen létezik $u \in H_0^1(\Omega)$ gyenge megoldása.

Az analitikus megoldást általában nem lehet kiszámolni, ezért numerikus módszerekkel véges dimenziós térben közelítő megoldást keresünk. A dolgozatban háromféle diszkretizációs módszert használok: véges differencia-módszert (FDM), standard végeselem-módszert (FEM) és áramvonal-menti végeselem-módszert (SDFEM), mely utóbbi kimondottan konvekció-dominált feladatokra lett kifejlesztve, és stabilizálja egy $\delta > 0$ paramétertől függő taggal a FEM konvergenciáját.

A diszkretizáció egy $Au = b$ lineáris egyenletrendszerhez vezet, ahol az A mátrix nem szimmetrikus. Az egyenletrendszer megoldásához iteratív módszereket használunk, ahol több lehetőség is felmerül. Két gyakran használt módszer a CGN-módszer, mely a konjugált gradiens-módszert alkalmazza a szimmetrizált normálegyenletre, és a GCR-módszer, amely a reziduális hibát minimalizálja a keresési altéren. A konvergencia gyorsítása érdekében az A mátrix szimmetrikus részével prekondicionálunk. A két algoritmus megtalálható a 2.5 fejezet végén. Ezek az algoritmusok és a három diszkretizációs módszer MATLAB-ban implementálva lett.

A dolgozatban azt a kérdést vizsgálom, hogy a differenciálegyenlet különböző paraméterei esetén melyik iteratív módszer konvergál kevesebb lépésben, azaz melyiknél éri el hamarabb a reziduális hibavektor normája a megadott toleranciaszintet. Ehhez bevezetjük a diffúziós tag nagyságát meghatározó ε , és a konvekciós tag nagyságát befolyásoló ρ pozitív konstans változót, és a peremérték-feladatot a (11)-ben megadott alakban vizsgáljuk. A mérések során ε értékét rögzítjük, és ρ -t elkezdjük növelni 0-ról, majd a kapott lépésszámot ábrázoljuk mindkét iteratív módszer esetén. A méréseket elvégezzük egy konstans és egy forgó vektormezővel is. Ezek a grafikonok láthatóak FDM esetén a 2. ábrán, FEM esetén a 3. ábrán, és SDFEM esetén a 4. és 5. ábrán.

Az FDM és FEM esetében nagyon hasonló grafikonokat kapunk. A ρ növelésével a lépésszám is növekszik, mely a CGN-módszernél közel lineáris, de a GCR-módszer görbéje konkáv alakú, ezért a két görbe metszi egymást egy pontban. A metszéspont előtt a CGN-módszer konvergál kevesebb lépésben, míg utána ez megfordul, és a GCR-módszer konvergál gyorsabban.

Az SDFEM esetében ρ növelésével a lépésszám hirtelen megnövekszik, majd eléri a csúcspontját, és utána gyorsan lecsökken. A 4. ábrán a δ stabilizáló paraméter kb. 0.02, és ρ értékétől függetlenül mindig a CGN-módszer konvergál gyorsabban. Ehhez képest amikor az 5. ábrán $\delta = 5 \cdot 10^{-5}$, a görbék kétszer metszik egymást. A metszéspontok között a GCR-módszer a gyorsabb, míg bárhol máshol a CGN-módszer a jobb.

A dolgozat további részében elméleti számítások segítségével bizonyítottam, hogy a grafikonok alakja mindig ilyen lesz. Ehhez először az iteratív módszerek (5) és (6)-nál ismertetett nevezetes lineáris konvergenciabecslését hasonlítottam össze, melyet a 3.1. tétel tartalmaz: a GCR-módszer lineáris becslése a k . iteratív lépésben pontosan akkor jobb a CGN-módszer becslésénél, ha $\frac{M}{m} > L_k$, ahol m és M a (4)-ben definiált alsó- és felső határ, L_k pedig a tételben definiált f_k harmadfokú polinom egyértelműen létező valós gyöke. A bizonyításban a harmadfokú polinom levezetése után meghatároztam a gyököt a Cardano-képlet segítségével, az egyértelműséghez pedig a diszkrimináns előjelét vizsgáltam.

A 3.2. állításban numerikusan ellenőriztem, hogy a gyökök sorozatának első 10^4 tagja szigorúan monoton növekszik (ez elég a dolgozat méréseihez), és a 3.3. megjegyzésben sejtést fogalmaztam meg arról, hogy a sorozat végig monoton növekszik, de ezt nem bizonyítottam. A sorozat lineárisan jól közelíthető, erre lineáris regresszióval képletet adtam. Az első tíz tagot az 1. táblázat tartalmazza. A 3.4. állításban bizonyítottam a monotonitás egy hasznos következményét: ha a CGN-módszer lineáris becslése jobb a GCR-módszer becslésénél egy adott k' indexre, akkor bármely $k > k'$ indexre is jobb lesz.

Ezután az általános eredményeket alkalmaztam a diszkretizációból kapott egyenletrendszerre. Ha a prekondicionáló S mátrix az A mátrix szimmetrikus része, akkor a 3.6. állítás szerint az $S^{-1}A$ mátrix m koercivitási együtthatója 1, ezért a 3.7. következmény szerint az $\|S^{-1}A\|_S < L_k$ feltétel teljesülése esetén a CGN-módszer lineáris becslése jobb lesz a GCR-módszer becslésénél a k . iterációs lépésben. A korábbi 3.4. állítás pedig garantálja, hogy ha már $k = 1$ esetén jobb, akkor bármely későbbi indexre is jobb lesz.

A peremérték-feladathoz tartozó bilineáris forma felső határának vizsgálatával és az előbb meghatározott feltétel előírásával sikerült a konvekciós tag ρ paraméterére olyan korlátot adni, melynek teljesülése esetén garantált, hogy a CGN-módszer konvergál gyorsabban. Standard FEM esetében a 3.8. állításban, SDFEM esetében pedig a 3.9. állításban a megfelelő bilineáris formák felső határát a Cauchy–Schwarz- és a Poincaré–Friedrichs-egyenlőtlenséggel becsültem, ami az ekvivalens prekondicionálás miatt átvihető az $\|S^{-1}A\|_S$ normára is, és erre már alkalmazható a korábbi feltétel, amely ρ -ra felső korlátot ad. Alsó korlát csak az SDFEM esetén létezik. Ezt a 3.11. állításban mondtam ki, ahol a bizonyításban egy módosított bilineáris formára alkalmaztam az úgynevezett áramvonal-menti Poincaré–Friedrichs-egyenlőtlenséget. Az SDFEM esetében a kétoldali korlát létezése miatt a 3.14. állításban levezettem egy alsó becslést a δ stabilizáló paraméterre, melynek teljesülése esetén ρ értékétől függetlenül valamelyik korlát biztosan teljesül, ezért mindig a CGN-módszer konvergál gyorsabban ilyenkor.

A 3.3. fejezetben a lineáris becsléseknél elért eredményekkel megmagyaráztam a 3.1. fejezet mérési grafikonjainak több részét is. Ennek alapját az képezi, hogy a 3.8. és 3.13. állításban a ρ -ra adott feltétel teljesülésekor a reziduális hibavektor normáját nagyon jól közelítik a lineáris becslések, így a becslések közti reláció eldönti, hogy melyik módszer konvergál gyorsabban. A tapasztalt reziduális normák és a lineáris becslésük a 6. és 7. ábrán látható. Ez tehát megmagyarázza, hogy a CGN-módszernek miért kell gyorsabban konvergálnia a grafikonok elején, amikor ρ elegendően kicsi (azaz teljesül a (12) feltétel), és SDFEM esetében a grafikonok végén is, amikor ρ elegendően nagy, és teljesül a (14) feltétel. Emellett arra is magyarázatot kapunk, hogy ha a δ stabilizáló paraméter elegendően nagy az SDFEM módszernél (azaz teljesül a (17) feltétel), akkor ρ -tól függetlenül mindig a CGN-módszer konvergál gyorsabban, mint ahogy azt a 4. ábra grafikonján is láttuk. A ρ -ra adott alsó- és felső korlátot kiszámoltam a $\mathbf{w}_0 = (1, 0)$ konstans vektormező esetében.

A lineáris becslések nem közelítenek jól a grafikonok azon részein, ahol a GCR-módszer konvergál gyorsabban, így ezeknél más becsléseket kell használnunk. A 3.4. fejezetben az iteratív módszerek nevezetes szuperlineáris konvergenciabecslését hasonlítottam össze, mely a (8) és (9)-nél található becsléseket jelenti, amennyiben a bennük szereplő E mátrix antiszimmetrikus. A 3.19. állításban levezettem egy feltételt az E mátrix szinguláris értékeire, melynek teljesülése esetén a GCR-módszer szuperlineáris becslése a k . iterációs lépésben jobb a CGN-módszer becslésénél. A lineáris becslésekhez képest itt nehézséget okoz, hogy ismerni kell az E szinguláris értékeit, amely függ az egyenlet paramétereitől is. Ezért itt azt az egyszerűbb esetet vizsgáltam részletesebben, amikor a $\mathbf{w}_0 = (1, 0)$ konstans vektormező szerepel a differenciálegyenletben, és a diffúziós tag ε paramétere 0.

A 3.21. állításban operátorszinten megoldottam a kapcsolódó sajátérték-feladatot SDFEM diszkretizáció esetén, és képletet adtam az E operátor sajátértékeire. A 3.22. megjegyzésben felírtam a sajátérték-feladatot az $\varepsilon > 0$ esetben is. Az E operátor antiszimmetrikus tulajdonságát felhasználva a 3.23. állításban képletet adtam az E operátor szinguláris értékeire. Ezt felhasználva a 3.24. állításban behelyettesítettem a szinguláris értékeket a 3.19. állításban levezetett feltételbe, mellyel ebben a speciális esetben ρ -ra felső korlátot adtam, melynek teljesülése esetén az adott lépésben a GCR-módszer szuperlineáris becslése a jobb. A felső korlátban szereplő összegekre az Euler–Maclaurin-képlet segítségével másodrendű közelítést adtam, melynek hibáját a 8. ábra szerint vizsgáltam.

Standard FEM diszkretizáció esetén a 3.27. állításban egy egyszerű számolással alsó korlátot vezettem le a ρ paraméterre, melynek teljesülése esetén a GCR-módszer szuperlineáris becslése jobb a k . iterációs lépésben. A 3.25. és 3.28. megjegyzésben megállapítottam, hogy a szuperlineáris becslésekkel kapott eredmények megerősítik azokat a mérési eredményeket, ahol a GCR-módszer konvergált gyorsabban.

A 3.5. fejezetben megvizsgáltam, hogy a szuperlineáris becslések mennyire közelítik meg az iteratív módszerek futása során tapasztalt reziduális normákat. A numerikus eredmények a 9. és 10. ábra grafikonjain látható. Ezek alapján megállapítható, hogy ρ azon értékeire, ahol a GCR-módszer konvergál gyorsabban, a szuperlineáris becslések (és a lineáris becslések is) nagyon rosszul közelítenek, és nem követik szorosan a reziduális normákhoz tartozó görbét. Ezért a szuperlineáris becsléseknél levezetett összefüggések önmagukban nem bizonyítják, hogy ezeken a részeken a GCR-módszernek mindig gyorsabban kell konvergálnia.

Végezetül kísérletet tettem a szuperlineáris becslések kapcsán felmerült probléma javítására. A GCR-módszernek van egy másik nevezetes szuperlineáris becslése, melyet a (10)-ben ismerttettem. Ez a szinguláris értékek összege helyett azok szorzatát használja, amely egy élesebb becslést eredményez a reziduális normára. Az egységnégyzet egy ritkább rácsfelosztását véve $\text{TOL} = 10^{-100}$ toleranciaszint mellett látjuk a konvergencia teljes folyamatát a 11. és 12. ábrán, ahol láthatóan ez a szuperlineáris becslés jól közelít. Mivel a szakirodalomban nem találtam a CGN-módszerhez olyan szuperlineáris becslést, mely hasonlóan a szinguláris értékek szorzatát használná, ezért nem volt azt mivel összehasonlítani, így ennél tovább nem jutottam, de bízunk benne, hogy ez az eredmény javítható, és meg tudja magyarázni a hiányzó részeket.