Numeričko rješavanje problema sedlaste točke

Telišman, Mihovil

SVEUČILIŠTE U ZAGREBU

PRIRODOSLOVNO–MATEMATIČKI FAKULTET

MATEMATIČKI ODSJEK

Mihovil Telišman

NUMERIčKO RJEŠAVANJE PROBLEMA SEDLASTE TOčKE

Diplomski rad

Voditelj rada:
doc. dr. sc. Nela Bosner

Zagreb, Studeni, 2016.

Ovaj diplomski rad obranjen je dana _____ pred ispitnim povjerenstvom u sastavu:

1. _____, predsjednik
2. _____, član
3. _____, član

Povjerenstvo je rad ocijenilo ocjenom _____.

Potpisi članova povjerenstva:

1. _____
2. _____
3. _____

Sadržaj

Poglavlje 1

Uvod

Problemi koji se svode na sustave sa sedlastom točkom dolaze iz širokog spektra primjena te su iz tog razloga predmet mnogih akademskih radova. U ovom radu iznjet će se kratak pregled najvažnijih primjena u kojima se sustavi takvog oblika prirodno javljaju s posebnim naglaskom na probleme inkompresibilnog toka i probleme optimizacije te njihovo rješavanje. Sustavi ovog oblika su iznimno numerčki zahtjevni zbog loših svojstava, prije svega loše uvjetovanosti. Najvažniji alat prilikom konstruiranja rješenja takvih sustava je Schurov komplement. Koristeći Schurov komplement u mnogim slučajevima je moguće transformirati početni sustav u ekvivalentan sa boljim teorijskim i numeričkim svojstvima. Schurov komplement je ujedno i krucijalan element prve numeričke metode koja je efikasno rješavala određenu klasu sustava sa sedlastom točkom - Uzawa metode. Osnovna ideja Uzawa metode je nepromijenjena do današnjih dana pa se shodno tome ona i dalje koristi zajedno sa mnogim varijacijama. Jedina razlika su moderniji rješavači za pomoćne sustave koje metoda generira u svakom koraku. Jedna od najvažnijih metoda za nesimetrične sustave je GMRES metoda koja je primjenjiva na široku klasu problema od kojih se mnogi svode na sustave sa sedlastom točkom. Nažalost nijedna od trenutnih metoda nije bez mane i često su u praksi potpuno neupotrebljive bez prekondicioniranja. Prekondicioniranje za sustave sa sedlastom točkom je iznimno složen postupak i u ovom radu će biti tek površno spomenut.

Napomenimo da su sustavi sa sedlastom točkom u praksi uglavnom realni te su zbog toga sva razmatranja iznesena u ovom radu valjana samo u realnom slučaju osim ako nije posebno drugačije navedeno.

Poglavlje 2

Problemi koji se svode na sustave sa sedlastom točkom

2.1 Problem sedlaste točke i klasifikacija

Neka je dan 2x2 blok sustav linearnih jednadžbi:

$$\begin{bmatrix} A & B_1^T \\ B_2 & -C \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} f \\ g \end{bmatrix} \tag{2.1}$$

pri čemu je $A \in R^{n\times n}, B_1, B_2 \in R^{m\times n}, C \in R^{m\times m}, n \geq m$.

Osnovni problem s kojim se susrećemo je rješenje danog sustava. Očito da primjenom adekvatnih transformacija, svaki linearni sustav je moguće napisati u obliku (2.1) te zbog toga uvodimo dodatne pretpostavke. Iz daljnih razmatranja izbacujemo slučajeve u kojima je A nul-matrica ili jedna od B_1 ili B_2 nul-matrica. Da bi opisani sustav opisivao generalizirani problem sedlaste točke potrebno je da zadavoljava jedan ili više od sljedećih uvjeta:

1. A je simetrična
2. simetrični dio od A, $H \equiv \frac{1}{2}(A + A^T)$ je pozitivna semidefinitna
3. $B_1 = B_2 = B$
4. C je simetrična i pozitivna semidefinitna
5. C je nul-matrica

Primjetimo da uvjet 5) implicira uvjet 4). Također, u literaturi postoji određeno odstupanje od navedene definicije, ali konstrukcija problema u ovoj formi je najopćenitija moguća i sadrži druge alternativne definicije kao specijalne slučajeve.

Ako pretpostavimo da su zadovoljeni svi uvjeti iz definicije tada dobivamo sustav sljedećeg oblika:

$$\begin{bmatrix} A & B^T \\ B & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} f \\ g \end{bmatrix} \tag{2.2}$$

pri čemu je A simetrična pozitivna semidefinitna.

Sistemi oblika (2.2) javljaju se u različitim problemima optimizacije, dinamici fluida (Stokesov problem), inkompresibilne elastičnosti, analizi električnih krugova, strukturalnoj analizi itd. Posebno, ovakav sustav se javlja kad promatramo sljedeći problem minimizacije:

$$J(x) \to min \tag{2.3a}$$

$$Bx = g \tag{2.3b}$$

pri čemu je $J(x) = \frac{1}{2}x^T Ax - f^T x$.

U tom slučaju varijabla y predstavlja vektor Lagrangeovih multiplilatora i vrijedi da je svako rješenje od (2.2) ujedno i sedlasta točka funkcionala:

$$L(x, y) = \frac{1}{2}x^T Ax - f^T x + (Bx - g)^T y$$

Definicija 2.1.1. *Neka je $L : R^{m+n} \to R$ proizvoljan funkcional. Točka $(x, y) \in R^{m+n}$ za koju vrijedi:*

$$L(x, z) \leq L(x, y) \leq L(w, y), \forall w \in R^m \; i \; \forall z \in R^n$$

zove se sedlasta točka. Ekvivalentno:

$$\min_w \max_z L(w, z) = L(x, y) = \max_z \min_w L(w, z)$$

Drugi slučaj od velikog značaja je kad su zadovoljeni uvjeti od 1-4, a ne vrijedi 5. U tom slučaju dobijemo sustav oblika:

$$\begin{bmatrix} A & B^T \\ B & -C \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} f \\ g \end{bmatrix} \tag{2.4}$$

Sistemi ovog oblika javljaju se kod mješovite metode konačnih elemenata za parcijalne diferencijalne jednadžbe, raznih problema optimizacije, posebno težinskih najmanjih kvadrata, diskretizacije jednadžbi koje opisuju kompresibilne fluide i kruta tijela itd.

Sustavi sa sedlastom točkom oblika (2.1) prirodno se javljaju u raznim primjenama. Spomenimo neka područja:

- dinamika fluida
- uvjetna i težinska aproksimacija metodom najmanjih kvadrata
- uvjetna optimizacija
- ekonomija
- električni krugovi i mreže
- elektromagnetizam
- financije
- rekonstrukcija slika
- prepoznavanje uzoraka u podacima
- interpolacija rasutih podataka
- linearna elastičnost
- generiranje mreža u računalnoj grafici
- mješovita metoda konačnih elemenata za eliptične parcijalne diferencijalne jednadžbe
- redukcija reda modela kod dinamičkih sistema
- optimalna kontrola
- problemi identifikacije parametara

2.2 Problemi inkompresibilnog toka

Stacionarna Navier-Stokesova zadaća

Promatrajmo stacionarnu Navier-Stokesovu zadaću:

$$\begin{gathered} -\mu\Delta u + (\nabla u)u + \nabla p = f \text{ na } \Omega \\ divu = 0 \text{ na } \Omega \\ Bu = g \text{ na } \Gamma \end{gathered}$$

pri čemu je $\Omega \subset R^d (d = 2, 3)$ ograničena, povezana domena sa dovoljno glatkom granicom Γ, a $f : \Omega \to R^d$ i $g : \Gamma \to R^d$ su poznate funkcije. Koeficijent $\mu > 0$ je konstanta koja opisuje kinematičku viskoznost fluida (broj je obrnuto proporcionalan Reynoldsovom broju). B je neki poznati operator koji djeluje na granici domene (npr. operator traga za Dirichletov rubni uvjet). Kako bi se riješio problem potrebno je odrediti brzinu $u : \Omega \to R^d$ i pritisak (tlak) $p : \Omega \to R^d$. Tako definirana jednadžba opisuje gibanje inkompresibilnog viskoznog fluida. Da bi se pritisak p mogao odrediti jedinstveno, moramo zahtijevati jedan dodatan uvjet:

$$\int_\Omega p dx = 0$$

Ako diskretizaciju zadaće izvršimo metodom konačnih elemenata dobijemo generalizirani sistem sa sedlastom točkom oblika (2.4).

Stacionarna Stokesova jednadžba

Promatrajmo stacionarnu Stokesovu zadaću:

$$\begin{gathered} \Delta u + \nabla p = f \text{ na } \Omega \\ divu = 0 \text{ na } \Omega \\ Bu = g \text{ na } \Gamma \end{gathered}$$

uz jedanke oznake kao i u prethodnom slučaju. Primjetimo da bez gubitka općenitosti ovdje možemo postaviti $\mu = 1$. Podijelimo cijelu jednadžbu sa μ te reskaliramo pritisak i prebacimo faktor $\frac{1}{\mu}$ na f. Ovako definirana zadaća opisuje protok sporogibajućeg vrlo viskoznog fluida. Diskretizacija ove zadaće vrši se analogno prethodnom slučaju, a dobiveni sustav je oblika (2.2) :

$$\begin{bmatrix} A & B^T \\ B & 0(C) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} f \\ g \end{bmatrix}$$

pri čemu je A blok dijagonalna matrica. Svaki dijagonalni blok matrice A predstavlja diskretizaciju Laplaceovog operatora $-\Delta$ sa odgovarajućim rubnim uvjetima. Dakle, A je simetrična pozitivno semidefinitna. Ukoliko se odabere neki postupak stabilizacije, moguće je da se pojavi sustav u kojem je $C \neq 0$.

2.3 Problemi optimizacije

U primjenama se često susreću situacije gdje je potrebno određeni izraz minimizirati uz neke uvjete koji obično slijede iz same prirode problema. Problemi takvog tipa spadaju u granu matematike koja se bavi optimizacijom. Ono što je zajedničko većini takvih problema je da se njihovo rješavanje svodi na linearne sustave sa sedlastom točkom što će se vidjeti iz sljedeća dva primjera.

Problem najmanjih kvadrata

Promatrajmo sljedeći problem najmanjih kvadrata sa pripadnim uvjetom:

$$\min_x \|c - Gy\|_2$$
$$\text{uz uvjet } Ey = d$$

pri čemu je $G \in R^{p\times m}$, $c \in R^p$, $y \in R^m$, $E \in R^{q\times n}$ i $q < m$. Da bi takav minimum postojao, moraju biti zadovoljeni uvjeti optimalnosti koji vode na sustav oblika:

$$\begin{bmatrix} I_p & 0 & G \\ 0 & 0 & E \\ G^T & E^T & 0 \end{bmatrix} \begin{bmatrix} r \\ \lambda \\ y \end{bmatrix} = \begin{bmatrix} c \\ d \\ 0 \end{bmatrix}$$

pri čemu je I_p $p \times p$ matrica identitete, a $\lambda \in R^q$ vektor Lagrangeovih multiplikatora. Očito da je dobiveni sustav specijalan slučaj simetričnog problema sedlaste točke (2.2). Problemi opisanog tipa javljaju se kada je potrebno nekom (često se zahtijeva dovoljna glatkoća) funkcijom interpolirati zadane podatke.

Metode unutarnje točke

Promatrajmo sljedeći problem konveksnog nelinearnog programiranja:

$$minf(x) \tag{2.5a}$$
$$c(x) \leq 0 \tag{2.5b}$$

pri čemu su $f : R^n \to R$ i $c : R^n \to R^m$ konveksne i klase C^2.

Uvođenjem pomoćne nenegativne varijable $z \in R^m$, moguće je uvjet nejednakosti zapisati u obliku jednakosti $c(x) + z = 0$. Na taj način dolazimo do pripadnog pridruženog problema prepreke:

$$minf(x) - \mu \sum_{i=1}^{m} lnz_i$$
$$\text{uz uvjet } c(x) + z = 0$$

Pripadni Langrangeov funkcional dan je sa:

$$L(x, y, z) = f(x) + y^T(c(x) + z) - \mu \sum_{i=1}^{m} lnz_i$$

Prisjetimo se da je točka minimuma ujedno i stacionarna točka danog funkcionala pa je potrebno odrediti pripadne parcijalne derivacije te ih izjednačiti s nulom. Dakle, dobijemo sljedeći sustav:

$$\nabla_x L(x, y, z) = \nabla f(x) + \nabla c(x)^T y = 0$$
$$\nabla_y L(x, y, z) = c(x) + z = 0$$
$$\nabla_z L(x, y, z) = y - \mu Z^{-1} e = 0$$

pri čemu je $Z = diag(z_1, z_2, ..., z_m)$ i $e = [1, 1, ..., 1]^T$. Ako uvedemo dijagonalnu matricu $Y = diag(y_1, y_2, ..., y_m)$, uvjet optimalnosti za promatrani problem prepreke daje sljedeći sustav:

$$\nabla f(x) + \nabla c(x)^T y = 0$$
$$c(x) + z = 0$$
$$YZe = \mu e$$
$$y, z \geq 0$$

Primjetimo da je dobiveni sustav nelinearan sa pripadnim uvjetom nenegativnosti. Moguće ga je rješiti Newtonovom metodom pri čemu se parametar μ postepeno smanjuje kako bi se osigurala konvergencija ka rješenju početnog problema (2.5). U svakom koraku iteracija, potrebno je rješiti sustav oblika:

$$\begin{bmatrix} H(x,y) & B(x)^T & 0 \\ B(x) & 0 & I \\ 0 & Z & Y \end{bmatrix} \begin{bmatrix} \delta x \\ \delta y \\ \delta z \end{bmatrix} = \begin{bmatrix} -\nabla f(x) - B(x)^T y \\ -c(x) - z \\ \mu e - YZe \end{bmatrix}$$

pri čemu je $H(x,y) = \nabla^2 f(x) + \sum_{i=1}^{m} y_i \nabla^2 c_i(x) \in R^{n\times n}$ i $B(x) = \nabla c(x) \in R^{m\times n}$. Napomenimo da $\nabla^2 f(x)$ označava Hesseovu matricu od funkcije f evaluiranu u točki x. Ako izdvojimo treću jednadžbu gornjeg sistema: $Z\delta y + Y\delta z = \mu e - YZe$, možemo zapisati $\delta z = \mu Y^{-1}e - Ze - ZY^{-1}\delta y$ te na taj način reducirati pripadni sustav za jednu dimenziju. Dobijemo sljedeće:

$$\begin{bmatrix} H(x,y) & -B(x)^T \\ B(x) & ZY^{-1} \end{bmatrix} \begin{bmatrix} \delta x \\ -\delta y \end{bmatrix} = \begin{bmatrix} -\nabla f(x) - B(x)^T y \\ -c(x) - \mu Y^{-1} e \end{bmatrix} \tag{2.6}$$

Prisjetimo se da smo zahtijevali konveksnost funkcije cilja f i pripadnih uvjeta $c_i(x)$. Zbog toga je simetrična matrica $H(x,y)$ ujedno i pozitivna semidefinitna te čak i pozitivno definitna ukoliko je f strogo konveksna. Dijagonalna matrica ZY^{-1} je očito pozitivno semidefinitna. Matrica koeficijenata u (2.6) ovisi o trenutnoj aproksimaciji para (x,y) te se mijenja prilikom svakog koraka Newtonove metode. Izuzev predznaka, dobiveni sustav je sustav sa sedlastom točkom oblika (2.4). Slični sustavi dobiju se analognim postupkom kad se metodama unutarnje točke rješavaju problemi linearnog i kvadratičnog programiranja. Tada se u svakom koraku Newtonovih iteracija dobije sustav oblika:

$$\begin{bmatrix} -H - D & B^T \\ B & 0 \end{bmatrix} \begin{bmatrix} \delta x \\ \delta y \end{bmatrix} = \begin{bmatrix} \epsilon \\ \nu \end{bmatrix}$$

pri čemu je $Hn \times n$ simetrična pozitivno semidefinitna ako je problem konveksan i D je dijagonalna matrica. U ovom slučaju matrice H i B su konstantne, a D se mijenja u svakom koraku Newtonovih iteracija. Primjetimo da je dobiveni sustav također sustav sa sedlastom točkom oblika (2.2)

Poglavlje 3

Svojstva matrica sa sedlastom točkom

Da bi se omogućio razvoj efikasnih algoritama za rješavanje sustava sa sedlastom točkom, potrebno je dobro razumijeti svojstva matrica u takvim sustavima. Poznavanje raznih faktorizacija, invertibilnosti, spektralnih svojstava, strukture i kondicioniranja su neka od svojstava koja se mogu iskoristiti u svrhu pronalaska i implementiranja traženih algoritama.

Promatrajmo naš početni sustav (2.2):

$$\begin{bmatrix} A & B^T \\ B & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} f \\ g \end{bmatrix}$$

Jedan od osnovnih problema kod rješavanja sustava ovog tipa je singularnost od A. Još veći problem je taj što se doista u praksi često i susreće singularna matrica A. Kad bismo mogli pretpostaviti da je A regularna matrica, imali bismo dvostruku korist:

- Poboljšanje svojstva promatranog sustava
- Olakšavanje teorijskih razmatranja

Dva pitanja koja se prirodno nameću su: koliko je zahtijev da je A regularna restriktivan? I drugo: da li je uopće moguće zamijeti singularnu matricu A nekom ekvivalentom regularnom matricom koja bi bila relativno jeftina u numeričkom smislu. Odgovor na oba ova pitanja je povoljan kao što će se pokazati primjenivši na naš početni sustav metodu proširenih Lagrangeovih multiplikatora. Dakle, neka je $A = A^T$ simetrična pozitivno semidefintna (dozvoljeno je da bude singularna) i B punog ranga. Tada je sustav (2.2) ekvivalentan sustavu:

$$\begin{bmatrix} A + B^T W B & B^T \\ B & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} f + B^T W g \\ g \end{bmatrix} \tag{3.1}$$

pri čemu je matrica W reda $m \times m$ simetrična pozitivno semidefinitna adekvatno odabrana. Najjednostavniji izbor za W jest $W = \gamma I$ pri čemu je I matrica identitete i $\gamma > 0$. U tom slučaju (1,1) blok u sustavu (3.1) je regularan i pozitivno definitan uz uvjet da je A pozitivno definitna na $\ker(B)$. Općenito, ideja je odabrati W tako da je novi sustav lakše riješiti od početnog, posebno upotrebom iterativnih metoda. Kad se koristi $W = \gamma I$, u praksi se pokazalo da je dobro postaviti $\gamma = \|A\|_2 / \|B\|_2^2$. Uvjet da je B punog ranga, koji je korišten prilikom izricanja gornje tvrdnje, bit će objašnjen u napomeni 3.2.5.

3.1 Schurov komplement

Neka je $Mn \times n$ matrica zapisana kao 2×2 blok matrica:

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

pri čemu je $Ap \times p$ matrica, $Bp \times q$ matrica, $Cq \times p$ matrica i $Dq \times q$ matrica. Ako probamo riješiti sustav:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} f \\ g \end{bmatrix},$$

to jest:

$$\begin{aligned} Ax + By &= f \\ Cx + Dy &= g \end{aligned}$$

primjenom Gaussovih eliminacija, uz pretpostavku da je D invertibilna, dobijemo za y:

$$y = D^{-1}(g - Cx)$$

Ako dobiveni izraz uvrstimo u prvu jednadžbu dobijemo:

$$Ax + B(D^{-1}(g - Cx)) = f$$

što nakon manjeg sređivanja po x daje:

$$(A - BD^{-1}C)x = f - BD^{-1}g$$

Ako je matrica uz x invertibilna, tada smo dobili rješenje početnog sustava:

$$x = (A - BD^{-1}C)^{-1}(f - BD^{-1}g)$$
$$y = D^{-1}(g - C(A - BD^{-1}C)^{-1}(f - BD^{-1}g))$$

Definicija 3.1.1. *Matrica $A - BD^{-1}C$ zove se Schurov komplement od D u M dok je matrica $D - CA^{-1}B$ Schurov komplement od A u M. U slučaju da su A ili D singularne matrice tada njihove inverze računamo u generaliziranom smislu te govorimo o generaliziranom Schurovom komplementu.*

Poznavajući Schurov komplement, sad smo spremni iskazati određene blok faktorizacije koje će nam biti korisne prilikom rješavanja sustava sa sedlastom točkom. Prisjetimo se da smo naš sustav u kojem je na (1,1) blok mjestu bila singularna matrica A zamijenili ekvivalentnim sustavom gdje je na istom mjestu sada regularna matrica tako da sada općenito možemo govoriti o nesingularnoj matrici A. Pod tim uvjetima za matricu sustava sa sedlastom točkom M vrijedi sljedeća blok triangulacijska faktorizacija:

$$M = \begin{bmatrix} A & B_1^T \\ B_2 & -C \end{bmatrix} = \begin{bmatrix} I & 0 \\ B_2A^{-1} & I \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & S \end{bmatrix} \begin{bmatrix} I & A^{-1}B_1^T \\ 0 & I \end{bmatrix} \tag{3.2}$$

pri čemu je $S = -(C + B_2A^{-1}B_1^T)$ Schurov komplement od A u M. Istaknimo i dvije korisne ekvivalentne faktorizacije:

$$M = \begin{bmatrix} A & 0 \\ B_2 & S \end{bmatrix} \begin{bmatrix} I & A^{-1}B_1^T \\ 0 & I \end{bmatrix}$$

te

$$M = \begin{bmatrix} I & 0 \\ B_2A^{-1} & I \end{bmatrix} \begin{bmatrix} A & B_1^T \\ 0 & S \end{bmatrix} \tag{3.3}$$

Napomena 3.1.2. *Promatrajući gornje faktorizacije, možemo zaključiti da je A regularna ako i samo ako je S regularna. Nažalost, trenutno se jako malo može reći o invertibilnosti od S u općenitom slučaju. Zbog toga je potrebno uvesti određene restrikcije na matrice A, B_1, B_2 i C.*

3.2 Uvjeti postojanja rješenja

Simetrični slučaj

Promatramo sustav sa sedlastom točkom definiran u drugom poglavlju (2.2). Tada je A simetrična pozitivno definitna, $B_1 = B_2$ i $C = 0$. Schurov komplement je tada $S = -BA^{-1}B^T$ simetrična negativna semidefinitna matrica. Primjetimo da je S invertibilna ako i samo ako je B^T punog ranga. Također, onda je jasno da analogno vrijedi i za M. Tada problemi (2.2) i (2.3) imaju jedinstveno rješenje i vrijedi: Ako je (x_*, y_*) rješenje od (2.2) , onda je x_* jedinstveno rješenje od (2.3). Također, moguće je pokazati da je x_* ortogonalna projekcija rješenja $\hat{x} = A^{-1}f$ bezuvjetnog problema (2.3) na skup uvjeta $C = \{x \in R^n; Bx = g\}$ uz skalarni produkt $< v, w >= w^T Av$.

Promotrimo slučaj kad je A simetrična pozitivno definitna, $B_1 = B_2 = B$ i $C \neq 0$ simetrična pozitivno semidefinitna. Ponovno je $S = -C - BA^{-1}B^T$ simetrična negativno semidefinitna te je invertibilna ako i samo je $ker(C) \cap ker(B^T) = 0$. Dakle, očito da je za invertibilnost potrebno zahtijevati da je C pozitivno definitna ili da je B punog ranga.

Teorem 3.2.1. *Neka je A simetrična pozitivno definitna, $B_1 = B_2 = B$ i C simetrična pozitivno semidefinitna. Ako vrijedi $ker(C) \cap ker(B^T) = 0$, onda je matrica M sa sedlastom točkom regularna. Posebno, M je invertibilna ako je B punog ranga.*

Napomena 3.2.2. *Pokušajmo oslabiti uvjet da je A pozitivno definitna. Ako je A indefinitna tada je moguće da M bude singularna čak i ako je B punog ranga što pokazuje sljedeći primjer:*

$$\left[\begin{array}{cc|c} 1 & 0 & -1 \\ 0 & -1 & 1 \\ \hline -1 & 1 & 0 \end{array}\right] = \begin{bmatrix} A & B^T \\ B & 0 \end{bmatrix}$$

Ipak, M će biti invertibilna ukoliko je A pozitivno definitna na $ker(B)$.

Teorem 3.2.3. *Neka je A simetrična pozitivno semidefinitna, $B_1 = B_2 = B$ punog ranga i $C = 0$. Tada je nužan i dovoljan uvjet da matrica M sa sedlastom točkom bude regularna $ker(A) \cap ker(B) = 0$.*

Dokaz. Dokaz ćemo podijeliti u dva dijela. Pokažimo prvo nužnost uvjeta. Pretpostavimo da je $ker(A) \cap ker(B) \neq 0$. Tada očito postoji $x \in ker(A) \cap ker(B)$ takav da je $x \neq 0$. Ako uzmemo da je $u = [x, 0]^T$ tada je $Mu = 0$ što implicira da je M singularna. Dakle, kontradikcija pa je

uvjet nužan. Pokažimo sada dovoljnost. Neka je $u = [x, y]^T$ takav da je $Mu = 0$. Raspišemo i dobijemo $Ax + B^T y = 0$ i $Bx = 0$. Iz prve jednadžbe slijedi:

$$Ax = -B^T y$$
$$x^T Ax = -x^T B^T y = -(Bx)^T y = 0$$

Budući da je A simetrična pozitivno semidefinitna, slijedi da $x^T Ax = 0$ implicira $Ax = 0$ to jest, $x \in ker(A) \cap ker(B)$ i u konačnici $x = 0$. Ako uvrstimo $Ax = 0$ u prvu jednadžbu dobijemo da je $B^T y = 0$ što povlači da je $y = 0$ jer je B punog ranga. Dakle, $u = 0$ i M je regularna. Time je pokazana dovoljnost i tvrdnja teorema. □

Napomena 3.2.4. *Iz dokaza prethodnog teorema jasno je da se uvjet pozitivne semidefinitnosti za A može dodatno oslabiti. Naime, dovoljno je zahtijevati da A bude definitna na ker(B). U biti, koristili smo samo $x^T Ax \neq 0$ za $x \in ker(B), x \neq 0$. Ta tvrdnja implicira da A mora biti ili pozitivno definitna ili negativno definitna na $ker(B)$. Oba ova uvjeta vode na isti zaključak: Da bi M bila regularna, nužno je zahtijevati da rang od A bude barem $n - m$.*

Napomena 3.2.5. *U do sada izrečenim tvrdnjama smo koristili da je B punog ranga. Pitanje koliko je takva pretpostavka restriktivna? Da bismo odgovorili na to pitanje, potrebno je promatrati samu prirodu problema koji je generirao sustav sa sedlastom točkom. Ukoliko B nije punog ranga, to znači da su neki uvjeti postavljeni na sustav redudantni. U praksi se pokazalo da je tu redudantnost relativno jednostavno eliminirati. Na primjer, u Stokesovoj jednadžbi , gdje B^T predstavlja diskretizirani gradijent, često se javlja dimenzija od $ker(B) = 1$. Dakle M ima jednu trivijalnu svojestvenu vrijednost koja odgovara hidrostatičkom pritisku, budući da je pritisak definiran do na aditivnu konstantu. Slična situacija javlja se i kod električnih mreža gdje vektor y predstavlja nodalne potencijale koji su također definirani do na aditivnu konstantu. Ovdje se B može proširiti do punog ranga tako da se točno precizira vrijednost potencijala u jednoj točki. Potencijalni problem sa ovim pristupom je da rezultirajući linearni sustav može ponovno biti loše uvjetovan. Doduše često to nije slučaj zbog konzistente konstrukcije sustava $Mu = b$ pa nije ni potrebno eliminirati singularitet u M. Također, postoje iterativne metode kao GMRES na koje uglavnom ne utječe pojavljivanje jedne trivijalne svojstvene vrijednosti, pogotovo kad se u prvom koraku koristi $u_0 = 0$.*

Općeniti slučaj

Teorem 3.2.6. *Neka je matrica*

$$M = \begin{bmatrix} A & B_1^T \\ B_2 & 0 \end{bmatrix}$$

regularna. Tada vrijedi: rang(B_1) = m i rang $\binom{A}{B_2} = n$.

Dokaz. Pretpostavimo da je rang(B_1) $< m$. Tada postoji vektor $0 \neq y \in R^m$ takav da je $B_1^T y = 0$. Ako definiramo da je $u = [0, y]^T$ dobijemo $Mu = 0$, kontradikcija. Pretpostavimo da je rang $\binom{A}{B_2} < n$. Tada postoji vektor $0 \neq x \in R^n$ takav da je $\binom{A}{B_2} x = 0$. Ako definiramo $u = [x, 0]^T$ dobijemo $Mu = 0$, kontradikcija. □

Kako bi se osigurala invertibilnost od M potrebno je uvesti dodatne uvjete. Sljedeći teorem daje nužne i dovoljne uvjete za invertibilnost od M kada je $B_1 = B_2$.

Teorem 3.2.7. *Neka je H , simetrični dio od A, pozitivna semidefinitna matrica, $B_1 = B_2 = B$ punog ranga i C simetrična pozitivna semidefinitna (moguće i nul-matrica). Tada vrijedi sljedeće:*

- *ker(H)∩ ker(B) = 0 ⇒ M je invertibilna*
- *M invertibilna ⇒ ker(A)∩ ker(B) = 0*

Dokaz. Dokažimo prvu tvrdnju. Neka je $u = [x, y]^T$ takav da je $Mu = 0$. Tada vrijedi $Ax + B^T y = 0$ i $Bx - Cy = 0$. Računamo:

$$\begin{aligned} Ax &= -B^T y \\ x^T Ax &= -x^T B^T y \\ x^T Ax &= -(Bx)^T y \\ x^T Ax &= -(Cy)^T y \\ x^T Ax + y^T Cy &= 0 \end{aligned}$$

Pokažimo da je $x^T Ax \geq 0$. Koristimo da je H pozitivna semidefintna i dobijemo da vrijedi sljedeće:

$$0 \leq x^T Hx \leq \tfrac{1}{2}(x^T Ax + x^T A^T x) = -\tfrac{1}{2} y^T Cy + \tfrac{1}{2} x^T A^T x.$$

Kako je $y^TCy \geq 0$ jer je C pozitvna semidefinitna zaključujemo da je x^TA^Tx realna što znači da vrijedi:

$$0 \leq x^THx \leq \tfrac{1}{2}(x^TAx + x^TA^Tx) = \tfrac{1}{2}x^TAx\,,$$

to jest $x^TAx \geq 0$. Dakle, imamo da je $x^TAx \geq 0$ i $y^TCy \geq 0$ pa zaključujemo $x^TAx = 0$ i $y^TCy = 0$. Budući da $x^TAx = 0 \implies x^THx = 0$, slijedi $x \in ker(H)$ zbog H simetrična pozitivno definitna. Slično, iz $y^TCy = 0$ zaključujemo da je $Cy = 0$. Uvrstimo u $Bx - Cy = 0 = Bx$. Slijedi da je $x = 0$ jer je $x \in ker(H) \cap ker(B) = 0$. Vratimo se u $Ax = -B^Ty$. Preostaje samo izraz $-B^Ty = 0$ odakle zaključujemo da je $y = 0$ jer je B punog ranga. Slijedi da je $u = 0$ što znači da je jedino rješenje jednadžbe $Mu = 0$ trivijalno. Dakle, M je regularna matrica. Dokaz druge tvrdnje ekvivalentan je dokazu druge tvrdnje u (3.2.3). □

Napomena 3.2.8. *Pokažimo primjerima da obrati iz prethodnog teorema ne vrijede. Kako bi pokazali da obrat prve tvrdnje nije istinit, promotrimo primjer:*

$$M = \left[\begin{array}{ccc|c} 1 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ \hline 0 & 0 & 1 & 0 \end{array}\right] = \begin{bmatrix} A & B^T \\ B & 0 \end{bmatrix}$$

Lako se provjeri da je M invertibilna te $ker(H) \cap ker(B) = span[0,1,0]^T \neq 0$.

Da bi provjerili neistinitost obrata druge tvrdnje, poslužimo se sljedećim primjerom:

$$M = \left[\begin{array}{cc|c} 0 & -1 & 0 \\ 1 & 1 & 1 \\ \hline 0 & 1 & 0 \end{array}\right] = \begin{bmatrix} A & B^T \\ B & 0 \end{bmatrix}$$

Ponovno se laganim računom provjeri da je $ker(A) \cap ker(B) = 0$ dok je M singularna.

3.3 Inverz matrica sa sedlastom točkom

Koristeći pretpostavku da je A regularna, pokazali smo da je M invertibilna ako i samo ako je $S = -(C + B_2A^{-1}B_1^T)$ regularna. U tom slučaju postoji egzaktna formula za inverz od M:

$$M^{-1} = \begin{bmatrix} A^{-1} + A^{-1}B_1^TS^{-1}B_2A^{-1} & -A^{-1}B_1^TS^{-1} \\ -S^{-1}B_2A^{-1} & S^{-1} \end{bmatrix} \tag{3.4}$$

Međutim, generalna formula za inverz nije od naročitog interesa za praktičnu primjenu. Od većeg interesa su posebni slučajevi. Jedan zanimljivi specijalan slučaj jest kad je A simetrična

pozitivno definitna, $B_1 = B_2 = B$, $C = 0$, $S = -BA^{-1}B^T$ regularna i $g = 0$. Tada koristeći formulu za egzaktan inverz od M^{-1}, možemo vidjeti da je rješenje (x_*, y_*) sustava (2.2) dano izrazom:

$$\begin{bmatrix} x_* \\ y_* \end{bmatrix} = \begin{bmatrix} (I + A^{-1}B^T S^{-1}B)A^{-1}f \\ -S^{-1}BA^{-1}f \end{bmatrix} \tag{3.5}$$

Laganim računom može se provjeriti da vrijedi sljedeće:

- $\Pi = -A^{-1}B^T S^{-1}B = A^{-1}B^T(BA^{-1}B^T)^{-1}B$

- $\Pi = \Pi^2$

Provjerimo da je $\Pi = \Pi^2$. Računamo:

$$\begin{aligned} \Pi^2 &= -A^{-1}B^T S^{-1}B(-A^{-1}B^T S^{-1}B) \\ &= A^{-1}B^T S^{-1}(BA^{-1}B^T)S^{-1}B \\ &= A^{-1}B^T S^{-1}(-S)S^{-1}B \\ &= -A^{-1}B^T S^{-1}B = \Pi \end{aligned}$$

Dakle, Π je projektor. Ali, vrijedi i više. Relacije:

$$\Pi v \in \mathscr{C}(A^{-1}B^T) \text{ i } v - \Pi v \perp \mathscr{C}(B^T) \text{ za svaki } v \in R^n$$

pokazuju da je Π projektor na $R(A^{-1}B)$ i ortogonalan na $R(B^T)$ pri čemu sa $R(A)$ označavamo vektorski prostor razapet stupcima matrice A. Koristeći ove relacije možemo zapisati prvu komponentu iz (3.5) kao:

$$x_* = (I - \Pi)\hat{x} \text{ pri čemu je } \hat{x} = A^{-1}f \text{ rješenje bezuvjetnog problema (2.3)}$$

Dodatno, vrijedi $\hat{x} = \Pi\hat{x} + x_*$ što znači da se rješenje bezuvjetnog problema (2.3) može rastaviti na dio koji je u $R(A^{-1}B)$ i dio koji je ortogonalan na $R(B^T)$. Koristeći $f - B^T y_* = Ax_*$ i $Bx_* = 0$ zaključujemo da vrijedi:

$$0 = Bx_* = (BA^{-1})(Ax_*) = (A^{-1}B^T)^T(f - B^T y_*) \tag{3.6}$$

Po pretpostavci je A^{-1} simetrična pozitivno definitna pa je funkcijom $< v, w >_{A^{-1}} := w^T A^{-1} v$ definiran skalarni produkt. Tada (3.6) pokazuje da je vektor $f - B^T y_* \in f + R(B^T)$ ortogonalan na prostor $R(B^T)$ s obzirom na A^{-1}-skalarni produkt. To znači da je y_* rješenje generaliziranog problema najmanjih kvadrata u odnosu na A^{-1}-normu induciranu A^{-1}-skalarnim produktom, $||v||_{A^{-1}} := (< v, v >_{A^{-1}})^{1/2}$:

$$||f - B^T y_*||_{A^{-1}} = \min_u ||f - B^T u||_{A^{-1}}$$

Sljedeći važan specijalan slučaj je kad su A i C simetrične pozitivno definitne i $B_1 = B_2 = B$. Tada je odgovarajuća matrica M sa sedlastom točkom kvazidefinitna neovisno o rangu od B. Diretktnom primjenom formule za inverz (3.4) možemo primjetiti da ukoliko je M kvazidefinitna tada je i M^{-1} kvazidefinitna.

Definicija 3.3.1. *Za matricu H kažemo da je kvazidefinitna ako postoji matrica permutacije P takva da vrijedi:*

$$H = P^T Q P = \begin{bmatrix} Q_{1,1} & Q_{1,2} \\ Q_{2,1}^* & -Q_{2,2} \end{bmatrix}$$

pri čemu su $Q_{1,1}$ i $Q_{2,2} + Q_{1,2}^* Q_{1,1}^{-1} Q_{1,2}$ pozitivno definitne. Posebno, u našem slučaju je $Q = M$ simetrična pa možemo reći da je simetrična matrica M kvazidefinitna ako postoji matrica permutacije P takva da vrijedi:

$$M = P^T M P = \begin{bmatrix} A & B^T \\ B & -C \end{bmatrix}$$

Alternativnu formulu za inverz matrica ovog tipa možemo dobiti i ako oslabimo zahtijev na A. U ovom slučaju dozvoljavamo da A bude singularna, ali pretpostavljamo da je $B_1 = B_2 = B$ punog ranga i $C = 0$. Sa $Z \in R^{n \times (n-m)}$ označimo proizvoljnu matricu čiji stupci razapinju $ker(B)$. Ukoliko je H, simetrični dio od A, pozitivna i semidefinitna tada primjenom teorema 3.2.7 zaključujemo da je $(n - m) \times (n - m)$ matrica $Z^T A Z$ invertibilna (njezin simetrični dio $Z^T H Z$ je pozitivno definitan). Označimo $W := Z(Z^T A Z)^{-1} Z^T$ i dobijemo sljedeći izraz za inverz od M:

$$M^{-1} = \begin{bmatrix} W & (I - WA)B^T (BB^T)^{-1} \\ (BB^T)^{-1} B(I - AW) & -(BB^T)^{-1} B(A - AWA)B^T (BB^T)^{-1} \end{bmatrix} \tag{3.7}$$

Provjerimo da je sa (3.7) dan izraz za inverz od M. Koristit ćemo pomoćnu tvdnju da je $B^T (BB^T)^{-1} B = I - ZZ^T$ koji slijedi iz jednakosti: $BB^T (BB^T)^{-1} B = B$ i $B(I - ZZ^T) = B$ jer je Z baza za $ker(B)$.

Izrazom MM^{-1} dobijemo četri jednakosti koje treba provjeriti:

- $AW + B^T[(BB^T)^{-1}B(I - AW)] = I$
- $A[(I - WA)B^T(BB^T)^{-1}] + B^T[-(BB^T)^{-1}B(A - AWA)B^T(BB^T)^{-1}] = 0$
- $BW = 0$
- $B[(I - WA)B^T(BB^T)^{-1}] = I$

gdje je treći uvjet trivijalno zadovoljen jer je $BZ = 0$.
Provjerimo 1:

$$\begin{aligned} AW + B^T[(BB^T)^{-1}B(I - AW)] &= \\ = AW + (I - ZZ^T)(I - AW) & \\ = AW + I - AW - ZZ^T + ZZ^TAW & \\ = I - ZZ^T + ZZ^TAZ(Z^TAZ)^{-1}Z^T & \\ = I - ZZ^T + ZZ^T = I & \end{aligned}$$

Provjerimo 2:

$$\begin{aligned} A[(I - WA)B^T(BB^T)^{-1}] + B^T[-(BB^T)^{-1}B(A - AWA)B^T(BB^T)^{-1}] &= \\ = (I - B^T(BB^T)^{-1}B)[(A - AWA)B^T(BB^T)^{-1}] & \\ = (I - I + ZZ^T)[(A - AZ(Z^TAZ)^{-1}Z^TA)B^T(BB^T)^{-1}] & \\ = ZZ^T[(A - AZ(Z^TAZ)^{-1}Z^TA)B^T(BB^T)^{-1}] & \\ = [ZZ^TA - ZZ^TAZ(Z^TAZ)^{-1}Z^TA]B^T(BB^T)^{-1} & \\ = (ZZ^TA - ZZ^TA)B^T(BB^T)^{-1} = 0 & \end{aligned}$$

Provjerimo 4:

$$
\begin{aligned}
B[(I - WA)B^T(BB^T)^{-1}] = \\
= (B - BWA)B^T(BB^T)^{-1} = \\
= BB^T(BB^T)^{-1} = I
\end{aligned}
$$

3.4 Spektralna svojstva matrica sa sedlastom točkom

Analizu spektralnih svojstava ograničavamo na dva standardna slučaja. Neka je A simetrična pozitivno definitna, $B_1 = B_2 = B$ punog ranga i C simetrična pozitivno semidefinitna (moguće i nul-matrica). Tada koristeći (3.2) vrijedi sljedeće:

$$
\begin{bmatrix} I & 0 \\ -BA^{-1} & I \end{bmatrix} \begin{bmatrix} A & B^T \\ B & -C \end{bmatrix} \begin{bmatrix} I & -A^{-1}B^T \\ 0 & I \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & S \end{bmatrix} \tag{3.8}
$$

gdje je Schurov komplement od A $S = -(C + BA^{-1}B^T)$ simetrična negativno definitna. Iz ovog raspisa možemo zaključiti da je M kongruentna blok dijagonalnoj matrici $\left(\begin{smallmatrix} A & 0 \\ 0 & S \end{smallmatrix}\right)$.

Teorem 3.4.1 (Sylvesterov zakon inercije)**.** *Neka je A simetrična kvadratna matrica reda n sa realnim vrijednostima. Ako regularna matrica R istog reda veličina transformira A u neku drugu simetričnu matricu $B = RAR^T$, također reda n, onda kažemo da su matrice A i B kongruentne. Koristeći ovu transformaciju, simetrična matrica A uvijek se može transformirati u dijagonalnu matricu D koja na dijagonali ima samo vrijednosti 0, 1 i -1. Sylvesterov zakon inercije tvrdi da ukupan broj pojavljivanja 0, 1 i -1 na dijagonali od D ne ovisi o matrici R kojom je obavljena transformacija. Ako sa n_+ označimo ukupan broj jedinica na dijagonali, sa n_- ukupan broj minus jedinica i sa n_0 ukupan broj nula tada očito vrijedi jedankost: $n_+ + n_- + n_0 = n$. Također, primjetimo da je broj n_0 dimenzija jezgre od A. U terminu svojstvenih vrijednosti, brojevi n_+ i n_- označavaju ukupan broj pozitivnih odnosno negativnih svojstvenih vrijednosti matrice A.*

Koristeći Sylvesterov zakon inercije, zaključujemo da je matrica M u (3.8) indefinitna sa n pozitivnih i m negativnih svojstvenih vrijednosti. Primijetimo da u našem zaključku nije nužno da je B punog ranga. Ako pretpostavimo da $S = -(C + BA^{-1}B^T)$ nije punog ranga onda vrijedi $rang(S) = m - r$ gdje je $r \leq m$, a M ima n pozitivnih, $m - r$ negativnih i r trivijalnih svojstvenih vrijednosti. Primijetimo da ovaj zaključak ostaje vrijediti čak i kad je A samo pozitivno semidefinitna, uz uvjet da je $ker(A) \cap ker(B) = 0$. Općenito, ukoliko m nije

puno manji od n, matrica M će imati visoki stupanj indefinitnosti to jest, mnogo svojstvenih vrijednosti oba predznaka. Nažalost, ovo je čest slučaj u praksi.

Teorem 3.4.2. *Neka je A simetrična pozitivno definitna, $B_1 = B_2 = B$ punog ranga i $C = 0$. Neka μ_1 i μ_n označavaju najveću i najmanju svojstvenu vrijednost od A, a σ_1 i σ_m najveću i najmanju singularnu vrijednost od B. $\sigma(M)$ označava spektar od M. Tada vrijedi:*

$$\sigma(M) \subset I^- \cup I^+$$

pri čemu je

$$I^- = \left[\tfrac{1}{2}\left(\mu_n - \sqrt{\mu_n^2 + 4\sigma_1^2}\right), \tfrac{1}{2}\left(\mu_1 - \sqrt{\mu_1^2 + 4\sigma_m^2}\right)\right]$$

i

$$I^+ = \left[\mu_n, \tfrac{1}{2}\left(\mu_1 + \sqrt{\mu_1^2 + 4\sigma_1^2}\right)\right]$$

Dokaz u [8].

Ograničenja iz prethodng teorema mogu se iskoristiti kod predviđanja brzine konvergencije kod nekih iterativnih metoda (naročito MINRES) kao i kod ocjenjivanja stabilnosti diskretizacije u mješovitoj metodi konačnih elemenata.

U općenitom slučaju malo je toga poznato o svojstvenim vrijednostima od M osim da u većini promatrnih slučajeva od interesa konveksna ljuska svojstvenih vrijednosti od M sadrži ishodište. Ako promatramo slučaj $A \neq A^T$, $B_1 = B_2 = B$ i $C = C^T$ tada dobijemo da je simetrični dio od M dan izrazom:

$$\tfrac{1}{2}(M + M^T) = \begin{bmatrix} H & B^T \\ B & -C \end{bmatrix}$$

gdje je H simetrični dio od A. Ako je H pozitivno definitna onda je simetrični dio od M indefinitan pa zaključujemo da M ima svojstvene vrijednosti sa obje strane imaginarne osi.

Općenito je indefinitnost matrice M negativno svojstvo te ga je potrebno popraviti. Jednostavnom transformacijom početnog sustava to je i moguće. Pretpostavimo da je $B_1 = B_2 = B$. Tada je sljedeći problem sa sedlastom točkom:

$$\begin{bmatrix} A & B^T \\ B & -C \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} f \\ g \end{bmatrix}$$

ekvivalentan sustavu:

$$\begin{bmatrix} A & B^T \\ -B & C \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} f \\ -g \end{bmatrix} \tag{3.9}$$

kojeg označavamo sa $\hat{M}u = \hat{b}$.

Primijetimo da vrijedi:

$$\hat{M} = JM \tag{3.10}$$

pri čemu je $J = \begin{bmatrix} I_n & 0 \\ 0 & -I_m \end{bmatrix}$

što znači da je $\hat{M}$ regularna ukoliko je M regularna.

Teorem 3.4.3. *Neka je $\hat{M}$ matrica sustava u (3.9). Pretpostavimo da je $H = \frac{1}{2}(A+A^T)$ pozitivna semidefinitna, $B_1 = B_2 = B$ punog ranga, C simetrična pozitivno definitna i $ker(H) \cap ker(B) = 0$. Neka je $\sigma(\hat{M})$ spektar od $\hat{M}$. Tada vrijedi:*

- *$\hat{M}$ je pozitivno semidefintina u smislu da za svaki $v \in R^{m+n}$ vrijedi $v^T \hat{M} v \geq 0$*
- *$\hat{M}$ je pozitivno semistabilna to jest, sve svojstvene vrijednosti od $\hat{M}$ leže u desnoj poluravnini: $Re(\lambda) \geq 0, \forall \lambda \in \sigma(\hat{M})$*
- *ako još vrijedi da je $H = \frac{1}{2}(A + A^T)$ pozitivno definitna, tada je $\hat{M}$ pozitvno stabilna: $Re(\lambda) > 0, \forall \lambda \in \sigma(\hat{M})$*

Dokaz. Da bi dokazali prvu tvrdnju pokažimo da vrijedi: $v^T \hat{M} v = v^T \hat{H} v, \forall v \in R^{m+n}$ pri čemu je

$$\hat{H} = \tfrac{1}{2}(\hat{M} + \hat{M}^T) = \begin{bmatrix} H & 0 \\ 0 & C \end{bmatrix}$$

simetrični dio od $\hat{M}$. Uzmimo proizvoljan $v = [v_1, ..., v_{m+n}]^T$ i označimo ga u obliku $[v_n, v_m]^T$ gdje je v_n prvih n komponenti, a v_m sljedećih m komponenti vektora v. Tada vrijedi:

$$v^T \hat{M} v = \begin{bmatrix} v_n & v_m \end{bmatrix} \begin{bmatrix} A & B^T \\ -B & C \end{bmatrix} \begin{bmatrix} v_n \\ v_m \end{bmatrix} =$$

$$
\begin{aligned}
&= v_n^T A v_n - v_m^T B v_n + v_n^T B^T v_m + v_m^T C v_m = \\
&= v_n^T A v_n + v_m^T C v_m \\
&= v_n^T \tfrac{1}{2}(A + A^T) v_n + v_m^T C v_m \\
&= v_n^T H v_n + v_m^T C v_m \\
&= v^T \hat{H} v
\end{aligned}
$$

Očito je $\hat{H}$ pozitivno semidefinitna pa vrijedi $v^T \hat{M} v \geq 0$.

Dokaz druge tvrdnje: Neka je (λ, v) svojstveni par od M takav da je $\|v\|_2 = 1$. Tada vrijedi $v^* \hat{M} v = \lambda$ i $(v^* \hat{M} v)^* = v^* \hat{M}^T v = \bar{\lambda}$. Zbrojivši ove jedankosti dobijemo $\frac{1}{2} v^* (\hat{M} + \hat{M}^T) v = \frac{\lambda + \bar{\lambda}}{2} = Re(\lambda)$. Primjetimo da je:

$$v^* (\hat{M} + \hat{M}^T) v = Re(v)^T (\hat{M} + \hat{M}^T) Re(v) + Im(v)^T (\hat{M} + \hat{M}^T) Im(v)$$

realna nenegativna veličina jer je $Re(v)^T (\hat{M} + \hat{M}^T) Re(v) \geq 0$ i $Im(v)^T (\hat{M} + \hat{M}^T) Im(v) \geq 0$ što znači $Re(\lambda) \geq 0$.

Dokaz treće tvrdnje: Neka je (λ, v) svojstveni par od $\hat{M}$ takav da je $v = [x, y]^T$. Tada vrijedi:

$$
\begin{aligned}
Re(\lambda) &= x^* H x + y^* C y \\
&= Re(x)^T H Re(x) + Im(x)^T H Im(x) + Re(y)^T C Re(y) + Im(y)^T C Im(y) \geq 0
\end{aligned}
$$

Jednakost se postiže samo kad je $x = 0$ i $Cy = 0$. Ali ako je $x = 0$ onda iz jednadžbe $\hat{M} v = \lambda v$ slijedi da je $B^T y = 0$ što, budući da je B punog ranga povlači $y = 0$. Dakle, $v = 0$ što je kontradikcija. □

U slučaju kad su A i C simetrične, moguće je konstruirati zanimljivu algebru. Primjetimo da je sada M simetrična indefinitna, a $\hat{M}$ nesimetrična pozitvno semidefintna i vrijedi:

$$J\hat{M} = \hat{M}^T J \text{ pri čemu je } J \text{ definirana kao u (3.10).}$$

Za $\hat{M}$ kažemo da je J-simetrična ili pseudosimetrična. Preciznije, $\hat{M}$ je simetrična u odnosu na skalarni produkt definiran na R^{m+n} kao $< v, w >:= w^T J v$. Obratno, svaka J-simetrična matrica je oblika $\begin{bmatrix} A & B^T \\ -B & C \end{bmatrix}$ za neke $A \in R^{n \times n}, B \in R^{m \times n}$ i $C \in R^{m \times m}$ pri čemu su A i C simetrične. Ako označimo skup svih J-simetričnih matrica kao:

$$\mathbb{J} = \left\{ \begin{bmatrix} A & B^T \\ -B & C \end{bmatrix}; A = A^T \in R^{n \times n}, B \in R^{m \times n}, C = C^T \in R^{m \times m} \right\}$$

tada se može pokazati da je trojka $(\mathbb{J}, +, *)$ neasocijativna, komutativna algebra nad poljem relanih brojeva uz operacije zbrajanja matrica i operaciju množenja Jordanovim produktom definiranog kao $F * G := \frac{1}{2}(FG + GF)$. Ovako definirana algebra poznata je kao Jordanova algebra pridružena realnoj Lievoj grupi $O(n, m, R)J$-ortogonalnih matrica (grupa svih matrica $Q \in R^{m+n}$ takve da vrijedi $Q^T JQ = J$).

Može se pokazati da u određenim specijalnim slučajevima svojstvene vrijednosti od $\hat{M}$ su sve realne i pozitivne što je odlično svojstvo sa stajališta iterativnih metoda. Da bi opravdali egzistenciju takvog slučaja potreban nam je sljedeći pomoćni rezultat:

Propozicija 3.4.4. *Neka je A simetrična, B^T punog ranga i $C = \beta I_m$, $\beta \in R$. Tada vrijedi:*

- *Ako je β svojstvena vrijednost od $\hat{M}$ sa pripadnim svojstvenim vektorom $x = [u, v]^T$, tada je $u \in ker(B)$ i β je svojstvena vrijednost od A pridružena u ako i samo ako je $v = 0$.*

- *Neka je $(\lambda, [u, v]^T)$ svojstveni par od $\hat{M}$ takav da je $u \neq 0$. Tada je β svojstvena vrijednost od A pridružena u ako i samo ako je $\lambda = \beta$. Tada je nužno $v = 0$ i $u \in ker(B)$.*

Dokaz u [3].

Teorem 3.4.5. *Neka je A simetrična pozitivno definitna, $B_1 = B_2 = B$ punog ranga i $C = \beta I_m$, $\beta \geq 0$. Neka je $(\lambda, [u, v]^T)$ svojstveni par od $\hat{M}$. Tada je $\lambda = \beta$ ili $\lambda \in R$ ako i samo ako*

$$(\frac{u^*(A+\beta I_n)u}{u^*u})^2 \geq 4\frac{u^*(B^TB+\beta A)u}{u^*u}$$

Dokaz. Ako je $\lambda = \beta \geq 0$ tada je $\lambda \in R$. U slučaju $\lambda \neq \beta$ koristimo $-Bu+Cv = \lambda v$ da dobijemo $v = \frac{-Bu}{\lambda-\beta}$ i uvrstimo u $Au + B^T v = \lambda u$. Nakon sređivanja dobijemo izraz:

$$\lambda^2 u - \lambda(A + \beta I_n)u + (B^T B + \beta A)u = 0. \tag{3.11}$$

Za $u \neq 0$ pomnožimo (3.11) sa lijeve strane sa u^* i dobijemo kvadratnu jednadžbu po λ sa realnim koeficijentima:

$$\lambda^2 u^*u - \lambda u^*(A + \beta I_n)u + u^*(B^T B + \beta A)u = 0$$

čija su rješenja dana sa:

$$\lambda_{1,2} = \frac{1}{2}\frac{u^*(A+\beta I_n)u}{u^*u} \pm \sqrt{(\frac{u^*(A+\beta I_n)u}{u^*u})^2 - 4\frac{u^*(B^TB+\beta A)u}{u^*u}}.$$

Zaključujemo da su rješenja realna ako i samo ako je zadovoljen uvjet teorema. □

Primjetimo da je uvjet u prethodnom teoremu moguće zapisati i na sljedeći način:

$$\frac{u^*(A+\beta I_n)u}{u^*u} \geq 4\frac{u^*(B^TB+\beta A)u}{u^*(A+\beta I_n)u}.$$

Budući da je $u^*(A+\beta I_n)u \geq u^*Au$ dobijemo da vrijedi:

$$\frac{u^*(B^TB+\beta A)u}{u^*(A+\beta I_n)u} \leq \frac{u^*(B^TB+\beta A)u}{u^*Au} = \frac{q^*(A^{-\frac{1}{2}}B^TBA^{-\frac{1}{2}}+\beta I_n)q}{q^*q}$$

pri čemu je $q = A^{\frac{1}{2}}u$ te λ_n i λ_1 minimalna i maksimalna svojstvena vrijednost od $\hat{M}$. Dakle, ukoliko je $\lambda_n(A+\beta I_n) \geq 4\lambda_1(BA^{-1}B^T + I_m)$, uvjet u teoremu 3.4.5 je zadovoljen budući da vrijedi:

$$\begin{aligned}\frac{u^*(A+\beta I_n)u}{u^*u} &\geq \lambda_n(A+\beta I_n)\\ &\geq 4\lambda_1(BA^{-1}B^T+\beta I_m)\\ &\geq 4\frac{q^*(A^{-\frac{1}{2}}B^TBA^{-\frac{1}{2}}+\beta I_n)q}{q^*q}\\ &\geq 4\frac{u^*(B^TB+\beta A)u}{u^*(A+\beta I_n)u}\end{aligned}$$

pri čemu je nejednakost

$$4\lambda_1(BA^{-1}B^T+\beta I_m) \geq 4\frac{q^*(A^{-\frac{1}{2}}B^TBA^{-\frac{1}{2}}+\beta I_n)q}{q^*q} \tag{3.12}$$

zadovoljena zbog sličnosti matrica i Rayleigh-Ritzovog teorema što će biti pokazano u napomeni 3.4.10.

Ovim računom dokazan je sljedeći rezultat.

Korolar 3.4.6. *Neka vrijede identične pretpostavke kao u teoremu 3.4.5. Neka je zadovoljeno $\lambda_n(A+\beta I_n) \geq 4\lambda_1(BA^{-1}B^T+\beta I_m)$. Tada matrica $\hat{M}$ ima samo realne svojstvene vrijednosti.*

Napomena 3.4.7. *Postoji nekoliko ekvivalentnih uvjeta za realnost svojstvenih vrijednosti od $\hat{M}$. Ekvivalentno je zahtijevati $\lambda_n(A) \geq 4\lambda_1(BA^{-1}B^T)+3\beta$. Također, u slučaju $\beta = 0$ moguće je pisati $\lambda_n(A) \geq 4\|S\|_2$.*

Također, napomenimo da je uvjet dovoljan, ali nije nužan. Na primjer promotrimo sljedeću matricu:

$$\hat{M} = \left[\begin{array}{cc|c} \frac{1}{2} & 0 & 0 \\ 0 & 3 & 1 \\ \hline 0 & -1 & 0 \end{array}\right]$$

Laganim računom provjeri se da matrica ima realni spektar, al ne zadovoljava uvjet teorema.

Napomenimo da su uvjeti iz teorema 3.4.5 doista i ispunjeni u praksi i to prilikom rješavanja stacionarne Stokesove zadaće različitim kombinacijama shema konačnih diferencija i konačnih elemenata.

Teorem 3.4.8 (Rayleigh-Ritz)**.** *Neka je M hermitska matrica sa svojstvenim vrijednostima poredanim od najmanje do najveće u poretku $\lambda_n, ..., \lambda_1$. Tada vrijedi:*

$$\lambda_n x^* x \leq x^* M x \leq \lambda_1 x^* x, \forall x \in C^n$$

Nadalje,

$$\lambda_1 = \max_{x \neq 0} \frac{x^* M x}{x^* x} = \max_{x^* x = 1} x^* M x$$
$$\lambda_n = \min_{x \neq 0} \frac{x^* M x}{x^* x} = \min_{x^* x = 1} x^* M x$$

Dokaz. Budući da je M hermitska, postoji unitarna matrica U koja dijagonalizira M:

$$M = U \Lambda U^*, \Lambda = diag(\lambda_n, ..., \lambda_1).$$

Za proizvoljni $x \in C^n$ vrijedi:

$$x^* M x = (x^* U) \Lambda (U^* x) = \sum_{i=1}^{n} \lambda_i |(U^* x)_i|^2.$$

U ovoj linearnoj kombinaciji svojstvenih vrijednosti λ_i, za koeficijente vrijedi $|(U^* x)_i|^2 \geq 0$. Iz poretka $\lambda_n \leq \lambda_i \leq \lambda_1$ slijedi:

$$\lambda_n \sum_{i=1}^{n} |(U^* x)_i|^2 \leq x^* M x \leq \lambda_1 \sum_{i=1}^{n} |(U^* x)_i|^2.$$

Kako je U unitarna, onda vrijedi:

$$\sum_{i=1}^{n} |(U^* x)_i|^2 = (x^* U)(U^* x) = x^* x.$$

Dakle, dokazano je

$$\lambda_n x^* x \leq x^* M x \leq \lambda_1 x^* x$$

□

Definicija 3.4.9. *Za matrice A i B kažemo da su slične ako postoji regularna matrica P takva da vrijedi* $A = P^{-1}BP$.

Napomena 3.4.10. *Vratimo se nejednadžbi (3.12). Kako bismo mogli primijeniti Rayleigh-Ritz teorem potrebno je pokazati sličnost odgovarajućih matrica. Promatrajmo:*

$$A^{\frac{1}{2}}/A^{-\frac{1}{2}}B^TBA^{-\frac{1}{2}}/A^{-\frac{1}{2}}.$$

Nakon odgovarajućeg množenja dobijemo B^TBA^{-1} *što znači da je preostalo pokazati sličnost te matrice sa matricom* $BA^{-1}B^T$. *To ćemo pokazati ako primjetimo da ove matrice imaju jednake svojstvene vrijednosti. Neka je* λ *svojstvena vrijednost za* $BA^{-1}B^T$. *Tada vrijedi:*

$$BA^{-1}B^Ty = \lambda y / B^T \text{ sa lijeva}$$
$$B^TBA^{-1}(B^Ty) = \lambda(B^Ty)$$

što pokazuje da je λ *svojstvena vrijednost i za* B^TBA^{-1} *ako je* $B^Ty \neq 0$ *što je istina jer bi u suprotnom vrijedilo:*

$$BA^{-1}B^Ty = 0 = \lambda y.$$

Dakle, time je pokazana tražena sličnost.

3.5 Uvjetovanost

Sistemi sa sedlastom točkom koji se javljaju u praksi često su loše uvjetovani te je zbog toga potrebno pažljivo konstruirati algoritme za njihovo rješavanje. U nekim slučajevima moguće je iskoristiti posebnu strukturu matrice M kako bi se poboljšala loša uvjetovanost. Nadalje, struktura desne strane b također ima važnu ulogu. Naime, u mnogim slučajevima u praksi je $f = 0$ ili $g = 0$. Tako je na primjer $g = 0$ kod problema inkompresibilnog toka i težinskih najmanjih kvadrata. U tom slučaju blokovi (1,2) i (2,2) matrice M^{-1} ne utječu na konačno rješenje $u = A^{-1}b$. Dakle, ukoliko je loša uvjetovanost sustava proizašla iz tih blokova, to neće utjecati na algoritam za traženje rješenja.

Promatrajmo sada najjednostavniji slučaj kad je A simetrična pozitivno definitna, $B_1 = B_2 = B$ punog ranga i $C = 0$. Sada je M simetrična, a broj uvjetovanosti dan je:

$$\kappa(M) = \frac{max|\lambda(M)|}{min|\lambda(M)|}\ .$$

Koristeći teorem 3.4.2 zaključujemo da broj uvjetovanosti matrice M neograničeno raste kako $\mu_n = \lambda_{min}(A)$ ili $\sigma_m = \sigma_{min}(B)$ padaju ka nuli (pretpostavljamo da su $\lambda_{max}(A)$ i $\sigma_{max}(B)$ konstante). Kad se koristi mješovita formulacija konačnih elemenata za eliptičke parcijalne diferncijalne jednadžbe, tada μ_n i σ_m teže ka nuli kad mrežni parametar h teži ka nuli. Zbog toga broj uvjetovanosti matrice M raste eksponencijalno. Takav rast ima za posljedicu da brzina konvergencije većine iterativnih metoda dramatično pada sa povećanjem veličine problema. Srećom, u mnogim slučajevima upotreba predkondicionera može reducirati ovisnost o h.

Prilikom korištenja metoda unutarnje točke, pridruženi sustavi sa sedlastom točkom imaju lošu uvjetovanost drugačijeg tipa. Ako uzmemo za primjer problem linearnog programiranja gdje je (1,1) blok matrica A dijagonalna, tada će iteracije koje generira metoda, mnoge vrijednosti od A težiti ka nuli ili beskonačnosti što se više približavamo rješenju te će zbog toga A postajati sve lošije uvjetovana. Precizije, norma od inverza Schurovog komplementa $S^{-1} = -(BA^{-1}B^T)^{-1}$ teži ka beskonačnosti. Ovaj problem može se djelomično izbjeći ako se iskoristi da je norma matrice $S^{-1}BA^{-1}$ ograničena neovisno o A.

Poglavlje 4

Numeričke metode

4.1 Redukcija dimenzije sustava pomoću Schurovog komplementa

Jedna od najranijih i najjednostavih ideja kako riješiti sustav sa sedlastom točkom jest rastaviti početni sustav na dva manja sustava koja je onda lakše riješiti. Promatrajmo najopćeniti sustav sa sedlastom točkom oblika (2.1) i zapišimo pripadne jednadžbe:

$$\begin{aligned} Ax + B_1^T y &= f \\ B_2 x - Cy &= g \end{aligned}$$

Kako bismo mogli napraviti redukciju sustava potrebna je pretpostavka da su obje matrice A i M regularne. Tada koristeći rastav (3.2) vrijedi da je $S = -(C+B_2A^{-1}B_1^T)$ također regularna. Pomnoživši prvu jednadžbu sa B_2A^{-1} dobijemo:

$$B_2 x + B_2A^{-1}B_1^T y = B_2A^{-1}f \ ;$$

odakle koristeći $B_2x = g + Cy$ dobijemo sustav za varijablu y:

$$(B_2A^{-1}B_1^T + C)y = B_2A^{-1}f - g \tag{4.1}$$

koji je sada dimenzije m. S obzirom da vrijedi $S = -(B_2A^{-1}B_1^T + C)$ možemo ekvivalento pisati:

$$-Sy = B_2A^{-1}f - g.$$

Neka je y_* izračunato rješenje sustava (4.1). Tada drugi dio rješenja dobijemo iz prve jednadžbe:

$$Ax = f - B_1^T y_* \tag{4.2}$$

gdje sada rješavamo sustav dimenzije n. Primijetimo da je ova metoda u suštini samo primjena Gaussovih eliminacija na blok 2×2 matricu što se može vidjeti primjenivši (3.3) na početni sustav:

$$\begin{bmatrix} I & 0 \\ -B_2A^{-1} & I \end{bmatrix} \begin{bmatrix} A & B_1^T \\ B_2 & -C \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} I & 0 \\ -B_2A^{-1} & I \end{bmatrix} \begin{bmatrix} f \\ g \end{bmatrix},$$

to jest:

$$\begin{bmatrix} A & B_1^T \\ 0 & S \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \begin{bmatrix} f \\ g - B_2A^{-1}f \end{bmatrix}$$

Rješavanje gornjeg sustava supstitucijom unatrag vodi na sustave reducirane dimenzije (4.1) i (4.2) za y i x. Dobivene sustave moguće je riješiti direktno ili iterativnim metodama. Od velikog interesa je specijalan slučaj kad su A i $-S$ simetrične pozitivno definitne. Tada je dane sustave moguće vrlo efikasno riješiti faktorizacijom Cholesky ili metodom konjugiranih gradijenata (CG).

Očita prednost ove metode očituje se u tome da je općenito puno lakše riješiti dva manja sustava nego jedan veći. Ova metoda je vrlo efikasna kad je m relativno mali. Tada je kompliciran sustav za y manje dimenzije. U slučajevima kad je jednostavno za izračunati A^{-1} metoda će raditi dobro.

S druge strane, restrikcija da A mora biti regularna je jako limitirajuća u praksi. Doduše, metodu je moguće upotrijebiti i za takve matrice, ali je tada potrebno raditi transformaciju početnog sustava kao i predkondicioniranje sustava za y koji je često loše uvjetovan. Čak i u slučaju kad je A regularna, može se dogoditi da nije lako izračunati A^{-1} dok je situacija sa S još i gora. Naime, iako početni sustav može imati lijepu strukturu (puno nula), S ne mora biti takva te može biti teška za izračunati i jako loše uvjetovana.

4.2 Metode nul-prostora

Pretpostavljamo da vrijedi $B_1 = B_1 = B$ punog ranga, $C = 0$ i $ker(H) \cap ker(B) = 0$ gdje je H simetričan dio od A. Tada je sustav sa sedlastom točkom koji promatramo oblika:

$$\begin{aligned} Ax + B^T y &= f \\ Bx &= g \ . \end{aligned}$$

Metoda prvo računa partikularno rješenje $\hat{x}$ jednadžbe $Bx = g$. Zatim traži matricu $Z \in R^{n\times(n-m)}$ takvu da je $BZ = 0$. Primjetimo da za takvu Z vrijedi $R(Z) = ker(B)$ to jest, stupci od Z razapinju jezgru od B. Skup rješenja jednadžbe $Bx = g$ dan je linearnom mnogostrukosti $x = Zv + \hat{x}$ gdje je $v \in R^{n-m}$. Uvrstivši $x = Zv + \hat{x}$ u $Ax + B^T y = f$ dobijemo $A(Zv + \hat{x}) = f - B^T y$. Pomnožimo obje strane sa Z^T i iskoristimo $Z^T B^T = 0$ tako da dobijemo reducirani sustav dimenzije $n - m$ za pomoćnu nepoznanicu v:

$$Z^T AZv = Z^T(f - A\hat{x})$$

Primjetimo da je matrica dobivenog sustava regularna. Kad se izračuna v_* tada konačno rješenje (x_*, y_*) slijedi iz jednadžbi:

$$\begin{aligned} x_* &= Zv_* + \hat{x} \\ BB^T y &= B(f - Ax_*) \end{aligned}$$

pri čemu je sustav za y reda m sa simetričnom pozitivno definitnom matricom BB^T.

Metoda ne zahtijeva računanje A^{-1} što može biti iznimno povoljno u određenim slučajevima, a samim time primjenjiva je i na singularne matrice A, doduše uvjet $ker(H) \cap ker(B)$ mora i dalje biti zadovoljen. Umjesto na cijelom sustavu velike dimenzije, metoda generira dva sustava manjih dimenzija. Posebno je povoljna kad je $n - m$ relativno malen broj. Tada je pomoćni sustav po varijabli v brže rješiv. Iako potreba za partikularnim rješenjem sustava $Bx = g$ ne mora biti jednostavna, često se može dobiti kao posljedica računanja Z što dodatno skraćuje posao. Ukoliko je A simetrična pozitivno semidefinitna tada je $Z^T AZ$ simetrična pozitivno definitna i postoji više efikasnih metoda za rješavanje. Općenito se za rješenje sustava koristi metoda konjugiranih gradijenata, ali moguće je primjeniti i druge iterativne rješavače.

Ukoliko je $n - m$ relativno velik, metoda je inferiorna drugim opcijama dok je u slučaju $C \neq 0$ potpuno neprimjenjiva. Ipak, glavni nedostatak je potreba za računanjem baze jezgre od B to jest, matricom Z. S obzirom da bismo željeli da dobivena baza ima što je moguće više dobrih svojstava (dobra uvjetovanost, puno trivijalnih vrijednosti, lagana za računanje,

dijagonalna dominantnost), pronalazak takve matrice nije jednostavan zadatak i poznat je u literaturi kao problem pronalaska dobre baze. Generalno, rješenje takvog problema je NP težine (nepolinomijalan problem) to jest, zasad ne postoji algoritam koji bi računao takvu bazu u polinomijalnom vremenu.

4.3 Iterativne metode

Najkvalitetnije metode za rješavanje sustava sa sedlastom točkom su iterativne metode. Podijenjene su u dvije grupe: metode stacionarnih iteracija i metode iz Krilovljevih potprostora. U današnje vrijeme metode iz obje grupe su u intenzivnoj upotrebi s tom razlikom da se stacionarne iteracije jako rijetko koriste kao samostalne metode, već više kao pomoćno sredstvo ili predkondicioneri za metode iz Krilovljevih potprostora. S obzirom da je velik broj ovih metoda u opticaju, ograničit ćemo se na najvažije iz svake grupe: Uzawa metodu i GMRES metodu (generalizirani algoritam minimalnog reziduala).

Uzawa metoda

Iako je algoritam moguće prilagoditi i na općeniti slučaj, kako bismo izbjegli komplikacije pretpostavimo da je $B_1 = B_2 = B$, $C = 0$ i A invertibilna pošto je to slučaj od najvećeg interesa. Tada su Uzawa iteracije dane s:

$$Ax_{k+1} = f - B^T y_k \tag{4.3a}$$

$$y_{k+1} = y_k + \omega(Bx_{k+1} - g) \tag{4.3b}$$

pri čemu su x_0 i y_0 inicijalne vrijednosti pogođenog rješenja , a ω relaksacijski parametar. Ideja iza ovih iteracija je sljedeća: Početnu matricu sustava M možemo rastaviti u obliku $M = P - Q$. Tada naš sustav izgleda $(P - Q)u = b$ to jest, $Pu = Qu + b$ što vodi na iteracije:

$$P_{u_{k+1}} = Q_{u_k} + b$$

pri čemu je:

$$P = \begin{bmatrix} A & 0 \\ B & -\frac{1}{\omega}I \end{bmatrix}, Q = \begin{bmatrix} 0 & -B^T \\ 0 & -\frac{1}{\omega}I \end{bmatrix} \text{ i } u_k = \begin{bmatrix} x_k \\ y_k \end{bmatrix}$$

Matrica iteracija dana je s:

$$T = P^{-1}Q = \begin{bmatrix} 0 & -A^{-1}B^T \\ 0 & I - \omega BA^{-1}B^T \end{bmatrix}$$

S druge strane ako upotrijebimo jednadžbu (4.3a) da eliminiramo x_{k+1} iz druge jednadžbe dobijemo:

$$y_{k+1} = y_k + \omega(BA^{-1}f - g - BA^{-1}B^T y_k)$$

što pokazuje da je Uzawa metoda jednaka Richardsonovoj metodi primjenjenoj na sustav sa Schurovim komplementom:

$$BA^{-1}B^T y = BA^{-1}f - g$$

Iz ove ekvivalentnosti slijedi ocjena za konvergenciju. Naime, Richardsonove iteracije za sustav $Mu = b$ su $u_{k+1} = u_k + \omega(b - Mx_k)$. Označimo izračunato rješenje s u_* i egzaktno s u. Označimo razliku rješenja sa $e_k = u_k - u$. Tada vrijedi:

$$e_{k+1} = e_k - \omega M e_k = (I - \omega M)e_k \text{ to jest,}$$

$$\|e_{k+1}\| = \|(I - \omega M)e_k\| \leq \|I - \omega M\|\|e_k\|$$

za bilo koju vektorsku i odgovarajuću induciranu matričnu normu. Zaključujemo da metoda konvergira kada je $\|I - \omega M\| < 1$. Neka je (λ_j, v_j) svojstveni par od M. Greška konvergira ka nula ako je $|1 - \omega\lambda_j| < 1$ za svaki λ_j. U slučaju kad je A simetrična pozitivno definitna onda je i $BA^{-1}B^T$ također takva pa su sve svojstvene vrijednosti realne te traženi uvjet možemo zadovoljiti ukoliko odaberemo ω tako da bude zadovoljeno:

$$0 < \omega < \frac{1}{\lambda_{max}(M)} .$$

Optimalan izbor za ω koji minimizira spektralni radijus matrice iteracija dan je s:

$$\omega = \frac{2}{\lambda_{min}(M) + \lambda_{max}(M)} .$$

U određenim specijalnim slučajevima optimalnu vrijednost za ω moguće je odrediti analitički. Primjer takvog slučaja je diskretizacija stacionarne Stokesove zadaće kad je zadovoljen LBB uvjet (Ladiženskaja-Babušksa-Brezzi uvjet). Tada je Schurov komplement spektralno ekvivalentan identiteti. To znači da su svojstvene vrijednosti od $BA^{-1}B^T$ ograničene konstantama koje ne ovise o mrežnom parametru h što za posljedicu ima konvergenciju Uzawa metode neovisno o h. To nažalost nije slučaj u mnogim drugim primjerima gdje je konvergencija metode relativno spora (npr. nestacionarna Stokesova zadaća). Ipak iz ove metode razvile su se mnoge varijacije, a istraživanja se nastavljaju i do današnjih dana.

Iteracije iz Krilovljevih potprostora

Metode koje konstruiraju rješenje iz Krilovljevih potprostora ne rade isključivo na sustavima sa sedlastom točkom već se primjenjuju na najrazličitije probleme što će jasno biti vidljivo iz načina na koji se konstruira rješenje. Pretpostavimo da je u_0 inicijalno pogođeno rješenje sustava (2.1) i definirajmo inicijalni rezidual $r_0 = b - Mu_0$. U k-om koraku algoritma vrijedi:

$$u_k \in u_0 + \mathcal{K}_k(M, r_0), k = 1, 2, ... \tag{4.4}$$

pri čemu je $\mathcal{K}_k(M, r_0) = span\{r_0, Mr_0, ..., M^{k-1}r_0\}k$-ti Krilovljev potprostor generiran M i r_0. Krilovljevi potprostori tvore ugnježđenu strukturu oblika:

$$\mathcal{K}_1(M, r_0) \subset ... \subset \mathcal{K}_d(M, r_0).$$

Za svaki $k \leq d$ Krilovljev potprostor $\mathcal{K}_k(M, r_0)$ je dimenzije k. Zbog k stupnjeva slobode prilikom izbora iteracije u_k potrebno je uvesti k uvjeta kako bi u_k bio jedinstven u svakoj iteraciji. To svojstvo se postiže zahtijevom da k-ti rezidual $r_k = b - Mu_k$ bude ortogonalan na k-dimenzionalan prostor Q_k:

$$r_k = b - Mu_k \in r_0 + M\mathcal{K}_k(M, r_0) \ , \ r_k \perp Q_k$$

gdje se pod pojmom ortogonalnosti smatra u odnosu na euklidski skalarni produkt.

Iz svojstava matrice M moguće je odrediti prostore Q_k koji vode na jedinstvene iteracije $u_k, k = 1, 2, ...$ To je posljedica sljedećeg teorema.

Teorem 4.3.1. *Neka su sa V i W označene baze za $\mathcal{K}_k$ i Q_k. Neka M , $\mathcal{K}_k$ i Q_k zadovoljavaju jedan od sljedeća dva uvjeta:*

- *M je pozitivno definitna i* $Q_k = \mathcal{K}_k(M, r_0)$

- *M je regularna i* $Q_k = M\mathcal{K}_k(M, r_0)$

Tada je matrica $U = W^T MV$ *regularna za bilo koji izbor baza V i W.*

Dokaz. Dokažimo prvu tvrdnju. Budući da su Q_k i $\mathcal{K}_k$ jednaki, moguće je izraziti W kao $W = VG$ pri čemu je G regularna $m \times m$ matrica. Tada vrijedi:

$$U = W^T MV = G^T V^T MV$$

Budući da je M pozitivno definitna slijedi da je i $V^T MV$ također. Dakle, U je regularna čime je pokazana prva tvrdnja.

Dokažimo drugu tvrdnju. Budući da je $Q_k = M\mathcal{K}_k(M, r_0)$, možemo W izraziti u obliku $W = MVG$ pri čemu je G regularna $m \times m$ matrica. Tada vrijedi:

$$U = W^T MV = G^T (MV)^T MV \ .$$

Kako je M regularna, matrica MV dimenzije $n \times m$ je punog ranga pa je $(MV)^T MV$ regularna. Ova tvrdnja zajedno sa G regularna pokazuje da je U regularna. □

GMRES

Metoda formira niz Krilovljevih prostora $\mathcal{K}_k$ tako da Arnoldijevim algoritmom računa pripadne ortonormirane baze Q_k.

Arnoldijev algoritam

Dan je vektor q_1 sa $\|q_1\|_2 = 1$

Za $j = 1, ..., n-1$

$\tilde{q}_j = Mq_j$

Za $i = 1, ..., j$

$h_{i,j} =< \tilde{q}_{j+1}, q_i >$

$\tilde{q}_{j+1} = \tilde{q}_{j+1} - h_{i,j}q_i$

$h_{j+1,j} = \|q_{j+1}\|_2$

$q_{j+1} = \frac{\tilde{q}_{j+1}}{h_{j+1,j}}$

U matričnom obliku Arnoldijev algoritam može se zapisati:

$$MQ_k = Q_k H_k + h_{k+1,k}q_{k+1}e_k = Q_{k+1}H_{k+1,k}$$

Ovdje je Q_k matrica $n \times k$ čiji stupci $q_1, ..., q_k$ čine ortonormuranu bazu, a H_k $k \times k$ gornja Hessenbergova matrica kojoj je (i, j)-ti element jednak $h_{i,j}$ za $j = 1, ...k$, $i = 1, ..., min j + 1, k$, a svi ostali elementi nula. e_k je jedinični vektor $[0, ..., 0, 1]$. Matrica $H_{k+1,k}$ ima na gornjem $k \times k$ bloku matricu H_k , a na zadnjem retku nule osim na mjestu $(k + 1, k)$ gdje se nalazi $h_{k+1,k}$. Rekurzivna definicija za matricu vektora ortonormirane baze $(k + 1)$-dimenzionalnog Krilovljevog potprostora Q_{k+1} može se zapisati:

$$[q_1 MQ_k] = Q_{k+1}R_{k+1}$$

pri čemu je R_{k+1} gornjetrokutasta matrica oblika $R_{k+1} = [e_1 H_{k+1,k}]$, a $e_1 = [1, 0, ..., 0]^T$. Primjetimo da je Arnoldijev algoritam zapravo rekurzivna QR faktorizacija matrice $[q_1 MQ_k]$.

Kako je aproksimacija rješenja oblika $u_k = u_0 + Q_k y_k$ za y_k takav da je $r_k = b - Ax_k = r_0 - MQ_k y_k$ minimalne euklidske norme, vektor y_k je rješenje problema najmanjih kvadrata:

$$\begin{aligned}\min_{y\in\mathbb{C}^k}\|r_0 - MQ_k y\|_2 &= \min_{y\in\mathbb{C}^k}\|r_0 - Q_{k+1}H_{k+1,k}y\|_2\\ &= \min_{y\in\mathbb{C}^k}\|Q_{k+1}(\beta e_1 - H_{k+1,k}y)\|_2\\ &= \min_{y\in\mathbb{C}^k}\|\beta e_1 - H_{k+1,k}y\|_2\end{aligned}$$

pri čemu je $\beta = \|r_0\|_2$ i $q_1 = \frac{r_0}{\|r_0\|}$

Ovaj problem rješavat ćemo faktoriranjem matrice $H_{k+1,k}$ na produkt $(k+1)\times(k+1)$ unitarne matrice $F^{(k)*}$ i $(k+1)\times k$ gornje trokutaste matrice $R^{(k)}$. Gornji $k\times k$ blok matrice $R^{(k)}$ je gornjetrokutast, a zadnji redak jednak je nul-vektoru. Ovo je ponovno *QR* faktorizacija. Ostvarit ćemo ju upotrebom Givensovih rotacija.

Pretpostavimo da imamo *QR* faktorizaciju matrice $H_{k+1,k}$. Tada ćemo *QR* faktorizaciju sljedeće matrice $H_{k+2,k+1}$ izračunati sljedećim koracima. Neka je F_i matrica rotacije koja rotira ravninu razapetu jediničnim vektorima e_i i e_{i+1} za kut θ_i:

$$F_i = \begin{bmatrix} I & & & \\ & c_i & s_i & \\ & -\bar{s}_i & c_i & \\ & & & I \end{bmatrix}$$

gdje je $c_i = cos(\theta_i)$ i $s_i = sin(\theta_i)$. Očito da dimenzija matrice F_i kao i dimenzija drugog bloka ovisi o kontekstu u kojem se koristi. Pretpostavimo da su rotacije $F_i, i = 1, ..., k$ u prethodnom koraku bile upotrijebljene na $H_{k+1,k}$ tako da je:

$$(F_k F_{k-1}...F_1)H_{k+1,k} = \begin{bmatrix} x & x & \dots & x \\ & x & \dots & x \\ & & \ddots & \vdots \\ & & & x \\ 0 & 0 & \dots & 0 \end{bmatrix}$$

gdje smo sa x označili netrivijalne elemente. Tada je $F^{(k)} = F_k F_{k-1}...F_1$. Da bismo dobili matricu $R^{(k+1)}$ koja je gornjetrokutasti faktor od $H_{k+2,k+1}$ potrebno je izmnožiti zadnji stupac od $H_{k+2,k+1}$ sa prethodnim rotacijama. Ako smo izračunali *QR* faktorizaciju za takvu matricu u k-tom koraku onda u $(k+1)$-om koraku moramo izračunati *QR* faktorizaciju matrice koja je jednaka prethodnoj matrici osim što ima jedan dodatan stupac. Trokutasti faktor matrice u $(k+1)$-om koraku je jednak trokutastom faktoru matrice iz k-og koraka kojemu je također

dodan još jedan stupac. Slijedi da je dovoljno obraditi samo novi $(k+1)$-i stupac matrice $H_{k+2,k+1}$. Dakle, dobivamo:

$$(F_k F_{k-1} ... F_1) H_{k+2,k+1} = \begin{bmatrix} x & x & \dots & x & x \\ & x & \dots & x & x \\ & & \ddots & \vdots & \vdots \\ & & & x & x \\ 0 & 0 & \dots & 0 & d \\ 0 & 0 & \dots & 0 & h \end{bmatrix}$$

gdje je $(k+2, k+1)$-i element h upravo $h_{k+2,k+1}$ jer na njega ne utječu prethodne rotacije. Sljedećom rotacijom F_{k+1} eliminiramo upravo taj element zahtijevajući da vrijedi $-\bar{s}_{k+1} d + c_{k+1} h = 0$. Ovaj uvjet bit će zadovoljen ako stavimo da vrijedi:

$$c_{k+1} = \frac{|d|}{\sqrt{|d|^2+|h|^2}} \, , \, \bar{s}_{k+1} = \frac{c_{k+1} h}{d} \text{ za } d \neq 0 \text{ i } h \neq 0$$
$$c_{k+1} = 0 \, , \, s_{k+1} = 1 \text{ za } d = 0$$
$$c_{k+1} = 1 \, , \, s_{k+1} = 0 \text{ za } h = 0$$

Primjetimo da ako je $h = 0$ egzaktno rješenje početnog sustava postignuto u $(k+1)$-om koraku. Tada je $F_{k+1} = I$ pa su prvih $k+1$ komponenti vektora $\beta F^{(k+1)} e_1$ dimenzije $k+2$ jednake vektoru $\beta F^{(k)} e_1$ iz prethodnog koraka, a zadnja komponenta je jednaka nuli. Budući da je apsolutna vrijednost te zadnje komponente jedanka euklidskoj normi reziduala u tom koraku, zaključujemo da je $r_{k+1} = 0$ to jest, postignuto je rješenje. S druge strane ako egzaktno rješenje nije postignuto, tada je $h \neq 0$ pa je $(k+1)$-i dijagonalni element od $R^{(k+1)}$ različit od nule. Za $d = 0$ taj element je jednak točno h , a za $d \neq 0$ jednak je $c_{k+1} d + s_{k+1} h = \frac{d}{|d|} \sqrt{|d|^2 + |h|^2}$.

Napomenimo da potpuno izvršavanje GMRES algoritma može biti nepraktično zbog velikog zahtjeva za memorijom. Zbog toga se uvodi dodatan parametar *restart* koji restarta GMRES algoritam kad dosegne zadani broj iteracija koristeći zadnju iteraciju prethodnog algoritma kao prvu iteraciju novog. Nažalost, ovo je veliki problem GMRES metode jer se može dogoditi da prilikom novog pokretanja algoritma pretražujemo smjerove koje smo pretražili u prethodnom ciklusu pa dolazi do stagnacije algoritma.

Očito da konvergencija GMRES metode ovisi o ponašanju reziduala. U k-tom koraku vrijedi $u_k = u_0 + z_k$ gdje je:

$$z_k = \sum_{j=0}^{k-1} \zeta_j M^j r_0 \in \mathcal{K}_k$$

korekcija i pripadni rezidual $r_k = b - Mu_k = r_0 - Mz_k$. Primjetimo da možemo pisati $Mz_k = q(M)r_0$ gdje je $q(t) = \sum_{j=1}^{k} \zeta_{j-1} t^j \in \mathcal{P}_k$ pri čemu je $\mathcal{P}_k$ prostor polinoma stupnja najviše k.

Propozicija 4.3.2. *U k-tom koraku GMRES metode vrijedi:*

$$\|r_k\|_2 = \min_{p \in \mathcal{P}_k, p(0)=1} \|p(M)r_0\|_2$$

Dokaz. Stavimo $p(t) = 1 - q(x)$. Odmah slijedi $p(0) = 1$, $r_k = p(M)r_0$ i jasno je da variranjem korekcije z_k po $\mathcal{K}_k$ svi mogući reziduali su oblika $p(M)r_0$, $p \in \mathcal{P}_k$, $p(0) = 1$. Obratno, svakom takvom polinomu odgovara korekcija koja reproducira $r_k = p(M)r_0$. □

Dakle, redukcija reziduala ovisi o ponašanju određene klase polinoma u varijabli M. Ako pretpostavimo da je M dijagonalizabilna onda možemo reći sljedeće:

$M = S\Lambda S^{-1}$, $\Lambda = diag(\lambda_i)$, $i = 1, ..., n$. Tada vrijedi:

$$p(M) = S\,p(\Lambda)S^{-1} = S \begin{bmatrix} p(\lambda_1) & & \\ & \ddots & \\ & & p(\lambda_n) \end{bmatrix} S^{-1}$$

Slijedi:

$$\begin{aligned} \|p(\Lambda)\|_2 &= \max_{i=1:n} |p(\lambda_i)| \\ \|p(M)\|_2 &\leq \|S\|_2 \|S^{-1}\|_2 \|p(\Lambda)\|_2 = \kappa_2(S)\|p(\Lambda)\|_2 \\ \|p(M)r_0\|_2 &\leq \kappa_2(S)\|p(\Lambda)\|_2 \|r_0\|_2 \end{aligned}$$

Sada zaključujemo da vrijedi:

$$\|r_k\|_2 = \min_{p \in \mathcal{P}_k, p(0)=1} \|p(M)r_0\|_2 \leq \kappa_2(S) \min_{p \in \mathcal{P}_k, p(0)=1} \|p(\Lambda)\|_2$$

što daje:

$$\frac{\|r_k\|_2}{\|r_0\|_2} \leq \kappa_2(S) \min_{p \in \mathcal{P}_k, p(0)=1} \max_{i=1:n} |p(\lambda_i)| \leq \kappa_2(S) \min_{p \in \mathcal{P}_k, p(0)=1} \max_{z \in \mathcal{D}} |p(z)|$$

gdje je $D \subset \mathbb{C}$ skup koji sadrži sve svojstvene vrijednosti od M. Primjetimo da ako je M normalna onda je S unitarna i $\kappa_2(S) = 1$. Vidimo da brzina redukcije reziduala ovisi o ponašanju određenih polinoma na spektru matrice M. Iz ove ocjene možemo intuitivno zaključiti da će metoda konvergirati brzo kad su svojstvene vrijednosti od M daleko od ishodišta i kad je M blizu normalne matrice.

4.4 Numerički primjeri

Prilikom provjeravanja algoritama egzaktno rješenje smo postavili na $[1, 1, ..., 1]^T$, a desnu stranu smo generirali u obliku $b = M * [1, 1, ..., 1]^T$. Kao početno pogođeno rješenje koristili smo $[0, 0, .., 0]^T$. Svi primjeri rađeni su u Matlabu 7.12.0(R2011a).

Kako bi generirali sustave oblika (2.2) i (2.4) poslužili smo se softverom IFISS koji je otvorenog tipa i dostupan ovdje: http://www.maths.manchester.ac.uk/ djs/ifiss/. Tim paketom moguće je generirati i rješavati različite oblike parcijalnih diferencijalnih jednadžbi. Mi smo generirali dvije stacionarne Stokesove zadaće (leaky lid-driven cavity). Kao što je već ranije spomenuto matrice dobivenih sustava su izrazito loše uvjetovane, gdje je tipično $\kappa(M) \geq 10^{10}$. Zbog toga je potrebno vršiti prekondicioniranje. Generalno, ideja je pronaći matricu kojom ćemo pomnožiti početni sustav kako bismo dobili novi sustav sa boljim svojstvima.

Primjer 1

U ovom primjeru matrica sustava je oblika (2.2) s razlikom da je A pozitivno semidefinitna, a promatrana metoda GMRES. Uvjetovanost matrice sustava je 10^{16}. Za matricu prekondicioniranja uzmemo

$$P = \begin{bmatrix} G & B^T \\ B & 0 \end{bmatrix} \tag{4.5}$$

gdje je G dijagonalna matrica za koju vrijedi $g_{i,i} = a_{i,i}$, $i = 1, ..., n$. Ovakav izbor za G moguć je samo ukoliko su sve dijagonalne vrijednosti od A pozitivne što je u ovom slučaju zadovoljeno. Za novu matricu sustava dobijemo

$$\hat{M} = \begin{bmatrix} G & B^T \\ B & 0 \end{bmatrix}^{-1} \begin{bmatrix} A & B^T \\ B & 0 \end{bmatrix} = \begin{bmatrix} (I - T)G^{-1}A + T & 0 \\ X & I \end{bmatrix} \tag{4.6}$$

gdje je $T = G^{-1}B^T(BG^{-1}B^T)^{-1}B$ i $X = (BG^{-1}B^T)^{-1}B(G^{-1}A - I)$.

Očito pitanje je koliko brzo se može izračunati P^{-1}? S obzirom da je u našem slučaju G dijagonalna i matrica $BG^{-1}B^T$ dobrih numerički svojstava, inverz od P moguće je odrediti efikasno. Napomenimo da u specijalnom slučaju kad je $n = m$ vrijedi $T = I$ te:

$$\hat{M} - I = \begin{bmatrix} 0 & 0 \\ X & 0 \end{bmatrix} \tag{4.7}$$

odakle zaključujemo da je $(\hat{M} - I)^2 = 0$. Dakle, minimalni polinom prekondicionirane matrice je stupnja 2 što znači da će GMRES pronaći rješenje u najviše 2 koraka neovisno o A i G.

Postavivši toleranciju na 10^{-12} te primjenivši ovaj algoritam na naš primjer, dobili smo konvergenciju GMRES metode u dva koraka sa postignutim rezidualom reda veličine 10^{-15} i normom greške izračunatog i postignutog rješenja reda veličine 10^{-9}.

Primjer 2

U ovom primjeru matrica sustava je oblika (2.2), a promatrana metoda Uzawa sa konjugiranim gradijentima. Također, u ovom primjeru postavit ćemo toleranciju na 10^{-8} ,a za optimalan ω uzeta je vrijednost 0.5. Dobijemo konvergenciju metode nakon 629 koraka sa rezidualom reda veličine 10^{-9} i normom greške izračunatog i egzaktnog rješenja reda veličine 10^{-2}.

Primjer 3

U ovom primjeru matrica sustava je oblika (2.4) pri čemu je A simetrična pozitivno definitna i C simetrična pozitivno semidefinitna, a promatrana metoda GMRES. Prekondicioniranje ćemo izvršiti slično kao u primjeru 1. Za matricu prekondicioniranja uzmemo:

$$P = \begin{bmatrix} A & B^T \\ B & -\epsilon I \end{bmatrix} \tag{4.8}$$

gdje je $\epsilon \geq 0$. U našem slučaju stavit ćemo $\epsilon = 5$. Nova matrica sustava onda je oblika:

$$\hat{M} = \begin{bmatrix} I & \hat{T} \\ 0 & (I + S^{-1}(C - \epsilon I))^{-1} \end{bmatrix} \tag{4.9}$$

gdje je $\hat{T}$ neka matrica različita od nul-matrice i $S = -BA^{-1}B^T$. Slično kao u primjeru 1 zbog dobrih svojstava inverz od P moguće je efikasno odrediti. Detalji u [1]. Postavimo toleranciju na 10^{-14} i novi sustav riješimo metodom GMRES te dobijemo rješenje u petom koraku

pri čemu je rezidual reda veličine 10^{-18}, a norma greške izračunatog i egzaktnog rješenja reda veličine 10^{-10}.

Primjer 4

U ovom primjeru matrica sustava je ck104 općenitog oblika (2.1) preuzeta s http://yifanhu.net. U ovom slučaju postavit ćemo traženu toleranciju na 10^{-16}. Sustav ćemo riješiti metodom GMRES i dobijemo rješenje u 6 koraka s rezidualom reda veličine 10^{-16} te normom greške izračunatog i egzaktnog rješenja reda veličine 10^{-10}.

Bibliografija

[1] Owe Axelsson, Maya Neytcheva *Preconditioning methods for linear systems arising in constrained optimization problems*, Numerical linear algebra with applications, 2003.

[2] Michele Benzi, Gene H. Golub, Jorg Liesen *Numerical solution of saddle point problems*, Cambridge University Press, Velika Britanija, 2005.

[3] Michele Benzi, Valeria Simoncini *On the eigenvalues of a class of saddlepoint matrices*, Springer-Verlag, 2005.

[4] Nela Bosner *Iterativne metode za rješavanje linearnih sustava*, Matematički odjel Prirodoslovno-matematičkog fakulteta, Zagreb, 2001.

[5] Zlatko Drmač *Numerička matematika*, dostupno na: https://web.math.pmf.unizg.hr/ drmac/na001.pdf, (studeni 2016.)

[6] Wilfried N. Gansterer, Josef Scheid, Christoph W. Ueberhuber *Mathematical properties of equilibrium systems*, Aurora 13, 2003.

[7] C.T. Kelley *Iterative methods for linear and nonlinear equations*, Society for industrial and applied mathematics, Philadelphia, 1995.

[8] Torgeir Rusten, Ragnar Winther *A preconditioned iterative method for saddlepoint problems*, SIAM Journal on Matrix Analysis and Applications 13, 1992., 887-904

[9] Yousef Saad *Iterative methods for sparse linear systems*, Society for industrial and applied mathematics, 2003.

Sažetak

U ovom radu opisani su neki problemi koji se svode na sustave sa sedlastom točkom među kojima su najvažniji problemi inkompresibilnog toka i problemi iz područja optimizacije. Da bi se mogli konstruirati efikasni algoritmi za rješavanje, bilo je potrebno iznjeti teorijsku podlogu matrica koji se javljaju prilikom diskretizacija promatranih problema. Iz teorijskih razmatranja zaključili smo da promatrane matrice imaju općenito jako loša spektralna svojstva zbog svoje blizine singularnim matricama. Ipak, koristeći se prije svega Schurovim komplementom, neke zapreke je moguće zaobići i u kombinaciji sa prekondiciniranjem konstruirati efikasne algoritme za rješevanje takvih sustava. Uzawa metoda je povijesno prva metoda koja je efikasno rješavala određenu klasu ovih problema dok su se kasnije pojavile modernije metode u potpunosti zasnovane na iteracijama iz Krilovljevih potprostora čiji je najznačajniji predstavnik GMRES metoda. Međutim, zbog loše uvjetovanosti matrica gotovo je uvijek potrebno provoditi prekondicioniranje. Teorija odabira matrice prekondicioniranja je iznimno složena sa mnogo otvorenih pitanja. Zbog toga smo mi obradili numerički jednostavnije primjere.

Summary

In this paper we described some of the problems that lead to saddle point systems with special emphasis on incompressible flow problems and optimization problems. In order to construct efficient solution algorithms, it was necessary to understand the properties of matrices that occur as a result of discretization of a continuous problems. The most important conclusion of this theory is that matrices from saddle point systems have very poor spectral properties due to their vicinity to singular matrices. Nevertheless, using Schur complement decomposition and other techniques, it is possible to overcome certain obstacles and in combination with preconditioning construct a viable solution algorithms for this ill-conditioned systems. The first effective method developed for solving a certain class of saddle point systems is Uzawa method. Later, many methods were developed that are primary based on Krylov subspace iterations with the most prominent of them GMRES method. Unfortunately, due to ill-conditioned nature of saddle point systems, some sort of preconditioning is almost always required. Since the theory of determining a precondition matrix is very complex with many unanswered questions, the numerical examples in this paper are a bit simpler.

Životopis

Rođen sam 31.03.1987. u Zagrebu. Iako živim u Samoboru, većinu svojeg obrazovanja sam stekao u Zagrebu pa sam tako završio I. tehničku školu Nikola Tesla 2005. Nakon kraće epizode na Fakultetu strojarstva i brodogradnje i rada u tvrtci Data telecom upisao sam 2009. Prirodoslovno-matematički fakultet u Zagrebu. Završio sam preddiplomski sveučilišni studij matematike 2013. te upisao diplomski studij primijenjene matematike koji trenutno završavam. Za vrijeme studiranja radio sam na mnogim poslovima preko student servisa.