SVEUČILIŠTE U ZAGREBU

PRIRODOSLOVNO–MATEMATIČKI FAKULTET

MATEMATIČKI ODSJEK

Maša Avakumović

KVADRATIČNI PROBLEM SVOJSTVENIH VRIJEDNOSTI

Diplomski rad

Voditelj rada:
Prof. dr. sc. Zlatko Drmač

Zagreb, rujan 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

Sadržaj **iii**

Uvod **2**

1 Primjena kvadratičnog problema svojstvenih vrijednosti **3**
1.1 Mehaničke oscilacije 4
1.2 Akustičke vibracije 6

2 Spektralna teorija **10**
2.1 Standardni problem svojstvenih vrijednosti 11
2.2 Kanonski oblici 12
2.3 Generalizirani problem svojstvenih vrijednosti 13
2.4 Polinomijalni problem svojstvenih vrijednosti 16

3 Perturbacijska analiza **20**
3.1 Greška unatrag 21
3.2 Osjetljivost 22

4 Potpuna metoda **24**
4.1 Prednosti 24
4.2 Linearizacija 25
4.3 Skaliranje 25
4.4 Deflacija 28
4.5 Svojstveni vektori 29
4.6 Algoritam 30
4.7 Eksperimenti 31

5 Iterativne metode **36**
5.1 Arnoldijeva metoda 36
5.2 SOAR 46

6 Zaključak **58**

Bibliografija **60**

Uvod

U ovom radu se bavimo kvadratičnim problemom svojstvenih vrijednosti ili kraće QEP *(eng. Quadratic Eigenvalue Problem)*, s kojim se susreću mnogi inžinjeri prilikom analize mehaničkih oscilacija, akustičkih vibracija, u mehanici fluida i mnogim drugim područjima. Primjena QEP-a u analizi mehaničkih oscilacija i akustičkih vibracija kratko je opisana u u poglavlju 1. Po definiciji, kvadratični problem svojstvenih vrijednosti predstavlja problem u kojem se traže skalari $\lambda \in \mathbb{C}$ (svojstvene vrijednosti) i nenul vektori $x \in \mathbb{C}^n$ (svojstveni vektori) tako da vrijedi

$$(M\lambda^2 + C\lambda + K)x = 0,$$

gdje su $M, C, K \in \mathbb{C}^{n\times n}$. Jedan od načina na koji se mogu podijeliti numeričke metode koje riješavaju ovaj problem je na potpune (direktne) i iterativne. Potpune metode računaju cijeli spektar, dakle sve svojstvene vrijednosti, dok iterativne metode koristimo za računanje samo dijela spektra koji je od interesa (npr. ako se promatra stabilnost sustava, od interesa će biti one svojstvene vrijednosti u lijevoj poluravnini i blizu imaginarne osi). Potpuna metoda koja će u ovom radu biti detaljnije obrađena je metoda *quadeig* (poglavlje 4), a što se tiče iterativnih metoda (poglavlje 5), glavnu ulogu će igrati Arnoldijeva metoda tj. SOAR *(eng. Second Order ARnoldi)* metoda koja je poopćenje Arnoldijeve metode za standardni svojstveni problem. S obzirom da je ljudski mozak naučen razmišljati linearno, a shodno tome su rađeni i numerički solveri; svaka numerička metoda koja pokušava riješiti QEP prvo mora kvadratični problem svojstvenih vrijednosti transformirati u njemu ekvivalentnu linearnu formu. Tu transformaciju kvadratičnog problema svojstvenih vrijednosti u njemu ekvivalentni generalizirani problem svojstvenih vrijednosti nazivamo *linearizacija*. Linearizacija ima beskonačno mnogo, a neke od najpopularnijih su navedene u poglavlju 2. Budući da se kod linearizacije kombiniraju matrice koje su mjerene u različitim fizikalnim veličinama što može dovesti do velikih razlika u normama a samim time i do loših aproksimacija svojstvenih vrijednosti, metoda *quadeig* nastoji izbjeći taj problem skaliranjem originalnih matrica. Osim toga, metoda *quadeig* na efikasan način računa svojstvene vrijednosti jednake beskonačnosti ili nuli, te pripadajuće svojstvene vektore. Točnost aproksimacija svojstvenih vrijednosti metode *quadeig* mjerimo analizom greške unatrag i koeficijentom osjetljivosti, koji su detaljnije opisani u poglavlju 3. S druge strane, kada su

originalne matrice M,C i K iznimno velikih dimenzija (npr. $n > 10^5$), potpunu metodu će u većini slučaja biti nemoguće primijeniti, a osim toga, često će nas zanimati samo par svojstvenih vrijednosti pa se u poglavlju 5 bavimo iterativnim metodama kojima je osnovna ideja projicirati početni problem u potprostor manjih dimenzija koji sadrži informaciju o željenom dijelu spektra. Naravno, ideja je da taj potprostor dobro aproksimira prostor u kojem se nalaze traženi svojstveni vektori. Osim toga, posebno ćemo obratiti pozornost na strukture matrica M, C i K koje ne samo da uvjetuju sam izgled spektra, već i numeričku metodu koja će se korisiti za računanje aproksimacija. Nakon svake obrađene metode prezentirat ćemo numeričke eksperimente u MATLAB-u koji će pokazati koliko rezultati određene metode zapravo dobro aproksimiraju tražene svojstvene vrijednosti.

Poglavlje 1

Primjena kvadratičnog problema svojstvenih vrijednosti

Kvadratični problem svojstvenih vrijednosti QEP (*eng. Quadratic Eigenvalue Problem*) je potraga za skalarom $\lambda \in \mathbb{C}$ i nenul vektorom $x \in \mathbb{C}^n$ tako da vrijedi

$$(M\lambda^2 + C\lambda + K)x = 0, \tag{1.1}$$

gdje su $M, C, K \in \mathbb{C}^{n\times n}$, te predstavljaju različita fizikalna svojstva (M najčešće predstavlja masu, K krutost, a C koeficijent prigušivanja). Najčešća pojava ovog tipa problema je kroz analize dinamičkih mehaničkih i akustičkih sustava, te mehanike fluida. Kako i na koji način se formulira kvadratični problem će biti detaljnije objašnjeno u nastavku. Da naglasimo važnost numeričke analize kvadratičnih problema svojstvenih vrijednosti prisjetimo se jednog od najpoznatijih primjera pogrešne analize kada je mogućnost isforsirane rezonance bila zanemarena te su krivi izračuni doveli do velike katastrofe kada se 7.11.1940. Tacoma most urušio samo zato jer je zapuhao malo jači vjetar. Vjetar koji je zapuhao tog nesretnog dana je imao jednaku frekvenciju kao i prirodna vibracija mosta što je dovelo do rezonance. Na isti način kako roditelj gura dijete na ljuljački odgovarajućom frekvencijom te namjerno izaziva rezonancu da bi ljuljačka išla sve više i više, tako je vjetar slučajno ”gurao” Tacoma most te izazvao rezonancu što je na kraju završilo s katastrofalnim posljedicama. Ovisno o problemu s kojim radimo, rezonanca može biti poželjan efekt (namještanje radio stanice) ili nepoželjan (Tacoma most). Još jedan zanimljiv most je Millenium most u centru Londona, koji se na dan otvorenja u srpnju 2000. godine počeo uvijati na neprirodan način. Razlog takvom ponašanju mosta je opet fenomen rezonanca. Tog dana, velika količina pješaka je svojim tempom hodanja slučajno uskladila korake prirodnim vibracijama mosta što je izazvalo zabrinjavajuće njihanje mosta. Millenium most je bilo potrebno odmah zatvoriti za daljnje korištenje sve dok se nisu dodali prigušivaći, koji danas svojim prigušivanjem vibracija izazvanih raznim faktorima (hodanje pješaka, puha-

nje vjetra,..) osiguravaju jedan stabilan most, koji se ne njiše. Millenium most je ponovo otvoren 2002. godine i dan danas je u funkciji. Sa kvadratičnim problemom svojstvenih vrijednosti se također susrećemo u analizi vibracija u automobilskoj ili zrakoplovnoj industriji. Kao što je zahtijevala i analiza prije gradnje Tacoma mosta i Millenium mosta, u automobilskoj industriji se kvadratični problem svojstvenih vrijednosti riješava kako bi se smanjile vibracije, a samim time i njihov neželjen efekt (npr. buka). Takva vrsta problema spada u podklasu akustičnih vibracija.

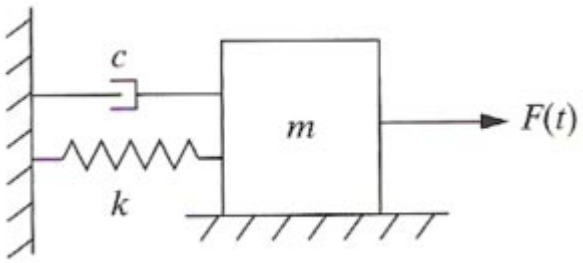
1.1 Mehaničke oscilacije

Kao prvi primjer pogledajmo klasu najpoznatijih jednadžbi koja je rezultat matematičkog modeliranja mnogih inžinjerskih problema, kao što su gradnja mostova, nebodera itd. Prigušeno titranje matematički formuliramo na sljedeći način:

$$M\ddot{q}(t) + C\dot{q}(t) + Kq(t) = F(t), \tag{1.2}$$

gdje su M,C,$K \in \mathbb{C}^{n\times n}$, a $q(t)$, $F(t) \in \mathbb{C}^n$. $q(t)$ opisuje gibanje, $F(t)$ je vanjska sila koja djeluje na sustav, M reprezentira masu, C svojstva prigušivača, a K krutost. U slučaju neprigušenog (ili slobodnog) titranja srednji član C se izostostavlja i problem znatno pojednostavljuje. U daljnjim razmatranjim pretpostavljamo da je titranje prigušeno tj. C nije nul matrica.

Krenimo s najjednostavnijim primjerom sa samo jednim stupnjem slobode. Ta situacija je prikazana na slici 1.1



Slika 1.1: Prigušene mehaničke oscilacije

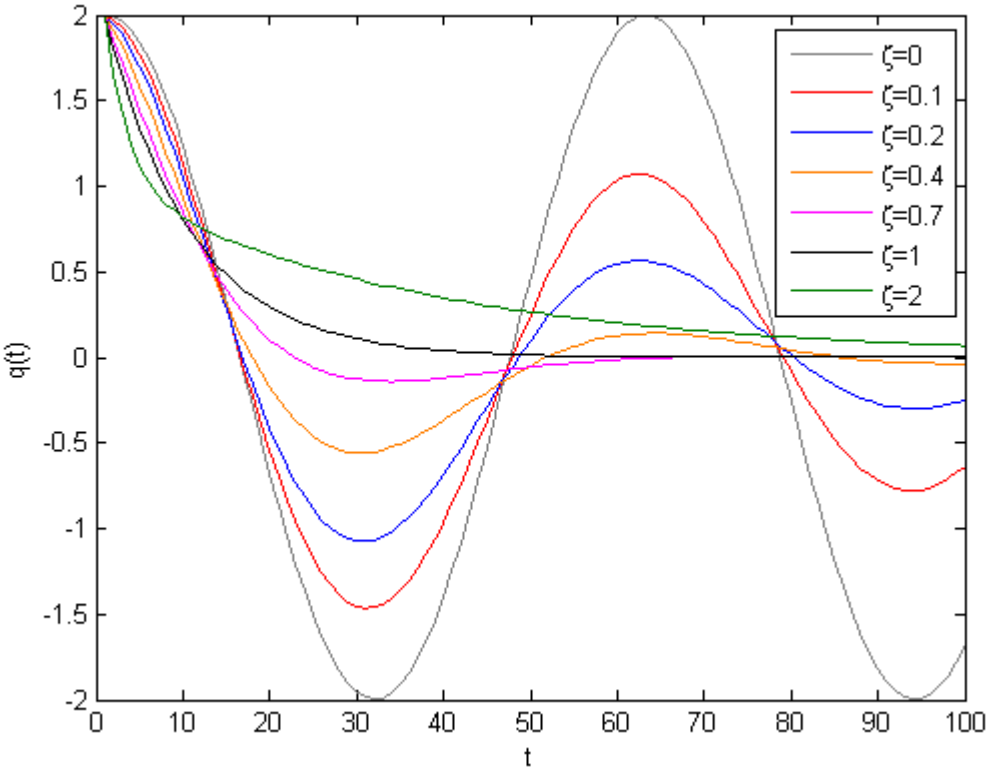
Zakon gibanja $q(t)$, rješenje je diferencijalne jednadžbe (1.2) kojom je opisano gibanje mase m. U slučaju da se radi o homogenoj jednadžbi, tj. nikakva sila ne djeluje na masu ($F(t) = 0$, $t \geq 0$). U tom slučaju je sustav sa slike 1.1 je opisan sljedećom matematičkom formulacijom:

$$\ddot{q}(t) + 2\zeta\omega\dot{q} + \omega^2 q(t) = 0, \quad \omega = \sqrt{\frac{K}{M}}, \quad \zeta = \frac{C}{2M\omega}, \tag{1.3}$$

gdje ω označava prirodnu frekvenciju, a ζ faktor viskoznog prigušivača.
Riješenje ovog sustava je dano sa

$$q(t) = \alpha_1 e^{\lambda_1 t} + \alpha_2 e^{\lambda_2 t}, \tag{1.4}$$

gdje su $\lambda_{1,2} = (-\zeta \pm \sqrt{\zeta^2 - 1})$ izračunate nultočke polinoma $\lambda^2 + 2\zeta\omega\lambda + \omega^2 = 0$, a α_1 i α_2 su određene ako su zadani inicijalni uvjeti $q(t_0)$ i $\dot{q}(t_0)$. To rješenje $q(t)$ je zapravo val, a ovisno o jačini prigušivača amplituda tog vala varira. Kao što i samo ime kaže, svrha prigušivača je da "guši" titranje tj. da ga smiruje s vremenom. U terminima već spomenutog vala, to gušenje znači da se amplituda polako, proporcionalno jačini prigušivača, približiva nuli, što se može vidjeti na slici 1.2.



Slika 1.2: Razlika u prigušenosti

Za složeniji sustav s n stupnjeva slobode (n masa, n opruga i n prigušivača), rješenje se dobiva analogno i zapisujemo ga u obliku:

$$q(t) = \sum_{k=1}^{2n} \alpha_k x_k e^{\lambda_k t} = X e^{\Lambda t} a, \tag{1.5}$$

gdje je $X = [x_1, ..., x_{2n}]$, $\Lambda = diag(\lambda_1, ..., \lambda_{2n})$ i $a = [a_1, .., a_{2n}]^T$ vektor proizvoljnih konstanti. Radi jednostavnosti smo pretpostavili da su sve svojstvene vrijednosti različite. Primjetimo da kada postoji svojstvena vrijednost λ_j takva da joj je realni dio strogo pozitivan ($Re(\lambda_j) > 0$) tada rješenje $\|q(t)\|$ raste prema beskonačnosti kada $t \to \infty$ što ukazuje na nestabilnost samog sustava. Slučaj kada $Re(\lambda_k) < 0, k = 1, ..2n$ je karakteristika stabilnog sustava u kojem rješenje $\|q(t)\| \to 0$ kada $t \to \infty$. Spomenimo još slučaj kada vrijedi $Re(\lambda_k) \leq 0, k = 1, ..2n$, tada kažemo da je sustav slabo stabilan, u smislu da je rješenje $\|q(t)\|$ ograničeno kada $t \to \infty$. Pretpostavimo li sada da na sustav djeluje harmonička sila $F(t) = f_0 e^{i\omega_0 t}$ gdje ω_0 predstavlja frekvenciju (u primjeru Tacoma mosta $F(t)$ predstavlja silu vjetra, a ω_0 njegovu frekvenciju). Tada je partikularno rješenje je dano

$$q_p(t) = e^{i\omega_0 t} \sum_{j=1}^{2n} \frac{y_j^* f_0}{i\omega_0 - \lambda_j} x_j. \tag{1.6}$$

Kada se $i\omega_0$ približava svojstvenoj vrijednosti λ_j, tada j-ti koeficijent $\frac{y_j^* f_0}{i\omega_0 - \lambda_j}$ iz rastava (1.6) teži k beskonačnosti. Na taj način matematički interpretiramo fenomen rezonance. U slučaju Tacoma mosta, frekvencija vjetra ω_0 se previše približila svojstvenoj vrijednosti sustava (tj. prirodnoj frekvenciji mosta) što je dovelo do rezonance koja je prouzrokovala rušenje mosta.

1.2 Akustičke vibracije

Ova klasa problema se sve češće javlja kroz analize svakodnevnih akustičkih pojava. Kao što je već spomenuto, rješavanjem kvadratičnog problema svojstvenih vrijednosti može se smanjiti buka (u avionima, vlakovima...) ili pojačati (npr. pri izradi instrumenata). Kao primjer ćemo detaljnije proučiti širenje zvuka u sobi gdje je jedan zid napravljen od materijala koji apsorbira zvuk, a drugi ga reflektira (detalji se mogu naći u [2]). Jednadžba koja opisuje takvo gibanje akustičnog vala (fluida) u $\Omega \subset \mathbb{R}^2$ je dana sa

$$\rho \frac{\partial^2 U}{\partial t^2} + \nabla P = 0, \tag{1.7}$$

gdje P predstavlja akustički tlak, U je pomak fluida u ovisnosti o prostoru x i vremenu t i ρ je gustoća fluida.

Rubna zadaća je formulirana na sljedeći način

$$\begin{cases} \rho \dfrac{\partial^2 \mathbf{U}}{\partial t^2} + \nabla P = \mathbf{0} & \text{u } \Omega \\ P = -\rho c^2 \operatorname{div} \mathbf{U} & \text{u } \Omega \\ \mathbf{U} \cdot \nu = 0 & \text{na } \Gamma_R \\ \alpha \mathbf{U} \cdot \nu + \beta \dfrac{\partial \mathbf{U}}{\partial t} \cdot \nu = P & \text{na } \Gamma_A, \end{cases} \tag{1.8}$$

gdje je c brzina zvuka u zraku, ν normala na granicu, α, β su koeficijenti akustičke impedancije[1], Γ_R predstavlja granicu tj. zid koji reflektira zvuk, a Γ_A predstavlja zid koji apsorbira zvuk. Prigušene vibracije fluida su dane kao kompleksna rješenja sustava (1.8) u obliku $P(x,t) = p(x)e^{\lambda t}$ i $U(x,t) = u(x)e^{\lambda t}$. Ta rješenja su definirana jednom kad nađemo $\lambda \in \mathbb{C}$, $\mathbf{u} : \Omega \to \mathbb{C}^n$ i $p : \Omega \to \mathbb{C}$, $(\mathbf{u}, p) \neq (\mathbf{0}, 0)$ tako da vrijedi:

$$\begin{cases} \rho\lambda^2 \mathbf{u} + \nabla p = \mathbf{0}, & \text{u } \Omega \\ p = -\rho c^2 \operatorname{div} \mathbf{u} & \text{u } \Omega \\ \mathbf{u} \cdot \nu = 0 & \text{na } \Gamma_R \\ p = (\alpha + \lambda\beta)\, \mathbf{u} \cdot \nu & \text{na } \Gamma_A. \end{cases} \tag{1.9}$$

Za rješavanje ovakvog sustava potrebno je definirati varijacijsku formulaciju, a za to prvo treba uvesti pojam prostora test-funkcija:

$$\mathcal{V} = \left\{ \phi \in H(\operatorname{div}, \Omega) : \phi \cdot \nu \in L^2(\partial\Omega) \quad i \quad \phi \cdot \nu = 0 \quad na \quad \Gamma_R \right\}, \tag{1.10}$$

gdje su prostori funkcija $L^2(\partial\Omega)$ i $H(\operatorname{div}, \Omega)$ definirani na sljedeći način:

$$L^2(\partial\Omega) = \left\{ f : \partial\Omega \to \mathbb{C}; f \text{ je izmjeriva i } \int_{\partial\Omega} |f|^2 \, \mathrm{d}x < \infty \right\}, \tag{1.11}$$

$$H(\operatorname{div}, \Omega) = \left\{ f \in L^2(\Omega) : \frac{\partial f}{\partial x_i} \in L^2(\Omega), 1 \leq i \leq d \right\}. \tag{1.12}$$

Nakon što prvu jednakost u sustavu (1.9) pomnožimo sa test funkcijom $\phi \in \mathcal{V}$, te prointegriramo po Ω dolazimo do varijacijske formulacije

$$\lambda^2 \int_\Omega \rho\, \mathbf{u} \cdot \phi \, \mathrm{d}x + \lambda \int_{\Gamma_A} \beta\, \mathbf{u} \cdot \nu \phi \cdot \nu \, \mathrm{d}S + \int_{\Gamma_R} \alpha\, \mathbf{u} \cdot \nu \phi \cdot \nu \, \mathrm{d}S + \int_\Omega \rho c^2 \operatorname{div} \mathbf{u} \operatorname{div} \phi \, \mathrm{d}x = 0, \tag{1.13}$$

[1] akustička impedancija je omjer zvučnog tlaka p i brzine čestice v u mediju u kojem se rasprostire zvučni val

Kako bismo riješili početni problem nekom od numeričkih metoda, potrebno je (1.13) diskretizirati. Za diskretizaciju ovakvih vrsta problema je najpopularnija metoda konačnih elemenata. Diskretizacija se uvodi kako bismo izbjegli beskonačnodimenzionalan prostor test funkcija, te ga zamijenili nekim konačnodimenzionalnim. Sama diskretizacija uključuje i konstrukciju mreže na domeni Ω po kojoj ćemo integrirati. Mreža je svaka konačna familija $\mathcal{T}$ podskupova od $\bar{\Omega}$ koja ima svojstva

- $\overline{\Omega} = \bigcup_{K\in\mathcal{T}} K$
- svaki $K \in \mathcal{T}$ je poliedarski skup i vrijedi $Int(K) \neq \emptyset$
- za svaka dva različita $K_1, K_2 \in \mathcal{T}$ vrijedi $Int(K_1) \cap Int(K_2) = \emptyset$.

Skupove iz $\mathcal{T}$ nazivamo elementima. Mreža na domeni Ω je dakle subdivizija domene na uniju disjunktnih poliedarskih skupova (elemenata) koji posve prekrivaju Ω. Elementi su najčešće trokuti ili četverokuti u $\mathbb{R}^2$, te tetraedri, heksaedri i prizme u $\mathbb{R}^3$. Jednom kada imamo mrežu na Ω, svakom elementu $K \in \mathcal{T}$ pridružujemo jedan konačnodimenzionalan linearni prostor funkcija koji označavamo sa $\mathcal{V}_h$. Dodatan uvijet regularnosti koji ćemo pretpostavljati je da je aproksimativna metoda konformna, što znači da vrijedi $\mathcal{V}_h \subset \mathcal{V}$, odnosno da aproksimativno rješenje tražimo u prostoru u kojem se nalazi i egzaktno rješenje. Za metodu kažemo da je nekonformna ako uvjet $\mathcal{V}_h \subset \mathcal{V}$ nije zadovoljen. Nekonformnu aproksimaciju koristimo tipično onda kada funkcije iz prostora aproksimativnih rješenja nemaju glatkoću koja se zahtijeva od funkcija iz $\mathcal{V}$. Detalji se mogu naći u [7]. Za $\phi_h \in \mathcal{V}_h$ analogno iz (1.13) imamo

$$\lambda_h^2 \int_\Omega \rho_h \mathbf{u}_h \cdot \phi_h \,\mathrm{d}x + \lambda_h \int_{\Gamma_A} \beta\, \mathbf{u}_h \cdot \nu\phi_h \cdot \nu \,\mathrm{d}S + \int_{\Gamma_R} \alpha\, \mathbf{u}_h \cdot \nu\phi_h \cdot \nu \,\mathrm{d}S + \int_\Omega \rho c^2 \,\mathrm{div}\, \mathbf{u}_h \,\mathrm{div}\, \phi_h \,\mathrm{d}x = 0. \tag{1.14}$$

Nadalje, neka je $\{\psi_j\}_{j=1}^N$ nodalna baza od $\mathcal{V}_h$ (vidi [7]). Sada svaku $\mathbf{u}_h$ možemo zapisati kao kombinaciju baznih funkcija tj. $\mathbf{u}_h = \sum_{j=1}^{N} u_j\psi_j$ te neka je $u = (u_1, u_2, ..., u_N)^T$ vektor njezinih nodalnih komponenti. Definiramo li matrice K,C i M na sljedeći način

$$K_{ij} := \int_\Omega \rho c^2 \,\mathrm{div}\psi_i \,\mathrm{div}\phi_j \,\mathrm{d}x, \quad B_{ij} := \int_{\Gamma_A} \psi_i \cdot \nu\, \phi_j \cdot \nu \,\mathrm{d}S, \quad M_{ij} := \int_\Omega \rho\, \psi_i \cdot \phi_j \,\mathrm{d}x,$$

tada možemo (1.14) zapisati u matričnom obliku

$$\lambda_h^2 Mu + \lambda_h \beta Cu + \alpha Cu + Ku = 0, \tag{1.15}$$

gdje dolazimo do našeg već poznatog kvadratičnog problema svojstvenih vrijednosti. Ovdje K i M predstavljaju krutost i masu fluida, a C efekt apsorpcijskog zida. Ovim primjerom je prikazana jedna druga vrsta formulacije problema (integralna) koja zahtijeva puno više posla da bi se svela na klasični kvadratični problem svojstvenih vrijednosti. Dodatni primjeri kvadratičnih problema svojstvenih vrijednosti mogu se naći u [11].

Poglavlje 2

Spektralna teorija

U ovom poglavlju ćemo se upoznati s osnovnim pojmovima i definicijama koji su potrebni za razumijevanje i rješavanje svakog svojstvenog problema. Definirat ćemo prvo najjednostavniju klasu problema svojstvenih vrijednosti tzv. standardni svojstveni problem, koja će poslužiti kao temelj za izgradnju kompliciranijih klasa problema svojstvenih vrijednosti.

Najprije definirajmo normu. l_p-norma vektora $x \in \mathbb{R}^n$,

$$||x||_p = (\sum_{i=1}^{n} |x_i|^p)^{1/p},$$

gdje je $p \geq 1$. U ostatku ovog rada koristit ćemo uglavnom 2-normu ($p = 2$).

Matrične norme definiraju se na sljedeći način. Neka je $A \in \mathbb{R}^{m\times n}$, tada je p-norma inducirana odgovarajućom vektorskom normom $\|\cdot\|_p$

$$||A||_p = \sup_{x\neq 0} \frac{||Ax||_p}{||x||_p}.$$

Posebni slučajevi su 1-norma, max-norma i Frobeniusova norma koje su definirane redom.

$$||A||_1 = \max_{1\leq j\leq n} \sum_{i=1}^{m} |a_{ij}|,$$

$$||A||_\infty = \max_{1\leq i\leq n} \sum_{j=1}^{n} |a_{ij}|,$$

$$||A||_F = (\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2)^{1/2}.$$

U ovom radu će se koristiti Frobeniusova norma (ukoliko nije naznačeno drugačije). Primjetimo da $\|\cdot\|_F$ nije inducirana vektorskom normom.

2.1 Standardni problem svojstvenih vrijednosti

Definicija 2.1.1. *Neka je A kvadratna matrica dimenzije $n \times n$. Standardni svojstveni problem je zadaća u kojoj se traži skalar λ, poznatiji pod nazivom **svojstvena vrijednost**, takav da je zadovoljena sljedeća jednakost*

$$Ax = \lambda x, \tag{2.1}$$

gdje je $x \neq 0$ odgovarajući desni svojstveni vektor. Lijevi svojstveni vektor $y \neq 0$ definiramo analogno:

$$y^T A = \lambda y^T. \tag{2.2}$$

U nastavku, kada spominjemo svojstveni vektor mislimo na desni. Ekvivalentna formulacija od (2.1) je traženje svojstvene vrijednosti λ koja je netrivijalno rješenje problema:

$$det(A - \lambda I) = 0. \tag{2.3}$$

Skup svih svojstvenih vrijednosti se naziva **spektar** *(eng. spectrum)* i označava sa $\sigma(A)$, a skup svih svojstvenih vektora neke svojstvene vrijednosti λ zajedno s nul vektorom formira potprostor u $\mathbb{C}^n$ koji se naziva **svojstveni potprostor**. Taj svojstveni potprostor odgovara jezgri od $A - \lambda I$.

Definicija 2.1.2. *Jezgra od $A \in \mathbb{R}^{n\times n}$ je potprostor definiran*

$$\mathcal{N}(A) = \{x \in \mathbb{R}^n : Ax = 0\}.$$

Definicija 2.1.3. *Karakteristični polinom od $A \in \mathbb{R}^{n\times n}$ je definiran s $\chi(\lambda) = det(\lambda I - A) = \lambda^n + a_{n-1}\lambda^{n-1} + ... + a_1\lambda + a_0$.*

Iz prethodne definicije vidi se da je traženje svojstvenih vrijednosti usko povezano sa traženjem nultočaka polinoma stupnja n. Prisjetimo se još samo da osnovni teorem algebre tvrdi da polinom stupnja n ima n nultočaka u $\mathbb{C}$, računato s kratnostima.

Kratnost nultočke λ u karakterističnom polinomu je algebarska kratnost svojstvene vrijednosti λ koju označavamo s α. Geometrijska kratnost γ te iste svojstvene vrijednosti λ je dimenzija prostora $\mathcal{N}(A - \lambda I)$. U slučaju kada su geometrijska (γ) i algebarska (α) kratnost jednake kažemo da je pripadajuća svojstvena vrijednost *polujednostavna*. Dodatno, ako su obje kratnosti jednake 1 onda se svojstvena vrijednost naziva *jednostavnom*.

2.2 Kanonski oblici

Čest je slučaj da određenu matricu pokušavamo svesti na jednostavniji oblik kako bi se njome lakše i efikasnije baratalo. U tu svrhu definiramo Schurovu dekompoziciju.

Teorem 2.2.1 (Schurova dekompozicija)**.** *Ako je $A \in \mathbb{C}^{nxn}$ onda postoji unitarna matrica $U \in \mathbb{C}^{nxn}$ takva da je*

$$T = U^* A U \tag{2.4}$$

gornje trokutasta matrica. Na dijagonali od T se nalaze svojstvene vrijednosti od A.

No, pitanje je sada možemo li se samo tako "igrati" s našom originalnom matricom A bez posljedica?

Definicija 2.2.2. *Matrica A i B su slične ako postoji regularna matrica S takva da vrijedi $B = SAS^{-1}$. Dodatno, ako je S unitarna tj. $S^*S = I$, onda govorimo o unitarnoj sličnosti.*

Unitarne transformacije su od velike važnosti u numeričkoj analizi zbog toga što su čuvaju konzistentnost grešaka. Za unitarnu matricu S vrijedi $\|S\|_2 = \|S^{-1}\|_2 = 1$ iz čega slijedi $\|A\|_2 = \|B\|_2$. Konkretno, kada se radi o perturbaciji početne matrice A (npr. greška zaokruživanja u konačnoj aritmetici) tada imamo slučaj

$$S(A + \delta A)S^{-1} = B + \underbrace{S\delta AS^{-1}}_{\delta B} = B + \delta B,$$

$$\|\delta B\|_2 = \|\delta A\|_2.$$

Dakle, greška od A se ne povećava nakon unitarne transformacije sličnosti, dok za druge vrste sličnih transformacija to ne možemo tvrditi. Unitarne transformacije su također korisne u slučajevima kada nam je bitno sačuvati strukturu od A. Npr. ako je A realna simetrična i S realna ortogonalna, onda je i S^TAS realna simetrična.

Teorem 2.2.3. *Slične matrice imaju jednake svojstvene vrijednosti s jednakim kratnostima. Ako je (λ, x) svojstveni par od A i $C = S^{-1}AS$, gdje je S regularna matrica, onda je $(\lambda, S^{-1}x)$ svojstveni par od C.*

No međutim, bitno je napomenuti da obrat ovog teorema ne vrijedi. Matrice koje imaju iste svojstvene vrijednosi nisu nužno slične, kao što je pokazano u sljedećem primjeru.

Primjer 2.2.4. *Neka su A i B definirane na sljedeći način*

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

Obje matrice imaju dvostruku svojstvenu vrijednost jednaku 1, no A ima jedan lijevi svojstveni vektor (e_2) i jedan desni (e_1), dok B ima dva lijeva (e_1 i e_2) i dva desna svojstvena vektora (e_1 i e_2). Zaključak je da matrice A i B ne mogu biti slične.

Što se tiče dijagonalizabilnosti matrice koja je također iznimno važna u razvoju numeričkih metoda, potrebno je analizirati Jordanovu formu.

Teorem 2.2.5 (Jordanova forma)**.** *Neka je A kvadratna matrica reda n. Tada postoji regularna matrica S takva da je*

$$S^{-1}AS = J, \tag{2.5}$$

pri čemu je J tzv. Jordanova forma tj. J je blok-dijagonalna

$$J = diag(J_1(\lambda_1), J_2(\lambda_2), ..., J_k(\lambda_k)),$$

gdje je J_i ,matrica reda m_i oblika

$$J_i = \begin{bmatrix} \lambda_i & 1 & & & & & \\ & \lambda_i & 1 & & & & \\ & & \lambda_i & 1 & & & \\ & & & \cdot & & & \\ & & & & \cdot & & \\ & & & & & \lambda_i & 1 \\ & & & & & & \lambda_i \end{bmatrix} \tag{2.6}$$

i vrijedi $m_1 + m_2 + ... + m_k = n$. Taj m_i je parcijalna kratnost od svojstvene vrijednosti λ_i. Broj parcijalnih kratnosti je geometrijska kratnost za tu svojstvenu vrijednost γ_i, a suma parcijalnih kratnosti je algebarska kratnost α_i. Matrica J je jedinstvena do na permutaciju blokova.

Imajući teorem o Jordanovoj formi u vidu, slijedi da je matrica A dijagonalizabilna tj. Jordanova forma J joj je dijagonalna, ako su sve svojstvene vrijednosti jednostruke tj. jednostavne ($\alpha = \gamma = 1$).

2.3 Generalizirani problem svojstvenih vrijednosti

Definicija 2.3.1. *Neka je A kvadratna matrica dimenzije $n \times n$. Generalizirani svojstveni problem je zadaća u kojoj se traži skalar λ takav da je zadovoljena sljedeća jednakost*

$$Ax = \lambda Bx, \tag{2.7}$$

gdje je $x \neq 0$ *odgovarajući desni svojstveni vektor. Lijevi svojstveni vektor* $y \neq 0$ *definiramo analogno*

$$y^T A = \lambda B y^T. \tag{2.8}$$

Ili ekvivalentno, traži se svojstvena vrijednost λ koja je netrivijalno rješenje sljedećeg problema

$$det(A - \lambda B) = 0. \tag{2.9}$$

Primjetimo da kada je $B = I$ imamo jedan obični standardni svojstveni problem. Svi kompliciraniji modeli, a pritom se misli na polinomijalne probleme koji su kasnije definirani, se pokušavaju svesti na ovu osnovnu jednadžbu.

Glavna razlika između standardnog problema i generaliziranog je da se u generaliziranom može pojaviti svojstvena vrijednost jednaka beskonačnosti (vidi primjer 2.3.2), dok su u standardnom problemu sve svojstvene vrijednosti nužno konačne. Beskonačna vrijednost se pojavljuje u slučajevima kada je matrica B singularna. Također, jedno bitno svojstvo koje vrijedi kod standardnog svojstvenog problema, a gubi se u generaliziranom problemu je da ukoliko je matrica A realna simetrična tada se spektar sastoji samo od realnih svojstvenih vrijednosti. U generaliziranom problemu svojstvenih vrijednosti to ne vrijedi, tj. matrice A i B mogu biti realne i simetrične, ali njihove svojstvene vrijednosti ne moraju biti nužno realne. No međutim, ako su A i B u nekoj linearnoj kombinaciji pozitivno definitne tj. ako postoje α i β takvi da vrijedi $\alpha A + \beta B > 0$, tada je spektar realan. Primjetimo da je dovoljno da je $A > 0$ jer ako β postavimo na 0, uvjet $\alpha A + \beta B > 0$ je automatski zadovoljen (ili obratno ako je $B > 0$ tada α postavimo na 0).

U slučaju da je B regularna ($det(B) \neq 0$), generalizirani problem možemo svesti na standardni tako da (2.7) pomnožimo s lijeva s B^{-1}:

$$B^{-1}Ax = \lambda x. \tag{2.10}$$

Ukoliko je B singularna, a A regularna istu transformaciju možemo izvesti množenjem s desna (2.7) s A^{-1}. U praksi se izbjegavaju ovakve vrste transformacija jer postoji mogućnost gubitka strukture matrica A i B (npr. simetričnost), a svaka informacija o strukturi može biti iznimno korisne prilikom određivanja spektra. Na primjer, u slučaju da su A i B realne simetrične matrice, ne mora nužno vrijediti da je $B^{-1}A$ simetrična matrica, što znači da su bitna svojstva spektra originalnog problema na neki način "zamagljena" što bi nas moglo odvratiti od korištenja već razvijenih algoritama za simetrične probleme. Da bi izbjegli taj problem gubitka simetričnosti koristit ćemo defintinost matrice B. Ako vrijedi da je B pozitivno definitna ($B > 0$) tada iz Cholesky dekompozicije slijedi da je $B = LL^T$, gdje je L donje trokutasta matrica, a samim time je L^T gornje trokutasta. Kada se (2.7) pomnoži s L^{-1} s lijeva, a potom s L^{-T} s desna, B s desne strane se pokrati i dobiva se standardni svojstveni problem

$$Hx = \lambda x. \tag{2.11}$$

gdje je $H = L^{-1}AL^{-T}$.

Dobivena matrica H simetrična što znači da možemo iskoristiti već dobro razvijene metode za simetrični problem. Ovakva vrsta transformacije generaliziranog u standardni svojstveni problem je također puno stabilnija, ali moguća je samo u slučaju pozitivne definitnosti od B (ili $A > 0$, kada sve ide analogno).

Što ako su i A i B singularne? Sljedeća dva primjera će ilustrirati kako riješiti generalizirani svojstveni problem $Ax = \lambda Bx$ i u tom slučaju.

Primjer 2.3.2. *Neka su A i B definirane na sljedeći način*

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix},$$

Tada je svojstvena vrijednost 0 i vrijedi

$$Ae_2 = 0 = 0 \cdot Be_2 = 0 \cdot e_2.$$

Neka je $\mu = \frac{1}{\lambda}$ tada se rješava recipročni problem $Bx = \mu Ax$

$$Be_1 = 0 = \mu Ae_1 = 0e_1.$$

Dakle, $\mu = 0$ je svojstvena vrijednost recipročnog. Sada iz definicije od $\mu = \frac{1}{\lambda}$ iz recipročnog problema slijedi da je $\lambda = \infty$ još jedna svojstvena vrijednost od početnog generaliziranog problema $Ax = \lambda Bx$.

Primjer 2.3.3. *Neka su A i B definirane na sljedeći način*

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad B = A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},$$

Iz (2.9) slijedi da tražimo λ tako da je determinanta od $A - \lambda B$ jednaka 0. Izračunajmo determinantu

$$det(A - \lambda B) = det\left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & 0 \end{bmatrix}\right) = det\left(\begin{bmatrix} 1 - \lambda & 0 \\ 0 & 0 \end{bmatrix}\right) \equiv 0.$$

Dakle, koju god λ uvrstimo u gornju jednadžbu determinanta će biti 0, iz čega slijedi da je svojstvena vrijednost λ proizvoljna vrijednost iz $\mathbb{C}$

2.4 Polinomijalni problem svojstvenih vrijednosti

U ovu kategoriju spada i naš kvadratični problem svojstvenih vrijednosti kao jednostavan primjer za polinomijalne probleme. Zadatak polinomijalnog problema svojstvenih vrijednosti (*eng. Polynomial Eigenvalue Problem* PEP) je naći svojstvenu vrijednost λ i odgovoarajući svojstveni (netrivijalan) vektor x koji zadovoljava

$$P(\lambda)x = 0, \tag{2.12}$$

gdje je P definiran kao polinom m-tog stupnja s matričnim koeficijentima

$$P(\lambda) = \lambda^m A_m + \lambda^{m-1} A_{m-1} + ... + \lambda_1 A_1 + \lambda_0 A_0 \tag{2.13}$$
$$A_k \in \mathbb{C}^{nxn}, k = 0, ..., m.$$

S ovakvim polinomijalnim problemima se susrećemo prilikom analiza sustava običnih diferencijalnih jednadžbi (reda m) s konstantnim koeficijentima

$$\sum_{i=0}^{m} A_i \left(\frac{\mathrm{d}}{\mathrm{d}t}\right)^i x(t) = 0. \tag{2.14}$$

Ideja kod rješavanja ovakvih sustava običnih diferencijalnih sustava je definirati substituciju tako da ovakav sustav diferencijalnih jednažbi m-tog reda svedemo na sustav diferencijalnih jednadžbi prvog reda. Više detalja se može naći u nastavku.

2.4.1 Linearizacija

Svaka metoda kojoj je cilj riješiti polinomijalni problem svojstvenih vrijednosti mora proći prvu fazu koja se naziva linearizacijom.

U literaturi se ponekad mogu pronaći metode koje se nazivaju direktnim. Takav naziv bi nas mogao prevariti da linearizacija nije potrebna tj. da se svojstvene vrijednosti direktno dobivaju iz polinomijalnog oblika (2.12) bez konačnog iterativnog postupka, no to je samo tradicionalan naziv za metodu koja računa sve svojstvene vrijednosti (uglavnom se primjenjuju na male i gusto[1] popunjene matrice) dok se iterativnim nazivaju one metode koje računaju samo dio spektra koji nam je od interesa. Svaki se polinomijalni problem prvo transformira u njemu ekvivalentni generalizirani svojstveni problem, upravo iz razloga što ljudi, a shodno tome i računala koja su kreirali, znaju razmišljati samo linearno.

Ovakva vrsta linearizacije je motivirana rješavanjem diferencijalnih jednadžbi višeg reda tj. redukcijom u diferencijalnu jednadžbu prvog reda. Vratimo li se na diferencijalnu jednadžbu m-tog reda definiranu u (2.14) i uvedemo supstituciju

[1] gusto popunjenim zovemo matrice kojima je omjer nenul i nul elemenata u korist nenul elemenata, dok praznim matricama (*eng. sparse*) nazivamo matrice koje imaju puno više nul elelemenata

$$x_0 = x, \quad x_1 = \frac{\mathrm{d}x_0}{\mathrm{d}t}, \quad \ldots, \quad x_{m-1} = \frac{\mathrm{d}x_{m-2}}{\mathrm{d}t}, \tag{2.15}$$

dolazimo do diferencijalne jednadžbe prvog reda

$$A_m \frac{\mathrm{d}x_{m-1}}{\mathrm{d}t} + A_{m-1}x_{m-1} + ... + A_1 x_1 + A_0 x_0 = 0. \tag{2.16}$$

Sada (2.16) i (2.15) zajedno čine novi sustav jednadžbi. Primjetimo da je dimenziju prostora u kojem sada tražimo rješenje znatno porasla, traži se *mn*-dimenzionalan vektor

$$\begin{bmatrix} x_0 \\ x_1 \\ . \\ . \\ x_{m-1} \end{bmatrix}.$$

U terminima matričnog polinomijalnog problema ovakva redukcija je zapravo jednaka traženju svojstvenih vrijednosti od generaliziranog svojstvenog problema $L(\lambda)x = 0$ gdje je $L(\lambda) = X\lambda + Y, X, Y \in \mathbb{C}^{mn\times mn}$. $L(\lambda)$ je zapravo linearizacija od $P(\lambda) = \sum_{i=0}^{m} \lambda^i A_i$ za koju vrijedi

$$X\lambda - Y \sim \begin{bmatrix} P(\lambda) & 0 \\ 0 & I \end{bmatrix}, \tag{2.17}$$

gdje $\sim$ predstavlja ekvivalenciju matričnih polinoma tj. egzistenciju matričnih polinoma $E(\lambda)$ i $F(\lambda)$ sa konstantnim nenul determinantama, koji zadovoljavaju

$$E(\lambda)(X\lambda - Y)F(\lambda) = \begin{bmatrix} P(\lambda) & 0 \\ 0 & I \end{bmatrix}.$$

Teorem 2.4.1. *Neka je dan matrični polinomijalni problem veličine stupnja m s* $n \times m$ *matričnih koeficijenata,* $P(\lambda) = I\lambda^m + \sum_{i=0}^{m-1} A_j\lambda^j$ *i neka je dana* $nm \times nm$ *matrica*

$$L = \begin{bmatrix} 0 & I & 0 & ... & 0 \\ 0 & 0 & I & ... & 0 \\ . & . & . & . & . \\ . & . & . & . & I \\ -A_0 & -A_1 & .. & & -A_{m-1} \end{bmatrix}.$$

Tada vrijedi

$$I\lambda - L \sim \begin{bmatrix} P(\lambda) & 0 \\ 0 & I \end{bmatrix}.$$

Linearizacija nije jedinstvena, štoviše za pojedini polinomijalni problem postoji beskonačno mnogo linearizacija. Bitno je napomenuti da su sve te linearizacije slične. Vrijedi i obrat, ako imamo linearizaciju L_1 određenog polinomijalnog problema i neku matricu L_2 njoj sličnu, tada znamo da je L_2 linearizacija tog istog polinomijalnog problema.

U ovom radu se fokusiramo na polinomijalni svojstveni problem drugog reda kojeg kraće nazivamo kvadratični problem svojstvenih vrijednosti i definiramo ga s

$$Q(\lambda)x = (\lambda^2 M + \lambda C + K)x = 0. \tag{2.18}$$

Njemu tražimo ekvivalentni linearizirani sustav (*eng. Linear Eigenvalue Problem LEP*)

$$(A - \lambda B)x = 0. \tag{2.19}$$

Glavni nedostatak linearizacije je što se dimenzija početnog problema dvostruko povećava, dakle rješavanje tog problema postaje memorijski zahtjevnije. Kao što je već napomenuto, postoji mnogo načina linearizacija i bitno je odabrati onaj koji će zadržati sva bitna svojstva početnog sustava, a pod svojstva podrazumijevamo određenu strukturu matrica M, C i K. Svaku informaciju o simetričnosti i sličnim strukturalnim svojstvima početnog problema bismo voljeli prenijeti i na linearizirani oblik. Prednost linearizirano oblika koji poštuje određenu strukturu (simetričnost, antisimetričnost,..) je da možemo primijeniti već razvijene numeričke metode za specifične strukture matrica. Također, važnost strukture matrica dolazi do izražaja u spektru tj. rasporedu svojstvenih vrijednosti. Na primjer, za Hermitski kvadratični problem ($Q(\lambda)^* = Q(\bar{\lambda})$) znamo da svojstvene vrijednosti moraju biti realne ili u paru $(\lambda, \bar{\lambda})$ što može poslužiti kao provjera jednom kada su svojstvene vrijednosti izračunate ili pak pojednostaviti račun na način da kad se izračuna jedna svojstvena vrijednost da se spektru odmah dodaje njen kompleksno konjugiran par.

Ovdje navodimo samo neke od najpopularnijih linearizacija kvadratičnog problema svojstvenih vrijednosti.

Linearizacija tipa 1:

$$L_1 = \begin{bmatrix} C & K \\ -N & 0 \end{bmatrix} - \lambda \begin{bmatrix} -M & 0 \\ 0 & N \end{bmatrix}; \tag{2.20}$$

Linearizacija tipa 2:

$$L_2 = \begin{bmatrix} C & -N \\ K & 0 \end{bmatrix} - \lambda \begin{bmatrix} M & 0 \\ 0 & -N \end{bmatrix}; \tag{2.21}$$

Linearizacija tipa 3:

$$L_3 = \begin{bmatrix} K & 0 \\ 0 & N \end{bmatrix} - \lambda \begin{bmatrix} -C & -M \\ N & 0 \end{bmatrix}; \tag{2.22}$$

Linearizacija tipa 4:

$$L_4 = \begin{bmatrix} K & 0 \\ 0 & N \end{bmatrix} - \lambda \begin{bmatrix} -C & N \\ -M & 0 \end{bmatrix}, \tag{2.23}$$

gdje je N regularna kvadratna matrica, te se obično uzima da je identiteta I. Prethodno navedeni tipovi linearizacija nisu strogo definirani, te razni autori različito definiraju linearizacije tipa 1,2,3 i 4.

Prednosti prethodno linearizacija tipova 1,2,3 i 4 su sljedeće:

- linearizacija je moguća čak kada Q(λ) nije regularan
- jednom kada su svojstvene vrijednosti poznate relativno je jednostavno izračunati pripadajuće svojstvene vektore,
- ako je Q(λ) dobro skalirana ($\|M\| = \|C\| = \|K\| \approx 1$) onda rješenje ima malu, *grešku unatrag* i mali *koeficijent osjetljivosti*.

Problem loših linearizacija je da i najmanje perturbacije lineariziranog problema mogu dovesti do ne tako malih perturbacija početnog kvadratičnog problema što vodi ka malim *greškama unatrag* linearizacije, ali puno većim *greškama unatrag* početnog problema.

Poglavlje 3

Perturbacijska analiza

U ovom poglavlju ćemo se upoznati s osnovnim pojmovima koji se javljaju kod analize rješenja. Ono što svaka metoda pokušava izbjeći je da mala promjena prilikom ulaznih podataka izazove veliku promjenu u izlaznim podacima. Osjetljivost rješenja koju mjerimo koeficijentom osjetljivosti, te kvaliteta i stabilnost koje mjerimo analizom greške unatrag su uz egzaktnost rješenja od velike važnosti. Relacija greške unaprijed, koeficijenta osjetljivost i greške unatrag je dana sljedećom relacijom

$$\text{greška unaprijed} \leq \text{koeficijent osjetljivost} \times \text{greška unatrag}. \tag{3.1}$$

Definirajmo prvo homogeni oblik kvadratičnog problema

$$Q(\alpha,\beta) = M\alpha^2 + C\alpha\beta + K\beta^2 \tag{3.2}$$

i njegovu homogenu lineariziraciju

$$L(\alpha,\beta) = A\beta - B\alpha, \tag{3.3}$$

gdje λ odgovara svojstvenom paru $(\alpha,\beta) \neq 0$ za koji vrijedi $\lambda = \frac{\alpha}{\beta}$. Matrice A i B se definiraju ovisno o tipu linearizacije. Neke od linearizacija, a time i matrice A i B, su dane u prethodnom poglavlju (vidi (2.20) (2.21),(2.22),(2.23)).

U daljnjim razmatranjima, perturbaciju kvadratičnog problema ćemo označavati sa $\Delta Q(\lambda)$ te ju definiramo na sljedeći način

$$\Delta Q(\lambda) = \Delta M\lambda^2 + \Delta C\lambda + \Delta K,$$

a perturbirane matrice

$$\widetilde{M} = M + \Delta M, \quad \widetilde{C} = C + \Delta C, \quad \widetilde{K} = K + \Delta K.$$

3.1 Greška unatrag

U praksi je egzaktno rješenje svojstvenog problema nepoznato tj. ne zna se egzaktna svojstvena vrijednost λ, ali se pokušava pronaći njena najbolja aproksimacija $\widetilde{\lambda}$. Postoji više načina na koji se mjeri koliko je dobivena aproksimacija blizu egzaktnog rješenja, a jedan je da izmjerimo koliko treba modificirati početnu matricu A, tako da je $\widetilde{\lambda}$ egzaktno rješenje modificirane matrice $A + \Delta A$. Ukoliko je ta perturbacija relativno mala, onda smatramo da je $\lambda = \widetilde{\lambda}$ prihvatljiva aproksimacija koja zadovoljava $Ax = \lambda x$. Ovakva analiza greške se naziva analizom *greške unatrag*. Ta vrsta analize je vrlo bitna jer se u praksi greške zaokruživanja interpretiraju kao greške podataka, pa podaci često rezultiraju netočnim rješenjima zbog načina spremanja u računalo ili prethodnog izračunavanja.

Definicija 3.1.1. *Neka je $(\widetilde{x}, \widetilde{\lambda})$ aproksimacija svojstvenog para od $Q(\lambda)$. Tada je greška unatrag za aproksimaciju svojstvenog para $(\widetilde{x}, \widetilde{\lambda})$ dana sa*

$$\begin{aligned}\eta(\widetilde{x}, \widetilde{\lambda}) := min\{\epsilon :&(Q(\widetilde{\lambda}) + \Delta Q(\widetilde{\lambda}))\widetilde{x} = 0,\\ &\|\Delta M\| \leq \epsilon\alpha_M, \|\Delta C\| \leq \epsilon\alpha_C, \|\Delta K\| \leq \epsilon\alpha_K\},\end{aligned} \tag{3.4}$$

gdje su α_M, α_C i α_K referentne vrijednosti u odnosu na koje ΔM, ΔC i ΔK moraju biti mali (npr. $\alpha_M = \|M\|, \alpha_C = \|C\|, \alpha_K = \|K\|$). Greška unatrag za aproksimaciju svojstvene vrijednosti $\widetilde{\lambda}$ je dana sa

$$\eta(\widetilde{\lambda}) := \min_{\widetilde{x} \neq 0} \eta(\widetilde{x}, \widetilde{\lambda}). \tag{3.5}$$

Da bi bolje razumijeli prethodnu definiciju, pogledajmo najjednostavniji slučaj kada su matrice M, C i K jednodimenzionalne tj. skalari iz $\mathbb{R}$, a njihova kombinacija jedna točka u prostoru $\mathbb{R}^3$. Tada se kvadratični problem svojstvenih vrijednosti svodi na jednostavnu kvadratičnu jednadžbu za čija rješenja je formula već dobro poznata:

$$\lambda_{1,2} = \frac{-C \pm \sqrt{C^2 - 4MK}}{2M}. \tag{3.6}$$

Kao što je već napomenuto $\widetilde{\lambda}$ je egzaktno rješenje perturbiranog problema, tj. kada su točke M,C i K malo promijenjene u $\widetilde{M} = M + \Delta M$, $\widetilde{C} = C + \Delta C$ i $\widetilde{K} = K + \Delta K$. To rješenje $\widetilde{\lambda}$ se nalazi u nekoj okolini egzaktnog rješenja originalnog problema λ (koje nama nije poznato). Analiza greške unatrag ide korak po korak kroz algoritam od krajnjeg rješenja $\widetilde{\lambda}$ ka početnim vrijednostima $\widetilde{M}, \widetilde{C}, \widetilde{K}$ za koje vrijedi $Q(\widetilde{\lambda})x = 0$ te mjeri udaljenost $\widetilde{M}, \widetilde{C}, \widetilde{K}$ od originalnih vrijednosti M, C i K. Točnije, koliko je potrebno promijeniti vrijednosti od M,C i K da bi vrijedilo $Q(\widetilde{\lambda})x = 0$. Naravno, u interesu nam je da je taj pomak što manji jer time osiguravamo stabilnost algoritma.

U računanju greške unatrag preko (3.4) i (3.5) problem nastaje kada se pojavi beskonačna svojstvena vrijednost. No to se lako riješava, ako QEP prebacimo u homogeni oblik (vidi (3.2)). U homogenom obliku, pretpostavka da je λ beskonačna odgovara pretpostavci da je $\beta = 0$.
Koristeći (3.2) dobiva se ekspilicitna formula za grešku unatrag kvadratičnog problema (dokaz se može naći u [10]).

$$\eta_Q(x, \alpha, \beta) = \frac{\|Q(\alpha, \beta)x\|_2}{(|\alpha|^2\|M\|_2 + |\alpha||\beta|\|C\|_2 + |\beta|^2\|K\|_2)\|x\|_2} \tag{3.7}$$

i njegove linearizacije

$$\eta_L(x, \alpha, \beta) = \frac{\|L(\alpha, \beta)z\|_2}{(|\beta|^2\|A\|_2 + |\alpha|^2\|B\|_2)\|z\|_2}. \tag{3.8}$$

3.2 Osjetljivost

Općenito u numeričkoj analizi, *koeficijent osjetljivosti* mjeri koliko je rješenje problema osjetljivo, tj. u kojoj mjeri će male perturbacije ulaznih podataka utjecati na konačno rješenje. Kažemo da je matrica slabo osjetljiva (ili dobro kondicionirana) za taj problem ukoliko male promjene ulaznih podataka uzrokuju male promjene izlaznih rješenja. Ako to nije slučaj, matrica je loše kondicionirana.

Definicija 3.2.1. *Neka je λ jednostavna nenul konačna svojstvena vrijednost regularnog QEP-a $Q(\lambda)$ sa pripadajućom desnom svojstvenim vektorom x i lijevim svojstvenim vektorom y. Koeficijent osjetljivosti od λ definiramo sa*

$$\kappa(\lambda) = \limsup_{\epsilon \to 0} \left\{ \frac{|\widetilde{\lambda} - \lambda|}{\epsilon|\lambda|} : (Q(\widetilde{\lambda}) + \Delta Q(\widetilde{\lambda}))\widetilde{x} = 0, \; \|\Delta M\| \leq \epsilon\alpha_M, \|\Delta C\| \leq \epsilon\alpha_C, \|\Delta K\| \leq \epsilon\alpha_K \right\}, \tag{3.9}$$

gdje su α_M, α_C i α_K referentne vrijednosti u odnosu na koje ΔM, ΔC i ΔK moraju biti mali.

Pogledajmo još jedanput jednodimenzionalan problem, tj. kada su M,C i K skalari iz $\mathbb{R}$. Sada oko svake te točke M,C i K zadajemo okolinu radijusa $\alpha_M\epsilon$, $\alpha_C\epsilon$ i $\alpha_K\epsilon$ koji određuju koliku perturbaciju originalnih točaka dopuštamo. Prijeđemo li u prostor rješenja, pitanje je koliko je konačno rješenje perturbirane kvadratične jednadžbe udaljeno od rješenja originalne kvadratične jednadžbe. Ta udaljenost se mjeri na način da zamislimo da je rješenje originalnog problema u središtu kruga, a radijus se povećava za ϵ dok taj isti krug ne pokrije i rješenje izračunate aproksimacije, tj. rješenje perturbiranog pronlema. Veličina

radijusa ($n\epsilon$) je upravo koeficijent osjetljivosti. Opet nam je u interesu da je taj interval mogućih rješenja što manji jer time spriječavamo preveliko udaljavanje od egzaktnog rješenja. To je način na koji prevodimo greške iz prostora ulaznih argumenata u prostor rješenja. Dakle, koeficijent osjetljivosti ispituje ponašanje samog kvadratičnog problema prilikom određenih perturbacija i neovisan je o algoritmu. Glavni fokus koeficijenta osjetljivosti je na ponašanje kvadratičnog problema, dok greška unatrag analizira sam algoritam.

Eksplicitna formula za *koeficijent osjetljivosti* kvadratičnog problema glasi (za detalje vidi [10]):

$$\kappa_Q(\alpha,\beta) = \frac{\sqrt{|\beta|^4\|M\|_2^2 + |\alpha|^2|\beta|^2\|C\|_2^2 + |\alpha|^4\|K\|_2^2}}{\left|y^*\left(\widetilde{\beta}\frac{\partial Q}{\partial \alpha} - \widetilde{\alpha}\frac{\partial Q}{\partial \beta}\right)\right|_{(\alpha,\beta)} x}\|y\|_2\|x\|_2 \tag{3.10}$$

gdje su x i y desni odnosno lijevi svojstveni vektor kvadratičnog problema definiranog s (3.2). Slično se definira i *koeficijent osjetljivosti* linearizacije:

$$\kappa_L(\alpha,\beta) = \frac{\sqrt{|\beta|^2\|A\|_2^2 + |\alpha^2|\|B\|_2^2}}{\left|w^*\left(\widetilde{\beta}\frac{\partial L}{\partial \alpha} - \widetilde{\alpha}\frac{\partial L}{\partial \beta}\right)\right|_{(\alpha,\beta)} z}\|w\|_2\|z\|_2 \tag{3.11}$$

gdje su z i w deesni odnosno lijevi svojstveni vektor lineariziranog oblika definiranog s (3.3).

Međutim, kada se radi o kvadratičnom problem i njegovoj linearizaciji može se dogoditi da se koeficijent osjetljivosti i greška unatrag tih dvaju (ekvivalentnih) problema razlikuju, tj. da su rezultati za linearizirani problem iznimno dobri, a za originalni kvadratični problem neprihvatljivi. Razlog takvom disbalansu je što linearizacija kombinira više matrica u jednu. Uzmimo za primjer linearizaciju tipa 2 definiranu u (2.21) i pretpostavimo da su elementi od C mali, npr. reda ϵ, a elementi od K u prosjeku reda 10 . U globalu, analiza osjetljivosti velike matrice A iz lineariziranog oblika će dati mali koeficijent osjetljivosti, u smislu, ako je malo perturbiramo (za neki ϵ), prevladat će ovi veliki elementi od K i koeficijent osjetljivosti će biti prihvatljiv. No, obratimo pozornost sada na svaku matricu zasebno; K ako promjenimo za ϵ, elementi matrice K će ostati relativno blizu svojim originalima, dok C kojemu su elementi reda ϵ promijenjen za ϵ rezultira novom matricom koja je u potpunosti drugačija od svog originala, pa će shodno tome i koeficijent osjetljivosti biti iznimno velik. Jedan takav slučaj ćemo vidjeti na primjeru koji je opisan u sljedećem poglavlju.

Poglavlje 4

Potpuna metoda

Potpuna (ili direktna) metoda je metoda koja računa sve svojstvene vrijednosti i pripadne vektore za postavljeni problem, dok metode koje nisu potpune računaju samo dio spektra (ovisno o tome što nas zanima). U MATLAB-u se može naći već implementirana jedna takva potpuna metoda koja ne samo da računa sve svojstvene vrijednosti za kvadratični problem već i za svaki polinomijalni svojstveni problem višeg stupnja. Ta metoda, koja je ujedno i prva metoda koja je riješavala ovakvu vrstu problema, je implementirana kao funkcija *polyeig*. U ovom poglavlju ćemo usporediti efikasnost i stabilnost metode *polyeig* s njezinom poboljšanom verzijom *quadeig* čiji algoritam su osmislili i implementirali S. Hammarling, C.J. Munro i F. Tisseur [5]. Metode *polyeig* i *quadeig* rješavaju kvadratični problem svojstvenih vrijednosti (QEP)

$$Q(\lambda) = (M\lambda^2 + C\lambda + K)x = 0 \tag{4.1}$$

bazirajući se na **3 osnovna koraka**, a to su:

1. Linearizacija: transformacija QEP u LEP.
2. Riješiti LEP nekom dobrom metodom (npr. QZ metoda).
3. Iz svojstvenih parova od LEP vratiti natrag svojstvene parove od QEP.

4.1 Prednosti

Noviteti, koji daju prednosti metodi *quadeig* pored metode *polyeig* su skaliranje matrica prije računanja svojstvenih vrijednosti i pripremni korak u kojem se otkrivaju i iz problema eliminiraju svojstvene vrijednosti jednake 0 i/ili beskonačnosti. Uz to *quadeig* bira linearizaciju koja smanjuje grešku unatrag i koeficijent osjetljivosti.Također, *quadeig* računa i lijeve i desne svojtvene vektore, dok je *polyeig* ograničen samo na desne.

4.2 Linearizacija

Kao što je već spomenuto, za kvadratični problem postoji beskonačno mnogo linearizacija. Za svaku linearizaciju nužno mora vrijediti det(Q(λ))=det(A-λB), te je ideja odabrati A i B tako da se sačuva struktura matrica početnog kvadratičnog problema što će pridonijeti minimizaciji greške u numerički izračunatim vrijednostima.

Metoda *quadeig* koristi linearizaciju tipa 2:

$$L_2 = L_2(\lambda) = \begin{bmatrix} C & -I \\ K & 0 \end{bmatrix} - \lambda \begin{bmatrix} -M & 0 \\ 0 & -I \end{bmatrix} = A_2 - \lambda B_2, \tag{4.2}$$

dok metoda *polyeig* koristi linearizaciju tipa 1:

$$L_1 = L_1(\lambda) = \begin{bmatrix} K & 0 \\ 0 & I \end{bmatrix} - \lambda \begin{bmatrix} -C & -M \\ I & 0 \end{bmatrix} = A_1 - \lambda B_1. \tag{4.3}$$

Kao što je već spomenuto, upravo izbor linearizacije daje jednu veliku prednost metodi *quadeig*. Linearizacija tipa 2 je bolja od linearizacije tipa 1 u terminima efikasnosti zbog toga što joj je matrica B_2 gornje blok trokutasta, što odmah povlači "jeftinije" računanje Hessenbergovog oblik koji QZ algoritam zahtijeva. Također, ne samo da je B_2 gornje trokutasta već se i cijela L_2 može lako svesti na blok gornje trokutasti oblik, nakon čega se beskonačne svojstvene vrijednosti ili pak one jednake nuli lako detektiraju.

Što se tiče svojstvenih vektora, ovakav tip linearizacije je iznimno pogodan za računanje svojstvenih vektora od $Q(\lambda)$ jednom kada su nam svojstveni vektori od L_2 poznati. U nastavku slijede detalji.
Neka je z desni svojstveni vektor, a w lijevi svojstveni vektor od L_2, tada vrijedi

$$z = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{cases} \begin{bmatrix} \alpha x \\ -\beta K x \end{bmatrix} & ako \quad \alpha \neq 0 \\ \begin{bmatrix} \beta x \\ \alpha M x + \beta C x \end{bmatrix} & ako \quad \beta \neq 0 \end{cases}, \qquad w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} \alpha y \\ \beta y \end{bmatrix}, \tag{4.4}$$

gdje su x i y desni i lijevi svojstveni vektor kvadratičnog problema (4.1) za odgovarajuću svojstvenu vrijednost λ. Iz (4.4) se jasno vidi da se x lako dobiva iz prvih n elemenata vektora z ili rješavanjem sustava $-\beta K x = z_2$ kad je K regularna, dok se lijevi svojstveni vektor y može dobiti lako iz prvih ili zadnjih n elemenata vektora w.

4.3 Skaliranje

Kako matrice M, C i K imaju fizikalno značenje, to nam može prouzrokovati novu vrstu problema gdje upotreba različitih mjernih jedinica može dovesti do velikih razlika u nor-

mama tih matrica, što pak može dovesti do malih greški unatrag linearizirano problema, ali velikih za originalni kvadratični problem. Htjeli bismo da su L_2 i originalni kvadratični problem $Q(\lambda)$ usklađeni u smislu osjetljivosti na male promjene, tj. ako kod rješavanja linearizirano problema $L_2 x = 0$ male promijene u ulaznim podacima uzrokuju male greške prilikom računanja svojstvenih vrijednosti, idealno bi bilo da takva osjetljivost vrijedi i za originalni problem $Q(\lambda)x = 0$, te da im je greška unatrag jednakog reda veličine.

Za homogeni oblik QEP-a koji smo definirali u (3.2) i homogeni oblik LEP-a (3.3) ($\lambda = \frac{\alpha}{\beta}$) vrijedi (za detalje vidi [5]):

$$\frac{1}{\sqrt{2}} \frac{1}{\|p(\alpha,\beta)\|_2} \leq \frac{k_{L_2}(\alpha,\beta)}{k_Q(\alpha,\beta)} \leq 2^{3/2} \frac{\max(1, \|M\|_2, \|C\|_2, \|K\|_2)^2}{\|p(\alpha,\beta)\|_2}, \tag{4.5}$$

$$\frac{1}{\sqrt{2}} \leq \frac{\eta_Q(z_1,\alpha,\beta)}{\eta_{L_2}(z,\alpha,\beta)} \leq 2^{3/2} \frac{\max(1, \|M\|_2, \|C\|_2, \|K\|_2)^2}{\|p(\alpha,\beta)\|_1} \frac{\|z\|_2}{\|z_1\|_2}, \tag{4.6}$$

gdje je

$$p(\alpha,\beta) = [|\alpha|^2\|M\|_2 \quad |\alpha||\beta|\|C\|_2 \quad |\beta|^2\|K\|_2],$$

a svojstvena vrijednost (α,β) je normirana tako da vrijedi $|\alpha|^2 + |\beta| = 1$.

Upravo zbog gornjih i donjih ograda u (4.5) i (4.6) vrijedi da kada $\|M\| = \|C\| = \|K\| \approx 1$ tada su koeficijent osjetljivosti i greška unatrag linearizacije LEP jednake koeficijentu osjetljivosti i greški unatrag originalnog problema QEP tj. $k_Q(\alpha,\beta) = k_{L_2}(\alpha,\beta)$ i $\eta_Q(\alpha,\beta) = \eta_{L_2}(\alpha,\beta)$.

4.3.1 Parametarsko skaliranje

Ideja parametarskog skaliranja je izbalansirati matrične norme na način da transformiramo originalni kvadratični problem $Q(\lambda) = \lambda^2 M + \lambda C + K$ u $\widetilde{Q}(\mu) = \mu^2\widetilde{M} + \mu\widetilde{C} + \widetilde{K}$ gdje je $\lambda = \gamma\mu$. Relacija kojom su $Q(\lambda)$ i $\widetilde{Q}(\mu)$ vezane je sljedeća

$$Q(\lambda)\delta = \mu^2(\gamma^2\delta M) + \mu(\gamma\delta C) + \delta K = \widetilde{Q}(\mu). \tag{4.7}$$

Važno svojstvo ovakve transformacije jest da greška unatrag i koeficijent osjetljivosti ostaju nepromijenjeni.

4.3.2 Fan, Lin i Van Doorenovo skaliranje

Neka su K i M nenul matrice. Fan, Lin i Van Dooren su rješavanjem minimazijskog problema maksimalne udaljenosti matričnih normi $\|M\|, \|C\|, \|K\|$ od 1 tj. problem optimizacije

$$\min_{\gamma,\delta} \max\{\|\widetilde{K}\|_2 - 1, \|\widetilde{C}\|_2 - 1, \|\widetilde{M}\|_2 - 1\}, \tag{4.8}$$

našli optimalne vrijednosti od γ i δ:

$$\gamma = \sqrt{\frac{\|K\|_2}{\|M\|_2}} := \gamma_{FLV}, \quad \delta = \frac{2}{\|K\|_2 + \|C\|_2} := \delta_{FLV}.$$

S ovim izborom parametara vrijedi da norme matrica nisu veće od 2 te je pokazano da koeficijent osjetljivosti $\kappa(\alpha,\beta)$ i grešku unatrag $\eta(\alpha,\beta)$ iz (4.5) i (4.6) možemo ograničiti preko teorijskog rezultat koji je dokazan [10], a glasi:

$$\frac{1}{\sqrt{2}} \leq \|p(\widetilde{\alpha},\widetilde{\beta})\|_2^{-1} \leq \frac{\sqrt{3}}{2} \min\left\{1 + \tau_Q, \frac{1}{|\widetilde{\alpha}\widetilde{\beta}|}\right\}. \tag{4.9}$$

No međutim, ovi parametri neće uvijek funkcionirati, tj. neće uvijek voditi ka boljem riješenju. Problem nastaje kada je početni sustav prejako pregušen. Ta pregušenost se mjeri sa

$$\tau_Q = \frac{\|C\|_2}{(\|M\|_2\|K\|_2)^2}. \tag{4.10}$$

Dva su slučaja:

- $\tau_Q \lessapprox 1$ znači da sustav nije prejako pregušen. Primjenimo li (4.9) na (4.5) i (4.6) slijedi da je $k_{L_2} \approx k_{\widetilde{Q}}$ i $\eta_{\widetilde{Q}} \approx \eta_{L_2}$ tj. da male greške unatrag stabilnih algoritama (npr. QZ metoda) primijenjenih na linearizirani oblik znače male greške i za naš početni problem $Q(\lambda)x = 0$.
- $\tau_Q > 1$ znači da je sustav prejako pregušen, što znači gornja ograda u (4.9) nije zagarantirana pa skaliranje ne vodi nužno poboljšanju rješenja.

4.3.3 Tropsko skaliranje

Za razliku od Fan, Lin i Van Doorenova skaliranja, tropsko skaliranje nalazi rješenje kako skalirati čak i u slučaju kad je sustav previše pregušen.
Definirajmo prvo tropski polinom

$$q_{trop}(x) = \max\{\|M\|_2 x^2, \|C\|_2 x, \|K\|_2\}, \quad x \in [0,\infty). \tag{4.11}$$

Kada je $\tau_Q \lessapprox 1$ ovaj polinom ima jedan dvostruki tropski korjen

$$\gamma_{trop}^{+} = \gamma_{trop}^{-} = \sqrt{\frac{\|K\|_2}{\|M\|_2}} = \gamma_{FLV},$$

a kada je $\tau_Q > 1$ tropski polinom ima dva različita tropska korijena

$$\gamma^+_{trop} = \frac{||C||_2}{||M||_2}, \quad \gamma^-_{trop} = \frac{||K||_2}{||C||_2}.$$

Ako su tropski korijeni dobro odvojeni ($\gamma^+_{trop} >> \gamma^-_{trop}$) i sve matrice dobro kondicionirane, onda je n najvećih (po modulu) svojstvenih vrijednosti veličine reda od γ^+_{trop}, a n najmanjih svojstvenih vrijednosti po veličini reda odgovara γ^-_{trop}.

Eksperimentalno je pokazano da za što veću točnost aproksimacija velikih svojstvenih vrijednosti treba uzeti parametre

$$\gamma = \gamma^+_{trop}, \quad \delta = (q_{trop}(\gamma^+_{trop})))^{-1},$$

a za što veću točnost aproksimacija malih svojstvenih vrijednosti treba uzeti parametre

$$\gamma = \gamma^-_{trop}, \quad \delta = (q_{trop}(\gamma^-_{trop}))^{-1}.$$

4.4 Deflacija

Ideja deflacije je detektirati beskonačne svojstvene vrijednosti ili pak one jednake 0 koje su zapravo posljedica singularnosti od $Q(\lambda)$ i prije ikakvih metoda za računanje svojstvenih vrijednosti, ugušiti (deflatirati) te svojstvene vrijednosti. Karakteristični polinom kvadratičnog problema svojstvenih vrijednosti je dan kao $det(Q(\lambda)) = det(M)\lambda^{2n}$ + *članovi manjeg reda*, iz čega slijedi da ukoliko je M regularna onda $Q(\lambda)$ ima $2n$ konačnih svojstvenih vrijednost, u suprotnom, kada je M singularna, onda $Q(\lambda)$ ima d konačnih i $2n$-d beskonačnih svojstvenih vrijednosti, gdje smo sa d označili stupanj polinoma $det(Q(\lambda))$. Primjetimo još da je λ svojstvena vrijednost od $Q(\lambda)$ ako i samo ako je $1/\lambda$ svojstvena vrijednost od recipročnog problema $\lambda^2 Q(\lambda^{-1}) = \lambda^2 K + \lambda C + M$ gdje je 0 recipročna svojstvena vrijednost od ∞ (i obrnuto). Označimo li sa $r_M < n$ rang od M i sa $r_K < n$ rang od K, slijedi da $Q(\lambda)$ ima barem $n - r_K$ svojstvenih vrijednosti jednakih 0 i $n - r_M$ beskonačnih svojstvenih vrijednosti. Kao što je već spomenuto, deflacija je važan pripremni korak u metodi *quadeig* koji bitno utječe na kvalitetu aproksimacija ostalih konačnih svojstvenih vrijednosti. U *quadeig-u* se provjeravaju rangovi od K i M koristeći QR dekompoziciju. Neka je

$$Q_M^* M P_M = \begin{bmatrix} R_{11}^M & R_{12}^M \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} R^M \\ 0 \end{bmatrix}, \qquad Q_K^* K P_K = \begin{bmatrix} R_{11}^K & R_{12}^K \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} R^K \\ 0 \end{bmatrix}. \tag{4.12}$$

QR faktorizacija za M i K, gdje $R^i_{11} \in \mathbb{R}^{r_i \times r_i}$ i $R^i_{12} \in \mathbb{R}^{r_i \times (n-r_i)}$, $i \in \{M, K\}$. Pomoću ovih faktorizacija, linearizaciju tipa 2 (vidi 4.2) koju *quadeig* koristi transformiramo u gornje

trokutastu formu

$$Q^* L_2 P = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ 0 & A_{22} & A_{23} \\ 0 & 0 & 0_{n-r_K} \end{bmatrix} - \lambda \begin{bmatrix} B_{11} & B_{12} & B_13 \\ 0 & 0_{n-r_M} & B_{23} \\ 0 & 0 & B_{33} \end{bmatrix}. \tag{4.13}$$

Kada su A_{22} i B_{33} singularne, onda (4.13) otkriva $n - r_K$ svojstvenih vrijednosti jednakih 0 i $n - r_M$ beskonačnih svojstvenih vrijednosti. Ostatak svojstvenih vrijednosti dobivamo riješavajuči generalizirani problem $A_{11}x = \lambda B_{11}x$ koji je manje dimenzije od početnog i čije svojstvene vrijednosti su konačne.

4.5 Svojstveni vektori

Kao što je već spomenuto, prednost našeg izbora linearizacije je jednostavan izračun pripadnih svojstvenih vektora. Način na koji računamo svojstvene vektore ovisi o tome radi li se o lijevom svojstvenom vektoru ili desnom, radi li se o deflatiranoj svojstvenoj vrijednosti ili "običnoj", te kakvo skaliranje je korišteno (ako uopće). Slijedi detaljniji opis kako računamo desne, a ako lijeve svojstvene vektore.

4.5.1 Desni svojstveni vektori

Kada su M i K singularne, vektori koji određuju desnu jezgru $\mathcal{N}(M)$ odnosno $\mathcal{N}(K)$ su ujedno i svojstveni vektori koji odgovaraju svojstvenoj vrijednosti ∞ odnosno svojstvenoj vrijednosti 0. Ta jezgra se lako dobiva iz (4.12) gdje Householderovim reflektorom svodimo R_{12} na nulu, na potpunu ortogonalnu dekompoziciju

$$Q_M^* M Z_M = \begin{bmatrix} T_{11}^M & 0 \\ 0 & 0 \end{bmatrix}, \qquad Q_K^* K Z_K = \begin{bmatrix} T_{11}^K & 0 \\ 0 & 0 \end{bmatrix}. \tag{4.14}$$

Zadnjih $n - r_K$ stupaca od Z_K odgovaraju svojstvenim vektorima od $Q(\lambda)$ sa svojstvenom vrijednosti $\lambda = 0$, a zadnjih $n - r_M$ stupaca od Z_M svojstvenim vektorima od $Q(\lambda)$ sa svojstvenom vrijednosti $\lambda = \infty$. Što se tiče desnih svojstvenih vektora od "običnih" svojstvenih vrijednosti tj. onih koje nisu deflatirane, oni se lako izračunaju jednom kada je Shurova dekompozicija od $A_{11} - \lambda B_{11}$ poznata. Računanje natrag svojstvenih vektora originalnog kvadratičnog problema iz linearizirane forme je objašnjeno ranije u ovom poglavlju (vidi (4.4)).

4.5.2 Lijevi svojstveni vektori

Ukoliko je K singularna tada zadnjih $n - r_K$ stupaca od Q_K u (4.12) reprezentira lijeve svojstvene vektore za $n - r_K$ deflatiranih svojstvenih vrijednosti jednakih 0, dok u slučaju

kada je M singularna zadnjih $n - r_M$ stupaca od Q_M u (4.12) reprezentira $n - r_M$ svojstvenih vektora od ∞. Lijeve svojstvene vektore za odgovarajuće svojstvene vrijednosti koje nisu deflatirane računamo na isti načina kao i desne. Analogno kao i za desne svojstvene vektore, prvo se traže svojstveni vektori za linearizaciju L_2 koristeći Shurovu dekompoziciju od $A_{11} - \lambda B_{11}$, a onda se iz (4.4) računaju natrag svojstveni vektori za originalni kvadratični problem.

4.6 Algoritam

Ulazni podaci su $n \times n$ matrice M, C i K koje definiraju kvadratični problem svojstvenih vrijednosti. Po izboru dodajemo ulaznim podacima strukturu *options* koja nudi sljedeće opcije skaliranja:

option.pscale	tip skaliranja
0	bez skaliranja
1	Fan, Lin i Van Dooren skaliranje ako vrijedi $\frac{\|C\|}{\sqrt{\|M\|\|K\|}} < 10$ u suprotnom bez skaliranja
2	Fan, Lin and Van Dooren skaliranje
3	tropsko skaliranje sa γ^{+}_{trop}
4	tropsko skaliranje sa γ^{-}_{trop}

kao i opcije za deflaciju:

options.tol	izvedba
tol	Korisnik može izabrati toleranciju *tol* koju koristi kod računanja ranga u smislu da sve što je manje od *tol* se smatra nulom. *Defaultna* tolerancija je dana sa $tol = nu * max\{\|M\|, \|C\|, \|K\|\}$ gdje je $nu = n * \epsilon$ (u MATLABU: $\epsilon = 2^{-52}$)
-1	nema deflacije

Izlazni podaci su izračunate svojstvene vrijednosti (njih $2n$), lijevi/desni svojstveni vektori, koeficijent osjetljivosti svojstvenih vrijednosti i greška unatrag svojstvenih vektora, te vrijednost τ_Q koja je definirana u (4.10)

QUADEIG metoda

1. Skalirati matrice M,C i K
2. Linearizirati QEP u L_2 te svesti linearizaciju L_2 na gornje trokutasti oblik (oblik definiran u (4.13)) koristeći *tol*
3. Izračunati svojstvene vrijednosti preko Shurove dekompozicije od $A_{11} - \lambda B_{11}$
4. (po izboru) Izračunati lijeve/desne svojstvene vektore
5. (po izboru) Izračunati koeficijent osjetljivosti svojstvenih vrijednosti i grešku unatrag svojstvenih vektora

4.7 Eksperimenti

Primjeri na kojima ćemo testirati efikasnost potpune metode *quadeig* se mogu naći u NLEVP kolekciji (*eng. NonLinear EigenValue Problem*), kolekciji koja nudi još mnoštvo primjera koji se bave problemima s kojima se u praksi susreću inžinjeri. Za detalje tih primjera vidi [3]. Kao što je ranije spomenuto, jedan od kriterija koji mjeri kvalitetu rješenja je analiza greške unatrag. Za svaku svojstvenu vrijednost, tj. za svaki pripadajući desni svojstveni vektor računamo grešku unatrag, a usporedbom najveće greške unatrag pokušat ćemo odgovoriti na pitanje je li *quadeig* zaista bolja metoda od *polyeig* metode. Tablica prikazana u nastavku prezentira neke primjere iz NLEVP kolekcije gdje je greška unatrag desnog svojstvenog vektora izračunatog *quadeig* metodom puno manja nego greška unatrag svojstvenog vektora dobivenog *polyeig* metodom. Vrijednosti koje se nalaze u tablici odgovaraju $max(\eta(x, \alpha, \beta))$. U metodi *quadeig* je korišteno Fan, Lin i Van Doorenovo skaliranje.

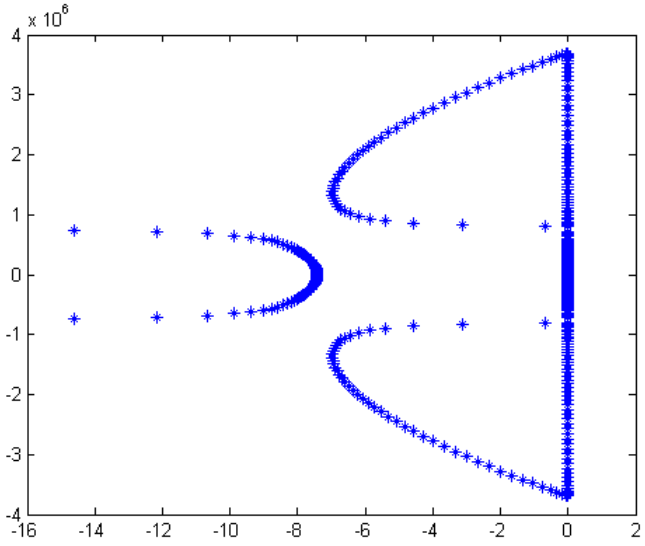
Problem	quadeig	polyeig
cd_player	9.6721e-16	1.6652e-10
hospital	6.9702e-16	5.9588e-13
power_plant	3.6830e-16	1.3037e-08
damped_beam	5.5467e-16	3.4251e-09

U ovim primjerima se jasno se vidi da je *quadeig* itekako efikasnija od *polyeig*-a. Razlog tome je skaliranje koje je u ovim slučajevima odigralo bitnu ulogu.

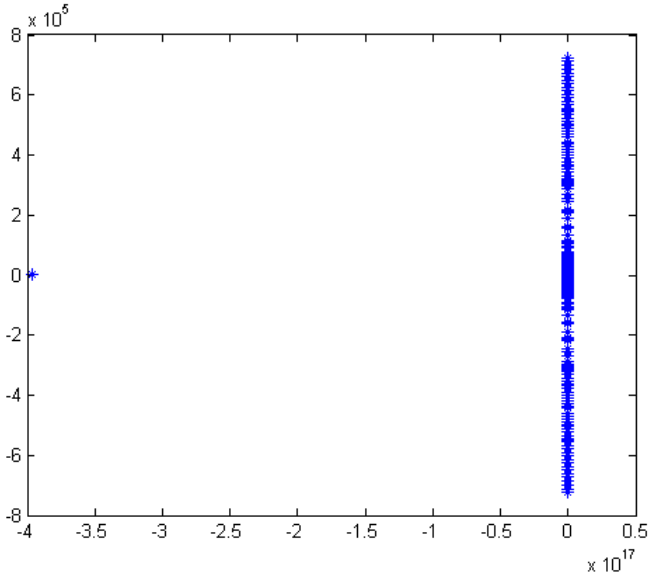
Primjer 1 Vratimo se sada primjeru mehaničkih oscilacija kojeg smo ilustrirali u prvom poglavlju. Taj je primjer dan u kolekciji NLEVP pod nazivom *damped_beam*, a sam model mehaničkih oscilacija je preuzet iz analize vibracija u prigušenoj gredi koja je pričvršćena za oba kraja, a prigušena u sredini. U ovom konkretnom primjeru,vrijedi da je $M > 0$ i $K > 0$, a $C = e_n c e_n^T \geq 0$. Jasno je da što je prigušenost veća da će sustav biti stabliniji, što znači da će veličina realnog dijela svojstvenih vrijednosti biti obrnuto proporcionalna jačini prigušivanja c tj. što je veći c, to će svojstvene vrijednosti više pobjeći u lijevo (vidi sliku 4.1). Slijedeći rezultati vrijede za $c = 5$, a $M, C, K \in \mathbb{R}^{200\times 200}$.

U nastavku su ilustrirane svojstvene vrijednosti izračunate koristeći metode *quadeig* i *polyeig*. Najveća greška unatrag za metodu *polyeig* je **3.4251e − 09** dok je za *quadeig*, gdje je korišteno Fan,Lin, Van Doorenovo skaliranje, greška samo **5.5467e − 16**. S druge strane, *quadeig* bez skaliranja daje puno gore rezultate od *polyeig*; najveća greška unatrag u tom slučaju je jednaka **2.1505e − 04**. Razlog ovako lošoj izvedbi *quadeig-a* je definitvno velika razlika matričnih normi; $||M|| = 0.0513$, $||C|| = 5$, $||K|| = 1.0645e+10$. Pogledamo li sad kakvu linearizaciju koristi *polyeig* (4.3), a kakvu *quadeig* (4.2) vidi se da *polyeig* u B_1 kombinira dvije matrice M i C kojima je veličina norme relativno blizu, tj. elementi im se ne razlikuju previše po modulu, pa možemo reći da je matrica B_1 u dobrom balansu, dok se u matrici B_2 miješaju C i K kojima je razlika u normama reda veličine 10, zbog čega je ona jako osjetljiva na bilo kakve perturbacije. Ovaj slučaj je jasni pokazatelj koliko važnu ulogu ima skaliranje. Na slici 4.2 se vidi da svojstvene vrijednosti izračunate metodom *quadeig* bez skaliranja odgovaraju svojstvenim vrijednostima neprigušenog sustava. Razlog simetričnosti spektra je taj da su M,C i K realne.

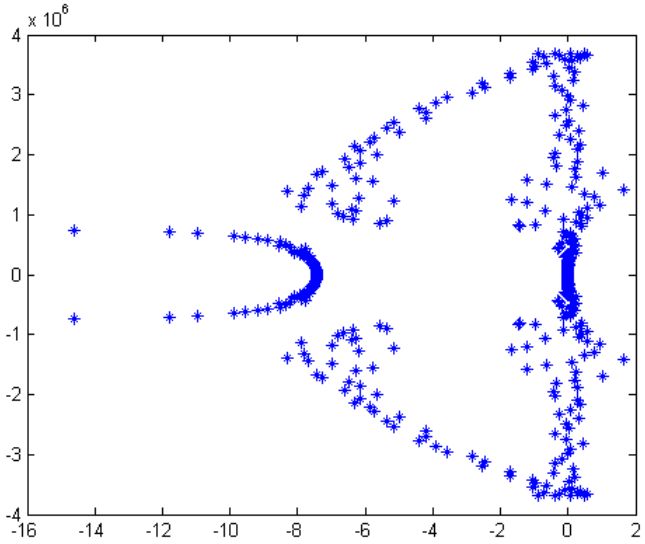
Izborom tropskog skaliranja umjesto Fan, Lin, Van Doorenove svojstvene vrijednosti sa slike 4.1 ostaju skoro pa iste (razliku je skoro nemoguće vidjeti) i greška unatrag se ne mijenja, ali koeficijent osjetljivosti se smanjuje sa $3.0842e + 6$ na $1.1404e + 4$ što sugerira da je tropsko skaliranje bolje u ovom slučaju.



Slika 4.1: Quadeig sa FVL skaliranjem u Primjeru 1.

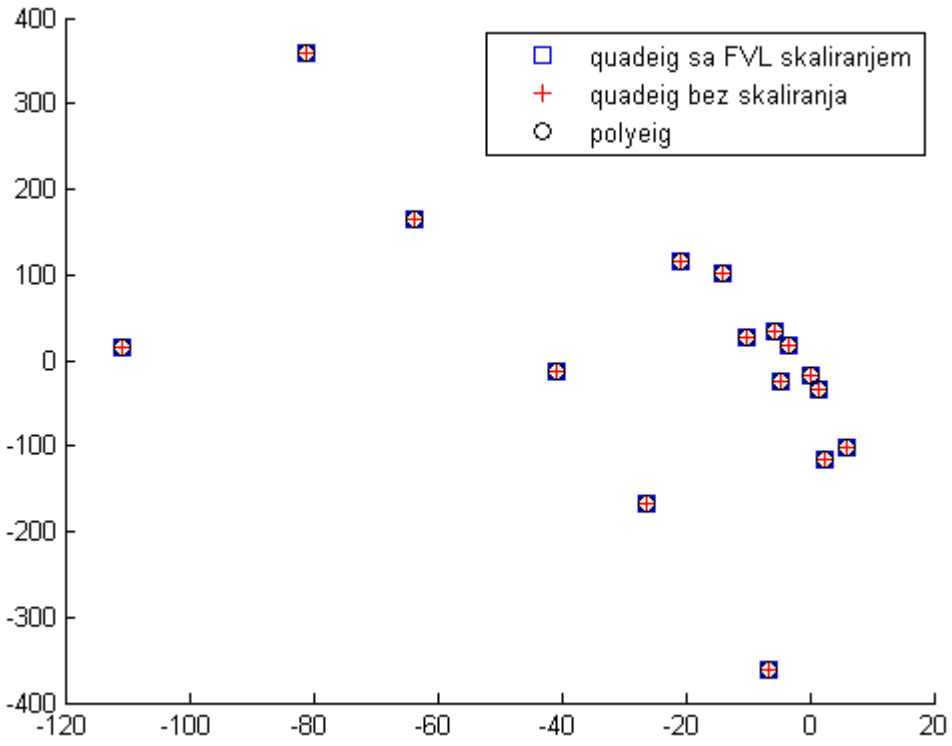


Slika 4.2: Quadeig bez skaliranja u Primjeru 1.



Slika 4.3: Polyeig u Primjeru 1.

Primjer 2 Primjer *power_plant* koji se također može naći u NLEVP kolekciji je dimenzija 8×8, a opisuje dinamičko ponašanje elektrane pojednostavljeno u sustav od 8 nepoznanica. Za razliku od prethodnog primjera, ovdje su veličine matrica u relativnom balansu; $||M||$=2.5154e+08, $||C||$=4.3710e+10, $||K||$=1.7168e+13. Nijedna norma ne odskače svojom veličinom od ostalih što znači da su elementi matrica bliski (za razliku od primjera 1). Kao što je već prezentirano prethodno u tablici, prilikom računanja svojstvenih vrijednosti metodom *polyeig* najveća greška unatrag je **5.9588e-13**, dok metodom *quadeig* (uz Fan, Lin i Van Doorenovo skaliranje) je samo **5.9702e-16**. Za razliku od prethodnog primjera, u slučaju da *quadeig* ne izvrši skaliranje, greška unatrag je **5.5858e-15** što je i dalje bolji rezlutat od *polyeig-a*. Slika 4.4 prikazuje svojstvene vrijednosti koje daju metode *quadeig* i *polyeig*. Primjetimo da razlike gotovo da ni nema, za razliku od prethodnog primjer gdje je razlika bila vidljiva. Mala dimenzija ovog problema također pridonosi točnosti rješenja.



Slika 4.4: Aproksimacije svojstvenih vrijednosti u Primjeru 2.

Poglavlje 5

Iterativne metode

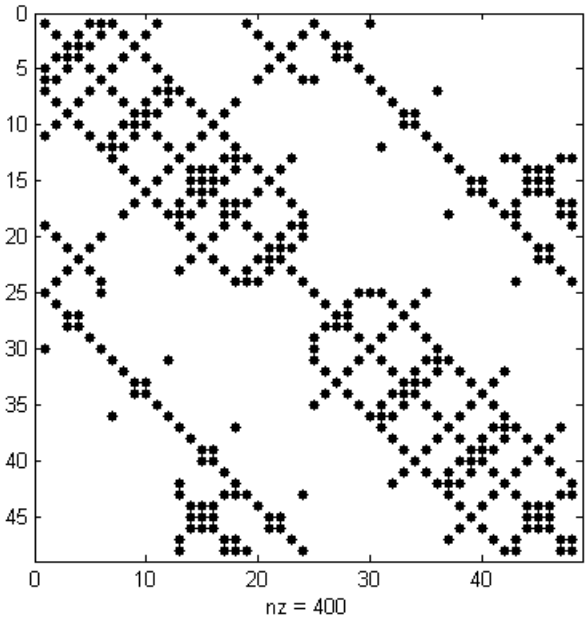
Kao što je već spomenuto, iterativna metoda, za razliku od potpune, računa samo dio spektra. Iterativnih metoda ima mnogo, a one su nam od velike koristi kada je početni kvadratični problem velikih dimenzija, a nas zanimaju samo određene svojstvene vrijednosti (najveće, najmanje ili pak unutar nekog određenog $\Omega \subset \mathbb{C}$). Na sličan način kako je u uvodnom primjeru o akustičnim vibracijama bilo potrebno diskretizirati beskonačnodimenzionalan prostor funkcija nekim konačnodimenzionalnim, metode koje će biti opisane u nastavku će aproksimirati prostor velikih dimenzija nekim prostorom manje dimenzije, koji sadrži dovoljno dobru informaciju o dijelu spektra koji nas zanima. Zbog praktičnosti i dobrih rezultata, nekoliko iterativnih metoda se razvilo tj. i dalje se razvija, no naš fokus će najviše biti na metodama baziranim na Krilovljevim potprostorima, koje računaju više svojstvenih vrijednosti istovremeno. Jedna takva metoda je Arnoldijeva metoda, koja riješava standardni svojstveni problem, tj. SOAR (*eng. Second Order ARnoldi*) zadužen za riješavanje kvadratičnog svojstvenog problema. Dakle, ideja je projicirati početni problem na adekvatno odabran potprostor manjih dimenzija iz kojeg se nekom od direktnih (potpunih) metoda računaju svojstvene vrijednosti tog reduciranog kvadratičnog svojstvenog problema. Izračunate svojstvene vrijednosti reduciranog kvadratičnog problema predstavljaju aproksimacije svojstvenih vrijednosti originalnog problema. U nastavku ovog poglavlja, prvo ćemo opisat osnovnu Arnoldijevu metodu za standardni svojstveni problem koja je motivirala razvitak SOAR metode za kvadratični problem svojstvenih vrijednosti.

5.1 Arnoldijeva metoda

Arnoldijevu metodu ćemo detaljnije objasniti na primjeru standardnog problema svojstvenih vrijednosti

$$Ax = \lambda x. \tag{5.1}$$

Što kada je A kvadratna matrica velikih dimenzija $n \times n$ i to rijetko popunjena (kao što je čest slučaj u praksi), a nas zanima samo podskup njenih svojstvenih vrijednosti? Struktura jedne takve matrice se može vidjeti na slici 5.1 Prikazana matrica dolazi iz dinamičke analize strukture sustava, a može se naći u Harwell-Boeing kolekciji na *Matrix Market*-u.



Slika 5.1: Rijetko popunjena matrica

Crne zvijezdice predstavljaju nenul elemente, a ostali elementi su nule. Primjetimo da je puno manje nenul elemenata, što znači da je matrica rijetko popunjena. Sada kada bismo na takvu matricu primijenili neku od direktnih metoda, nule bi prešle u nenul element i ova elegantna struktura bi bila izgubljena. Važnost i glavna prednost rijetko popunjenih matrica je ušteda memorije, jer se pohranjuju samo nenul elementi, a operacija $v \to Av$ je zbog strukture iznimno efikasna.

Ideja Arnoldijeve metode je projicirati početni problem velikih dimenzija u pogodno dobar potprostor manjih dimenzija gdje ćemo lakše i efikasnije doći do željenih svojstvenih vrijednosti. Odmah se postavlja pitanje, koji će to prostor dobro aproksimirati prostor rješenja?

Definicija 5.1.1. *Neka je $A \in \mathbb{R}^{n\times n}$ i $b \in \mathbb{R}^n \backslash \{0\}$. Krilovljeva matrica reda i je definirana s $K_i \equiv K_i(A, b) = [b, Ab, ..., A^{i-1}b]$, a i-ti Krilovljev potprostor $\mathcal{K}_i$ je definiran kao slika od K_i, $\mathcal{K}_i \equiv \mathcal{K}_i(A, b) = \mathcal{R}(K_i)$.*

Pri računanju projekcije, zbog numeričke stabilnosti je bitno zadati $\mathcal{K}_i$ u ortonormiranoj bazi koja se može dobiti QR faktorizacijom. Pretpostavimo sada da su svojstvene vrijednosti od A jednostruke, tj. neka je matrica A dijagonalizabilna s n linearno nezavisnih svojstvenih vektora $v_1, v_2..., v_n$ i n svojstvenih vrijednosti koje numeriramo tako da vrijedi $|\lambda_1| \geq |\lambda_2| \geq ... \geq |\lambda_n|$. Kako su v_j, $j = 1, .., n$ svojstveni vektori znamo da vrijedi $Av_j = \lambda_j v_j$, također svaki vektor b možemo zapisati kao linearnu kombinaciju $b = \beta_1 v_1 + \beta_2 v_2 + ... + \beta_n v_n$. Kada primjenimo A na b i iskoristimo definiciju svojstvenih vrijednosti dobiva se

$$\begin{aligned} Ab &= \beta_1 Av_1 + \beta_2 Av_2 + ... + \beta_n Av_n \\ &= \beta_1 \lambda_1 v_1 + \beta_2 \lambda_2 v_2 + ... + \beta_n \lambda_n v_n. \end{aligned} \tag{5.2}$$

Nadalje, za potenciranu matricu A^k vrijedi

$$\begin{aligned} A^k b &= \beta_1 \lambda_1^k v_1 + \beta_2 \lambda_2^k v_2 + ... + \beta_n \lambda_n^k v_n \\ &= \lambda_1^k \left(\beta_1 v_1 + \beta_2 \underbrace{\left(\frac{\lambda_2}{\lambda_1}\right)^k}_{\to 0} v_2 + ... + \beta_n \underbrace{\left(\frac{\lambda_n}{\lambda_1}\right)^k}_{\to 0} v_n \right), \quad \text{uz pretpostavku da vrijedi } |\lambda_1| > |\lambda_2|. \end{aligned} \tag{5.3}$$

Dakle, kada je $\beta_1 \neq 0$, smjer vektora $A^k b$ zapravo konvergira ka smjeru svojstvenog vektora v_1. Ovakva metoda potencija je efikasan način za računanje dominantnog svojstvenog para (λ_1, v_1). Što bliže je λ_1 ostalim svojstvenim vrijednostima u apsolutnoj vrijednosti, to će konvergencija ka dominantnom svojstvenom paru biti sporija. No istovremeno, kako se k povećava tako se i kut koji vektor $A^k b$ zatvara s v_1 smanjuje i metoda postaje nestabilna tj. povećava se osjetljivost QR faktorizacije. Ideja Arnoldijevoga algoritma je ortonormirati niz $b, Ab, ...A^k b$ tj. pronaći ortonormiranu bazu $\{q_1, q_2, ..., q_k\}$ za k-ti Krilovljev potprostor u kojem ćemo tražiti željene svojstvene vrijednosti.

Definicija 5.1.2. *Kažemo da je $n \times n$ matrica H u Hessenbergovoj formi (Hessenbergova matrica) ako je $H_{ij} = 0$ za $i > j + 1$*

Hessenbergova matrica H dimenzije 6×6 je sljedećeg oblika:

$$H = \begin{bmatrix} * & * & * & * & * & * \\ * & * & * & * & * & * \\ 0 & * & * & * & * & * \\ 0 & 0 & * & * & * & * \\ 0 & 0 & 0 & * & * & * \\ 0 & 0 & 0 & 0 & * & * \end{bmatrix}.$$

Propozicija 5.1.3. *Za svaki indeks $i = 1, 2, ..$ je $dim(\mathcal{K}_i) \leq i$. Nadalje, $A\mathcal{K}_i \subseteq \mathcal{K}_{i+1}$, i postoji indeks $l \leq n$ za kojeg je $\mathcal{K}_1 \subset \mathcal{K}_2 \subset ... \subset \mathcal{K}_i \subset \mathcal{K}_{i+1} ... \subset \mathcal{K}_l = \mathcal{K}_{l+1}, A\mathcal{K}_l \subseteq \mathcal{K}_l$. Potprostor $\mathcal{K}_l$ je najmanji A–invarijantni potprostor koji sadrži vektor b. Osim toga,*

- *Ako je Q_l ortonormirana baza dobivena QR faktorizacijom matrice K_l, onda je u toj bazi $P_{\mathcal{K}_l} A_{\mathcal{K}_l}$ reprezentiran gornje Hessenbergovom matricom $\hat{H}_l = Q_l^* A Q_l$ s $(\hat{H}_l)_{i+1,i} \neq 0, i = 1, ..., l-1$. Ako je $A^* = A$ onda je $\hat{H}_l$ tridijagonalna.*
- *Ako je A regularna, onda $\mathcal{K}_l$ sadrži $x = A^{-1}b$ i vrijedi formula $A^{-1}b = \|b\|_2 Q_l(\hat{H}_l^{-1} e_1)$.*
- *Sa svakim proširenjem matrice Q_l do unitarne matrice $Q = [Q_l, Q_l^{\perp}]$ je $Q^* A Q$ 2×2 blok gornje trokutasta matrica s $\hat{H}_l$ na poziciji (1,1). Specijalno je svaka svojstvena vrijednost od $\hat{H}_l$ ujedno i svojstvena vrijednost od A.*

Iz prethodne propozicije 5.1.3 slijedi da je naš tražen Krilovljev potprostor zapravo najmanji invarijanti prostor operatora A koji sadrži $A^{-1}b$. Dokaz propozicije, kao i više detalja o Krilovljevim potprostorima, se može naći u [4]. To znači da kada djelujemo s matricom A na vektore iz Krilovljevog potprostora, htjeli bismo da oni ostaju u istom tom potprostoru. Jasno je da to neće biti baš moguće u konačnoj aritmetici, pa ćemo umjesto zahtjeva da su Krilovljev prostori $\mathcal{K}_i$ i $\mathcal{K}_{i+1}$ jednaki zahtjevati da su vrlo blizu jedan drugoga. Idealna situacija bi bila $\mathcal{K}_l = \mathcal{K}_{l+1}$ (u konačnoj aritmetici $\mathcal{K}_l \approx \mathcal{K}_{l+1}$) gdje je l puno manji od n, tj. da je pronađen invarijantni potprostor malih dimenzija.

Pokažimo sada kako se dobiva tražena ortonormirana baza za Krilovljev prostor. Jedan od načina je formirati niz matrica $K_1, ..., K_l$ i rekurzivno izračunati pripadne QR-faktorizacije. Ako je $K_i = [k_1, ...k_i] = Q_i R_i$ već izračunata, onda slijedi:

$$\begin{aligned} K_{i+1} &= \begin{bmatrix} K_i & k_{i+1} \end{bmatrix} = \begin{bmatrix} Q_i & k_{i+1} \end{bmatrix} \begin{bmatrix} R_i & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} Q_i, k_{i+1} - Q_i Q_i^* k_{i+1} \end{bmatrix} \begin{bmatrix} R_i & Q^* k_{i+1} \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} Q_i & q_{i+1} \end{bmatrix} \begin{bmatrix} R_i & v_i \\ 0 & \gamma_{i+1} \end{bmatrix}, \end{aligned} \tag{5.4}$$

gdje je $\gamma_{i+1} = \|k_{i+1} - Q_i v_i\|_2$ i $v_i = Q_i^* k_{i+1}$. Ako je $\gamma_{i+1} \neq 0$ onda je $q_{i+1} = \frac{k_{i+1} - Q_i v_i}{\gamma_{i+1}}$ i $Q_{i+1} = [Q_i, q_{i+1}]$ je tražena baza. Bitno je još napomenuti da je prethodno definirana ortonormirana baza jednistveno određena (do na dijagonalnu unitarnu transformaciju).

Propozicija 5.1.4. *Ako je $Q_m = [q_1, q_2, ..., q_m]$ ortonormalna matrica takva da je $Q_i = [q_1, ..., q_i]$ baza za $K_i, i = 1, ..., m$, onda je Q_m jedinstvena do na množenje s desna dijagonalnom ortogonalnom (unitarnom) matricom.*

Dakle, glavni nedostatak metode potencija koja kreira niz $b, Ab, A^2b, ..$ je taj da se kutevi među vektorima smanjuju što pridonosi nestabilnosti same metode. Taj nedostatak

rješavamo rekurzijom kojom generiramo niz ortonormiranih baza za $\mathcal{K}_1 \subset \mathcal{K}_2$.... U svakom novom koraku $i+1$ se prethodno generiranoj bazi dodaje novi vektor q_{i+1} koji je nastao tako da prvo na q_i djelujemo s A i potom takav novonastali vektor ortgonaliziramo u odnosu na ostatak vektora iz baze. Upravo o tome govori sljedeća propozicija 5.1.5. Međutim, kod ortogonalizacije u konačnoj aritmetici postoji jedna opasnost; uvjet da su vektori u i v numerički okomiti je ekvivalentan uvjetu da $|u^T v|/\|u\|\|v\| \leq \epsilon$ gdje $\epsilon \approx 10^{-8}$ (u jednostrukoj preciznosti) ili $\epsilon \approx 10^{-16}$ (u dvostrukoj preciznosti). Zbog greške zaokruživanja postoji mogućnost da konačna baza nije ortogonalna, pogotovo kod ortogonalizacije vektora koji zatvaraju jako mali kut. U takvim slučajevim bi bilo poželjno reortogonalizirati. Dokaz zašto je dvostruka ortogonalizacija dovoljna (*twice is enough*) i više detalja o samoj proceduri se mogu naći u [12].

Propozicija 5.1.5. *Neka je $\mathcal{K}_i \subset \mathcal{K}_{i+1}$ te neka su $Q_i = [q_1, ..., q_i]$ i $Q_{i+1} = [Q_i, q_{i+1}]$ odgovarajuće baze. Tada je $\mathcal{K}_{i+1}$ generiran s $[Q_i, Aq_i]$ Dakle, za svaki i je*

$$\mathcal{K}_{i+1} = \mathcal{R}([q_1, Aq_1, Aq_2, ..., Aq_i]).$$

Ovakvom vrstom Gram-Schmidtove ortogonalizacije $v = Aq_i - Q_i(Q_i^* Aq_i)$, u konačnici dolazimo do Arnoldijevog algoritma, algoritma koji je ujedno i najpoznatiji algoritam koji koristi Krilovljeve potprostore.

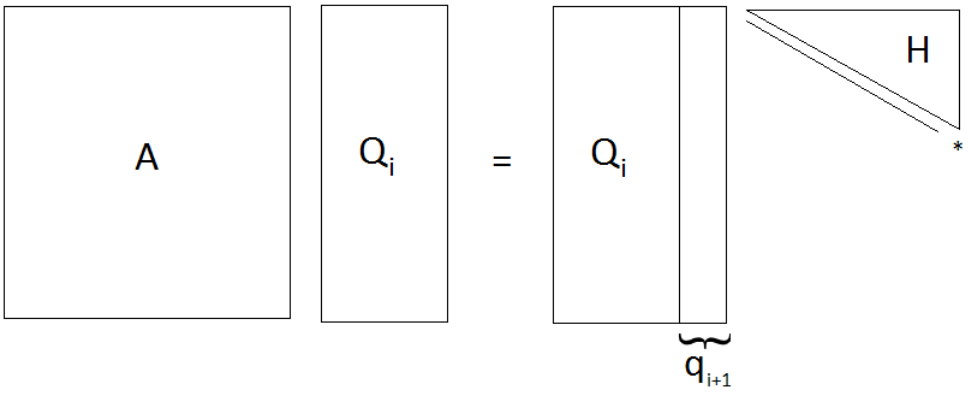
Algoritam $ARNOLDI(A, b, m)$ za zadane $A \in \mathbb{R}^{n\times n}$ i $b \in \mathbb{R}^n \backslash \{0\}$ računa ortonormirane baze $Q_i = [q_1, ..., q_i]$ za $\mathcal{K}_i$, te Hessenbergove matrice $H_{1:i,1:i}$ za koje je

$$AQ_i = Q_{i+1}H_{1:i+1,1:i}, \quad Q_i^* AQ_i = H_{1:i,1:i}, i = 1, ..., m. \tag{5.5}$$

Ako je za neki $l \leq m$, $\mathcal{K}_l = \mathcal{K}_{l+1}$, onda algoritam završava u koraku $i = l$ i vraća vrijednost l. Inače završava u zadanom m–tom koraku i stavlja $l = m$.

Algoritam 1: ARNOLDI(A, b, m)

$q_1 = \|b\|_2$;
for $i = 1, .., m$ **do**
  $v = Aq_i$;
  for $j = 1, .., i$ **do**
    $h_{ji} = q_j^* v$;
    $v = v - q_j h_{ji}$;
  end
  $h_{i+1,i} = \|v\|_2$;
  if $h_{i+1,i} = 0$ **then**
    $l = i$;
    $Q = [q_1, ..., q_l]$;
    $H = (h_{ij})_{(l+1)\times l}$; STOP
  end
  $q_{i+1} = v/h_{i+1,i}$;
end
$l = m$;
$Q = [q_1, ..., q_m]$; $H = (h_{ij})_{(m+1)\times m}$



Slika 5.2: i-ti korak Arnoldijevog algoritma

Dvije su mogućnosti završetka iteracija u Arnoldijevom algoritmu: ili je nađen invarijantan potprostor ($h_{j+1,j} = 0$) ili je završio korak m koji označava maksimalnu dopuštenu dimenziju Krilovljevog potprostora (zadana apriori).

Kao rezultat Arnoldijevog algoritma dobiva se Hessenbergova matrica H i ortonormirana baza l-tog Krilovljevog prostora. Pitanje je sada kako doći do željenih svojstvenih

vrijednosti? U (5.5) smo definirali $Q_i^* A Q_i = H_{1:i,1:i}$, a prema teoremu 2.2.3 su svojstvene vrijednosti od $H_{1:i,1:i}$ ujedno i aproksimacije od i svojstvenih vrijednosti od A. Te svojstvene vrijednosti θ_i od $H_{1:i,1:i}$ nazivamo *Ritzovim vrijednostima*, a svojstveni vektor g od $H_{1:i,1:i}$ daje *Ritzov vektor* y koji je definiran sa $y = Q_i^* g$. U konačnici, aproksimacija svojstene vrijednosti (λ, x) je Ritzov par (θ, y).

5.1.1 Restart

Kada je Arnoldijev algoritam došao do kraja iteracija (do m-tog koraka), a rezultati, tj. izračunate aproksimacije svojstvenih vrijednosti, nisu zadovoljavajući predlaže se restartati metodu. Kod restartanja je ideja iskoristiti informacije koje su prikupljene u prethodnim iteracijama te nove iteracije započeti vektorom koje će pogurati iteracije u pravom smjeru. Vratimo se sada rastavu početnog vektora b na svojstvene vektore. Pretpostavimo da svojstvenih vrijednosti ima n koje smo numerirali kao i prije $|\lambda_1| \geq |\lambda_2| \geq ... \geq |\lambda_n|$, neka su svi svojstveni vektori linearno nezavisni te neka je djelovanje matrice A na taj vektor b definiriano kao i u (5.2). Odemo li sada korak dalje i primijenimo matrični polinom $p(A) = \sum_{i=1}^{n} \alpha_i A^i$ na taj isti vektor b, tada se po uzoru na (5.3) dobiva

$$p(A)b = \sum_{i=1}^{n} \beta_i p(\lambda_i) v_i. \tag{5.6}$$

Zadatak koji nam sada prethodi je odabrati adekvatan polinom koji će ubrzati konvergenciju ka smjeru svojstvenih vektora koji su nam od interesa tj. polinom koji će naglasiti željene svojstvene vrijednosti, a ugušiti neželjene. Pošto su točne vrijednosti svojstvenih vrijednosti λ_i nepoznate (jer bismo inače bili gotovi), u praksi biramo polinome koji će poprimati velike vrijednosti na čitavom dijelu ravnine koji sadrži željene svojstvene vrijednosti, a male vrijednosti na dijelu ravnine koji sadrži neželjene svojstvene vrijednosti. Pretpostavimo sada da nas zanima r najvećih po modulu svojstvenih vrijednosti. Označimo dio ravnine koji sadrži neželjene svojstvene vrijednosti $\lambda_{r+1}, \lambda_{r+2}, .., \lambda_n$ sa E. Sada se problem traženja optimalnog polinoma svodi na problem optimizacije kojeg definiramo na sljedeći način

$$\min_{\substack{p \in \mathcal{P}_n \\ p(\lambda_r)=1}} \max_{\lambda \in E} |p(\lambda)|. \tag{5.7}$$

Među polinomima koji poprimaju najveću vrijednost u skupu E, traži se minimalni stupnja n. Zahtjevat ćemo da je traženi polinom normiran tj. da je $p(\lambda_r) = 1$. Za elipsku imamo elegantno rješenje pomoću Čebiševljevog polinoma, zbog čega definiramo elipsu E sa središtem na realnoj osi d i fokusima $d - c, d + c$ tako da sadrži sve neželjene svojstvene vrijednosti $\lambda_{r+1}, \lambda_{r+2}, .., \lambda_n$. Optimalno rješenje problema minimizacije (5.7) tada je dano

sa

$$p_n(\lambda) = \frac{T_n[(\lambda - d)/c]}{T_n[(\lambda_r - d)/c]}, \tag{5.8}$$

gdje je T_n Čebiševljev polinom reda n, koji je definiran sljedećom rekurzijom

$$\begin{aligned} T_0(\lambda) &= 1, \\ T_1(\lambda) &= \lambda, \\ T_{n+1}(\lambda) &= 2\lambda T_n(\lambda) - T_{n-1}(\lambda), n = 1, 2.. \end{aligned} \tag{5.9}$$

Detalji o Čebiševljevim iteracijama mogu se naći u [9].
Ovakvim manipulacijama dolazimo do Arnoldi-Čebiševljevog algoritma koji je definiran u nastavku. Ulazni podaci su početni vektor b ($\|b\| = 1$), prirodni broj m koji definira broj Arnoldijevih iteracija, prirodni broj n koji definira broj Čebiševljevih iteracija tj. stupanj Čebiševljevog polinom kojim ćemo modificirati početni vektor b, te prirodni broj r koji određuje broj željenih svojstvenih vrijednosti (zasada nas zanima r najvećih po modulu).

Arnoldi-Čebiševljeva metoda (b, m, n, r)

---

1. Izvrši m koraka Arnoldijevog algoritma s početnim vektorom b. Izračunaj m svojstvenih vrijednosti dobivene Hessenbergove matrice. Odaberi r najvećih svojstvenih vrijednosti $\widetilde{\lambda_1}, \widetilde{\lambda_2}, .., \widetilde{\lambda_r}$ i neka je ostatak $R = \{\widetilde{\lambda_r}, \widetilde{\lambda_{r+1}}, .., \widetilde{\lambda_m}\}$. Ako ovako dobivene aproksimacije zadovoljavaju uvjet stani.

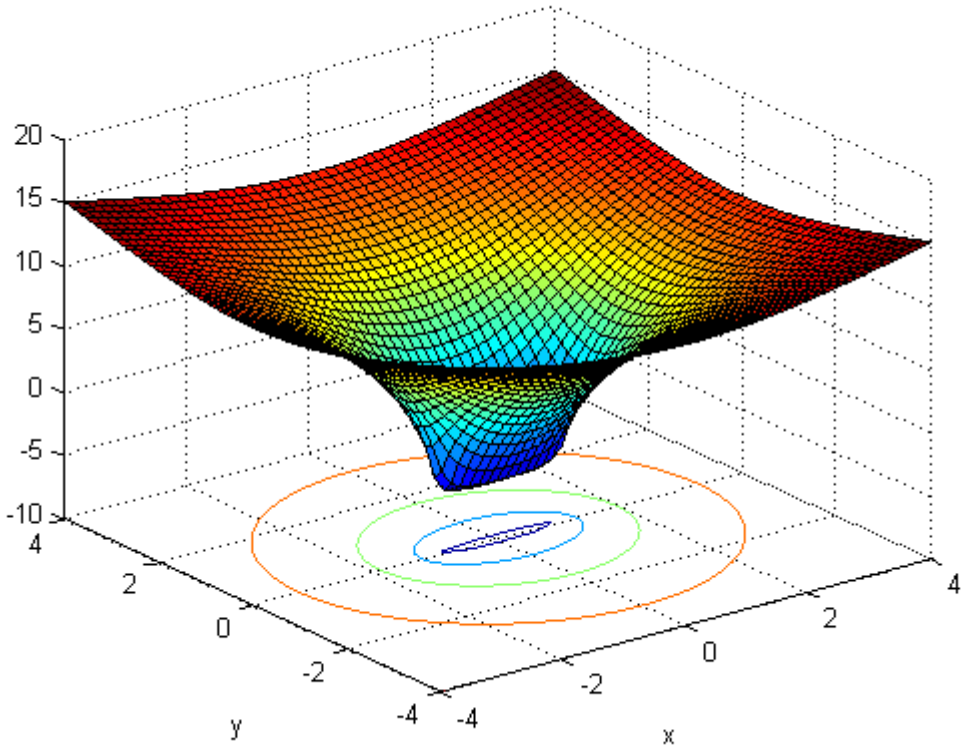
2. Koristeći R izračunaj optimalne d i c ta najbolju elipsu E tako da ona obuhvati sve vrijednosti iz R. Tada izračunaj početni vektor z_0 kao linearnu kombinaciju aproksimacije svojstvenih vektora $\widetilde{u_i}, i = 1, ..r$. Taj vektor z_0 iskoristi kao početni vektor za Čebiševljeve iteracije.

3. Izvrši n koraka Čebiševljevih iteracija ([9]) koje su bazirane na rekurziji definiranoj u (5.9), a u konačnici daju novi vektor b_n koji je jednak $p(A)b$ i definiran s (5.6). Normiraj dobiveni vektor $b = b_n/\|b_n\|$ i vrati se na prvi korak.

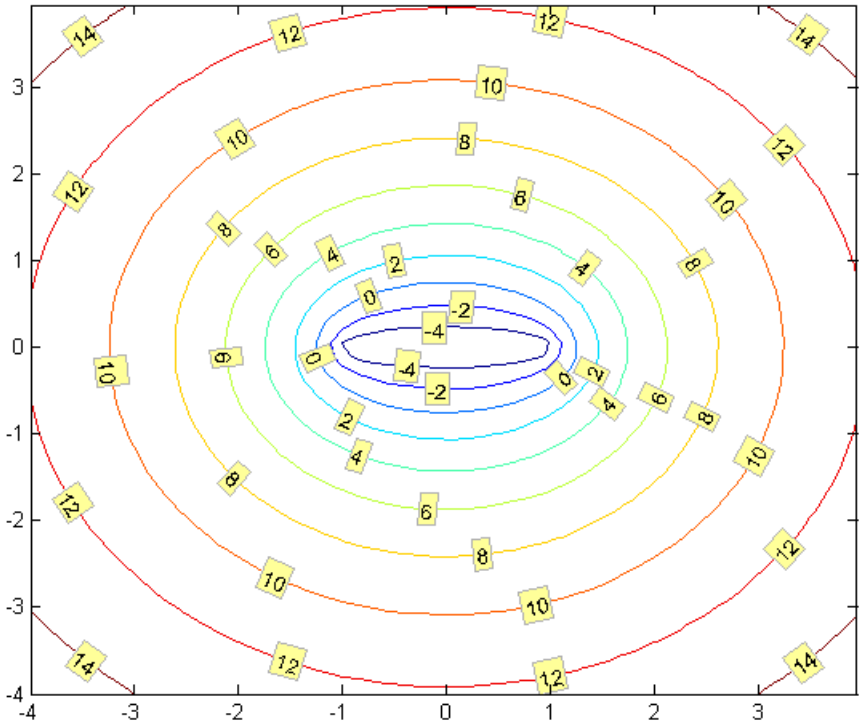
Glavna ideja ovakvog polinomijalnog filtriranje je pronaći Čebiševeljev polinom koji će vrijednosti unutar elipse E umanjiti, a vrijednosti izvan nje amplificirati. Nakon što je takav polinom pronađen, primijenjujemo ga na našu matricu A, a zatim s $p(A)$ djelujemo na početni vektor b. Sada će iteracije s novim početnim vektorom $p(A)b$ dobivenim ovakvim polinomijalnim filtriranjem brže iskonvergirati ka željenim svojstvenim vrijednostima.
Sljedeće vizualizacije će dati jasniju predodžbu zašto su baš Čebiševljevi polinomi dobar

izbor za polinomijalno filtriranje. Pošto su nam od interesa svojstvene vrijednosti velike po modulu, htjeli bismo da naš odabrani polinom ima velike vrijednosti daleko od ishodišta, a male vrijednosti blizu ishodišta. Upravo je to slučaj na slici 5.3 koja ilustrira vrijednosti Čebiševljevog polinoma stupnja 10 na elipsama s fokusima u 1 i -1, te središtem u ishodištu.

Na slici 5.4 su vrijednosti Čebiševljevog polinoma na samim elipsama (u log 10 skali), te se još jasnije vidi kako vrijednost polinoma brzo raste udaljavanjem od ishodišta.



Slika 5.3: Čebiševljev polinom stupnja 10 ($|T_{10}(z)|$)



Slika 5.4: Vrijednosti Čebiševljevog polinoma na elipsama (*log10*)

U konačnici, nakon ažuriranja početnog vektora b se ponovo poziva Arnoldijev algoritam. U slučaju da ni nakon novih m iteracija rezultati i dalje nisu zadovoljavajući, postupak se ponavlja na isti način koristeći novodobivene informacije.

Čest je slučaj da metoda nađe dovoljno dobre aproksimacije koje odgovaraju zadanim kriterijima za l svojstvenih vrijednosti (od traženih r), pa je preporučljivo smjerove tih odgovarajućih svojstvenih vektora zaključati *(eng.locking)* dok ostale svojstvene vrijednosti ne iskonvergiraju. Zaključavanje se vrši prije sljedećeg restarta kako dobre aproksimacije ne bi bile izgubljene, tj. da bi se uštedilo na vremenu. Ostatak svojstvenih vrijednosti koje ne odgovaraju kriterijima treba očistit *(eng. purging)* prije sljedećeg restarta. Za detalje vidi [8].

5.2 SOAR

SOAR metoda (*eng. Second Order ARnoldi method*) je poopćenje Arnoldijeve metode opisane u prethodnom poglavlju. Kao i kod Arnoldijeve metode, glavni korak SOAR metode je generiranje ortonormirane baze, ali ovog puta baze Krilovljevog potprostora drugog reda, nakon čega se na sličan način Rayleigh-Ritz aproksimacijama računaju željene svojstvene vrijednosti. Ova metoda je idealna kada nas zanima samo određen dio spektra i to po mogućnosti svojstvene vrijednosti kvadratičnog problema koje se nalaze na samom rubu spektra (npr. najveća ili najmanja svojstvena vrijednost po modulu). S druge strane u slučaju kad nas zanima unutrašnjost spektra (npr. frekvencije blizu nekog τ) se preporučuje korištenje Jacobi-Davidsonove metode koja pak računa jednu po jednu svojstvenu vrijednost (za razliku od Arnoldijevih metoda gdje se istovremeno računa više svojstvenih vrijednosti).

5.2.1 Linearizacija

Kao i svaka metoda za rješavanje kvadratičnog problema svojstvenih vrijednosti, SOAR započinje linearizacijom. Linearizacija koju koristi SOAR je varijanta linearizacije tipa 3, koja je definirana u poglavlju 2 (2.22). Neka su

$$F = \begin{bmatrix} -C & -K \\ I & 0 \end{bmatrix}, \quad G = \begin{bmatrix} M & 0 \\ 0 & I \end{bmatrix},$$

gdje pretpostavljamo da je M regularna. Sada možemo početni kvadratični problem svojstvenih vrijednosti

$$(\lambda^2 M + \lambda C + K)x = 0 \tag{5.10}$$

transformirati u njemu ekvivalentni generalizirani problem svojstvenih vrijednosti

$$Fx = \lambda Gx. \tag{5.11}$$

Za ovako generalizirani problem se traži odgovarajući Krilovljev potprostor, ali ovog puta drugog reda (definirano u sljedećoj sekciji). Jednom kada je izračunata ortonormirana baza za Krilovljev potprostor drugog reda vrši se projekcija originalnog kvadratičnog problema (5.10), a ne linearizacije (5.11) i na taj način je sačuvana originalna struktura.

5.2.2 Krilovljev potprostor drugog reda

U SOAR metodi glavnu ulogu igra Krylovljev potprostor drugog reda tj. njemu pripadajuća ortonormirana baza. Za generalizirani svojstveni problem

$$Ax = \lambda Bx \tag{5.12}$$

Krilovljev potprostor drugog reda definiramo na sljedeći način.

Definicija 5.2.1. *Neka su A i B kvadratne matrice reda n i neka je u≠**0** vektor dimenzije n. Krilovljev potprostor drugog reda je definiran s*

$$\mathcal{G}_n(A, B; u) = span^1\{r_0, r_1, ...r_{n-1}\},$$

gdje je niz vektora $r_0, r_1, ..r_{n-1}$ definiran na sljedeći način

$$\begin{aligned} r_0 &= u, \\ r_1 &= Ar_0, \\ r_j &= Ar_{j-1} + Br_{j-1} \end{aligned}$$

Ukoliko je $B = \mathbf{0}$ tada se radi o standardnom Krilovljevom potprostoru s kojim smo već dobro upoznati

$$\mathcal{G}_n(A, 0; u) = \mathcal{K}_n(A; u).$$

Jednom kad transformiramo kvadratični problem u njemu ekvivalentni generalizirani problem svojstvenih vrijednosti ((5.10) → (5.11)) odgovarajući *m*-ti Krylovljev potprostor lineariziranog problema ((5.10) je dan sa

$$\mathcal{K}_m(H; v) = span\{v, Hv, H^2v, ..., H^{m-1}v\}, \tag{5.13}$$

gdje je

$$H = G^{-1}F = \begin{bmatrix} -M^{-1}C & -M^{-1}K \\ I & 0 \end{bmatrix} \equiv \begin{bmatrix} A & B \\ I & 0 \end{bmatrix}. \tag{5.14}$$

Dakle, vrijedi $A = -M^{-1}C$, $B = -M^{-1}K$. Definiramo li još $v = [u^T 0]^T$ odmah je jasno da Krilovljevi vektori drugog reda $\{r_j\}$ duljine *n* iz definicije 5.2.1 i standardni Krylovljevi vektori $\{H^j v\}$ duljine $2n$ poštuju sljedeću relaciju

$$\begin{bmatrix} r_j \\ r_{j-1} \end{bmatrix} = H^j v, \tag{5.15}$$

što implicira da prostor $\mathcal{G}_m(A, B; u)$ od $\mathbb{R}^n$ daje dovoljno informacija za direktni rad s kvadratičnim oblikom, umjesto korištenja Krylovljevog potprostora $\mathcal{K}_m(H; v)$ od $\mathbb{R}^{2n}$ koji je dobiven iz lineariziranog oblika i koji je memorijski više zahtjevan. Da bismo se u to uvjerili pogledajmo još sljedeću relaciju

$$\mathcal{K}_m(H, v) \subseteq \mathcal{G}_m^2(A, B; u), \tag{5.16}$$

[1] *span je oznaka za linearnu ljusku, npr. span{a,b} predstavlja sve linearne kombinacije vektora a i b*

gdje je $\mathcal{G}_m^2(A, B; u)$ prostor generiran vektorima iz skupa

$$\left\{\begin{bmatrix} r_0 \\ 0 \end{bmatrix}, \begin{bmatrix} r_1 \\ 0 \end{bmatrix}, ..., \begin{bmatrix} r_{m-1} \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ r_0 \end{bmatrix}, .., \begin{bmatrix} 0 \\ r_{m-1} \end{bmatrix}\right\}.$$

Zbog ekvivalencije kvadratičnog problema i njegove linearizacije, relacija (5.16) pokazuje da ako se svojstveni vektor $[\lambda x^T, x^T]^T$ nalazi u prostoru $\mathcal{K}_m(H, [u^T, 0]^T)$, onda se svojstveni vektor x kvadratičnog problema (5.10) nalazi u $\mathcal{G}_m(A, B; u)$. Nadalje, zbog neprekidnosti, ako u prostoru $\mathcal{K}_m(H, [u^T, 0]^T)$ možemo naći dobru aproksimaciju za svojstveni vektor $[\lambda x^T, x^T]^T$, onda u prostoru $\mathcal{G}_m(A, B; u)$ postoji dobra aproksimacija svojstvenog vektora x.

Prednost ovakve generalizacije Krilovljevog potprostora je čuvanje strukture matrica koja je, kao što smo već napomenuli, vrlo bitna za računanje spektra i provjeru točnosti rješenja. Također, memorijski je manje zahtjevno jer $\mathcal{G}_m(A, B; u) \subset \mathbb{R}^n$, a $\mathcal{K}_m(H; v) \subset \mathbb{R}^{2n}$. U mnogim inžinjerskim problemima, n raste i do milijun, a s takvim kvadratičnim problemom je skoro nemoguće raditi, pa se uzima $m \ll n$ i generiraju Krilovljevi potprostori dimenzije m za koje se nadamo da sadrže dovoljno informacija da nađemo dobre aproksimacije rješenja.

Kao i kod Arnoldijevog algoritma, SOAR metoda Gram-Schmidtovom ortogonalizacijom traži ortonormiranu bazu $\{q_1, q_2, ..., q_m\}$ od prostora koji razapinju prvih n komponenti vektora $r_0, r_1, .., r_{m-1}$ koji pak definiraju prostor $\mathcal{G}_m(A, B; u)$. Taj postupak ortogonalizacije je opisan sljedećim algoritmom.

Algoritam 2: SOAR procedura(A, B, u, m)

$q_1 = u/\|u\|_2$; $p_1 = 0$;
for $j = 1, .., m$ **do**
  $v = Aq_j + Bp_j$;
  $s = q_j$;
  for $i = 1, .., j$ **do**
    $t_{ij} = q_i^* r$;
    $r = r - q_i t_{ij}$;
    $s = s - p_i t_{ij}$;
  end
  $t_{j+1,j} = \|r\|_2$;
  if $t_{j+1,j} = 0$ **then**
    STOP
  end
  $q_{j+1} = r/t_{j+1,j}$;
  $p_{j+1} = s/t_{j+1,j}$;
end

U prethodnom algoritmu niz vektora $\{p_j\}$ predstavlja pomoćne vektore koji generiraju prostor $\mathcal{G}_{m-1}(A, B; u)$.

Teorem 5.2.2. *Neka je* $Q_m = [q_1, ..., q_m]$ *i* $P_m = [p_1, ..., p_m]$ *i* $\widetilde{T}_m = \begin{bmatrix} T_m \\ t_{m+1,m} e_m^* \end{bmatrix} = [t_{ij}] \in \mathbb{R}^{(m+1)\times m}$. *Ako SOAR procedura ne stane prije koraka m, tada vrijedi*

$$span\{Q_m\} = \mathcal{G}_m(A, B; u)$$

i m-ti korak SOAR dekompozicije zadovoljava relaciju

$$H \begin{bmatrix} Q_m \\ P_m \end{bmatrix} = \begin{bmatrix} Q_{m+1} \\ P_{m+1} \end{bmatrix} \widetilde{T}_m,$$

gdje je $Q_{m+1} = [Q_m, q_{m+1}], P_{m+1} = [P_m, p_{m+1}]$.

Jednom kada je ortonormirana baza $\{q_1, q_2, ..., q_m\}$ poznata, Rayleigh-Ritzovom procedurom aproksimiramo početni kvadratični problem velikih dimenzija jednim kvadratičnim problemom puno manjih dimenzija.
Usporedimo sada SOAR metodu s *quadeig* metodom, koja je detaljnije opisana u prethodnom poglavlju. U *quadeig* metodi se originalni kvadratični problem transformira u linearizirani oblik što znači da se početne matrice kombiniraju u jednu matricu dvostruke dimenzije te se time gubi njihova originalna struktra (npr.simetričnost) i sve daljnje transformacije koje su potrebne za računanje svojstvenih vrijednosti primijenjuju se na tu linearizaciju. Nedostatak ovakve metode je što se gubitkom strukture ne mogu koristiti numeričke metode koje su već razvijene za rješavanje sustava posebnih struktura. S druge strane, SOAR metoda prvo kreira lineariziran oblik koji je samo pomoćni alat u računanju ortonormirane baze prostora u koji ćemo projicirati naš originalni problem. Dakle, takav projiciran kvadratični problem je i dalje kvadratični, struktura originalnih matrica je i dalje sačuvana, ali dimenzija je puno manja. Tek sada se primijenjuje neka od potpunih metoda kao što je *quadeig* ili *polyeig* kako bi se izračunale željene svojstvene vrijednosti.

Matrica Q_m dobivena SOAR procedurom sadrži m linearno nezavisnih stupaca koji predstavljaju ortonormiranu bazu od $\mathcal{G}_m(A, B; u)$. Kada projiciramo originalni kvadratični svojstveni problem na prostor $\mathcal{G}_m(A, B; u)$ dobiva se reducirani kvadratični problem

$$(\theta^2 M_m + \theta C_m + K_m)g = 0, \tag{5.17}$$

gdje je

$$M_m = Q_m^T M Q_m, \quad C_m = Q_m^T C Q_m, \quad K_m = Q_m^T K Q_m. \tag{5.18}$$

Sada kada smo sveli originalni kvadratični svojstveni problem na kvadratični problem puno manjih dimenzija, taj reducirani problem rješavamo nekom od potpunih metoda za

guste matrice. U našim eksperimentima je korištena *quadeig* metoda koja je ranije detaljno opisana, jer je u usporedbi s *polyeig*-om dala bolje aproksimacije. Svojstveni parovi (θ, g) od (5.17) definiraju Ritzove parove (θ, z) gdje je $z = Q_m g / \|Q_m g\|_2$. Te Ritzove vrijednosti su zapravo aproksimiracije svojstvenih vrijednosti početnog kvadratičnog problema definiranog u (5.10). Jednom kada imamo izračunate aproksimacije, njihova točnost se ocjenjuje kroz norme rezidualnih vektora

$$\frac{\|(\theta^2 M + \theta C + K)z\|_2}{|\theta|^2\|M\|_1 + |\theta|\|C\|_1 + \|K\|_1}. \tag{5.19}$$

U koraku redukcije kvadratičnog problema bitno je napomenuti da matrice M_m, C_m, K_m poštuju strukturu (npr. simetrija, antisimetrija,..) svojih originala M, C, K, pa će i skup aproksimacija svojstvenih vrijednosti tog reduciranog problema poštovati strukturu svojstvenih vrijednosti originalnog problema.

5.2.3 Algoritam

Dvije najbitnije komponente SOAR metode za računanje svojstvenih vrijednosti su SOAR procedura koja računa ortonormiranu bazu od $\mathcal{G}_m(A, B; u)$ (analogon Arnoldijevom algoritmu za standardni svojstveni problem) i Rayleigh-Ritzova projekcija koja slijedi odmah iza SOAR procedure te računa aproksimacije svojstvenih vrijednost. Svi koraci jedne SOAR metode dani su u nastavku, ali prije toga definiramo argumente što ih SOAR metoda prima, te šta ona vraća.

Ulazni podaci su:

- $n \times n$ matrice M, C i K koje definiraju kvadratični problem svojstvenih vrijednosti
- početni vektor $v = [u^T, \mathbf{0}]^T$
- prirodni broj $m \ll n$ kojim određujemo maksimalnu dimenziju Krilovljevog potprostora drugog reda tj. maksimalan broj iteracija SOAR procedure
- prirodni broj $r < m$ koji određuje koliko svojstvenih vrijednosti nas zanima
- *atribut* koji opisuje uvjet pod kojim se traže svojstvene vrijednosti (najveće po modulu, najmanje po modulu, blizu neke vrijednosti τ,..)

Izlazni podaci su:

- aproksimacije r svojstvenih vrijednosti koje odgovaraju *atributu*
- aproksimacije odgovarajućih svojstvenih vektora
- relativne norme rezidualnih vektora definirane u (5.19)

SOAR metoda $(M, C, K, m, r, v, atribut)$

1. (*SOAR procedura*) Izračunati ortonormiranu bazu za $\mathcal{G}_m(A, B; u)$ gdje je $A = -M^{-1}C$ i $B = -M^{-1}K$
2. Projicirati M, C i K na $\mathcal{G}_m(A, B; v)$ kako je definirano u (5.18)
3. (*Rayleigh-Ritz procedura*) Riješiti reducirani QEP (5.17) za (θ, g) i izračunati Ritzove parove (θ, z) koji predstavljaju aproksimaciju svojstvenih vrijednosti i odgovarajućih svojstvenih vektora od QEP-a (5.10)
4. Izračunati relativne norme rezidualnih vektora definiranih u (5.19) kako bi se testirala točnost Ritzovih parova (θ, z)

Komentar: Matrice A i B se ne računaju eksplicitno, nego se u SOAR proceduri (vidi algoritam 2) gdje računamo vektore Aq_j i Bp_j, umjesto $-M^{-1}Cq_j$ prvo djelujemo sa C na q_j, a potom rješavamo sustav $-Mx = Cq_j$. Vektor Bp_j računamo na analogan način. Ovakvo implicitno računanje se koristi zbog efikasnosti i veće točnosti.

5.2.4 Deflacija i prekid

Prisjetimo se klasičnog Arnoldijevog algoritma za računanje Krilovljevog potprostora; iteracije prestaju onog trenutka kad smo našli invarijantan potprostor tj. $H\mathcal{K}_j \subset \mathcal{K}_j$, no kod generiranja Krilovljevog potprostora drugog reda treba biti oprezan jer postoje dvije mogućnosti prestanka iteracija; *prekid* i *deflacija*.

Definicija 5.2.3. *Ako su* $\{r_i\}, i = 0, 1...j$ *linearno zavisni, ali* $\{[r_i^T, r_{i-1}^T]\}, i = 0, ...j$ *gdje je* $r_{-1} = 0$ *linearno nezavisni, takva situacija se naziva deflacija; ako su oboje* $\{r_i\}$ *i* $\{[r_i^T, r_{i-1}^T]\}$ *linearno zavisni, takvu situaciju nazivamo prekid.*

Prethodna definicija govori o tome da u slučaju deflacije vrijedi

$$\mathcal{G}_{j+1}(A, B; u) = \mathcal{G}_j(A, B; u)$$

, ali $\mathcal{K}_{j+1}(H, v) \neq \mathcal{K}_j(H, v)$ iz čega slijedi da kada dođe do deflacije u koraku $j < m$, $\mathcal{K}_j$ nije dobar prostor za egzaktne svojstvene vektore od H, pa samim time ni $\mathcal{G}_j(A, B; u)$ nije dobar prostor za traženje rješenja (zbog relacije (5.15)). Iz tog razloga nakon deflacije treba nastaviti s iteracijama. Ako se deflacija dogodi u j-tom koraku SOAR procedure definirane algoritmom 2 tada se q_{j+1} postavlja na $\mathbf{0}$, a $t_{j+1,j}$ na 1. Posljedica ovakve modifikacije je

da će matrica Q_m čiji su stupci zapravo vektori ortonormirane baze prostora $\mathcal{G}_m(A, B; u)$ sadržavati $\mathbf{0}$, ali to se lako rješava tako da se stupci matrice Q_m filtriraju i ostave samo oni koji odgovaraju koracima u kojima se nije dogodila deflacija.
S druge strane, kada se prekid dogodi iteracije se zaustavljaju jer $\mathcal{G}_j(A, B; u)$ predstavlja dovoljno dobar potprostor za traženje rješenja. Kao i prije, stupci matrice Q_m definiraju ortonormiranu bazu tog prostora. Više detalja o prekidu i deflaciji se može naći u [1].

5.2.5 Restart

Kako je to slučaj i kod običnog Arnoldijevog algoritma, postoji mogućnost da u SOAR metodi iteracije ne iskonvergiraju, tj. da nije pronađen invarijantan potprostor. Također, kako m raste ta metoda postaje skuplja i nepraktična zbog zauzimanja velike količine memorije. Ta dva razloga dovoljna su motivacija za restartati metodu. S druge strane, nakon m iteracija obično već imamo nekakvo saznanje o smjerovima svojstvenih vektora i njihovim svojstvenim vrijednostima, te bismo voljeli te informacije iskoristiti prilikom ponovnog pozivanja novih iteracija. Te informacije se implementiraju ažuriranjem početnih vektora q_1 i p_1 u nove vektore q_1^+ i p_1^+ s kojima kreću nove iteracije definirane algoritmom 2. Međutim, primjetimo odmah da se restart neće moći direktno primijeniti na SOAR proceduru s obzirom da ažurirani p_1^+ više neće biti $\mathbf{0}$. Kako bi se riješio taj nedostatak modificira se SOAR procedura na način da se umjesto inicijalizacije $p_1 = \mathbf{0}$ zadaje nenul vektor p_1 kao što je to slučaj i za q_1. Takvu modificiranu SOAR proceduru nazivamo GSOAR (*eng.Generalized Second Order Arnoldi*), a prostor koji razapinju vektori dobiveni takvom procedurom se naziva generalizirani Krilovljev potprostor drugog reda te se definira na sljedeći način

Definicija 5.2.4. *Neka su A i B kvadratne matrice reda n i neka je $u_1, u_2 \in \mathbb{C}^n \backslash \{0\}$ vektor dimenzije n. Generalizirani Krilovljev potprostor drugog reda je definiran sa*

$$\mathcal{G}_n(A, B; u_1, u_2) = span\{r_0, r_1, ...r_{n-1}\},$$

gdje je niz vektora $r_0, r_1, ..r_{n-1}$ definiran na sljedeći način

$$\begin{aligned} r_0 &= u_1, \\ r_1 &= Ar_0 + Bu_2, \quad . \\ r_j &= Ar_{j-1} + Br_{j-1} \end{aligned}$$

Primjetimo da u slučaju kada je $u_2 = \mathbf{0}$ radi se o standardnom Krilovljevom potprostoru drugog reda defiranog u 5.2.1 tj. $\mathcal{G}_n(A, B; u_1, \mathbf{0}) = \mathcal{G}_n(A, B; u_1)$. Kako je to već definirano u algoritmu 2, naši vektori q_1 i p_1 koji će kasnije biti ažurirani su samo normirani vektori u_1 i u_2, respektivno. Dakle, nakon ažuriranja početnih vektora poziva se ponovo

GSOAR procedura koja kreira novi generalizirani Krilovljev potprostor drugog reda, potprostor koji je bogatiji s informacijama o smjeru željenih svojstvenih vektora. Sada preostaje samo naći način kako efikasno ažurirati vektore q_1 i p_1 tako da amplificiraju komponente željenih svojstvenih vektora i simultano guše komponente neželjenih. Za razliku od Arnoldi-Čebiševljeve metode u kojoj se ažuriranje vrši eksplicitno, metode Krilovljevih potprostora drugog reda preferiraju implicitni restart. Ideja implicitnog restarta je primijeniti pomake $\mu_1, .., \mu_p$ na dobivenu Hessenbergovu matricu T_m, tj. izvršiti p implicitno pomaknutih QR iteracija na T_m definiranu u teoremu 5.2.2

$$(T_m - \mu_1 I) \cdots (T_m - \mu_p I) = V_m R, \tag{5.20}$$

gdje je V_m $m \times m$ ortogonalna matrica i R gornje trokutasta. Sada je najteži zadatak odabrati najbolje pomake $\mu_1, \mu_2, ..., \mu_p$ tako da se ažuriranjem smjerovi neželjenih svojstvenih vektora guše, a da smjerovi željenih svojstvenih vektora dođu više do izražaja. Idealno bi bilo za pomake $\mu_1, \mu_2, ..., \mu_p$ odaberati baš p neželjenih svojstvenih vrijednosti, jer se onda u potpunosti uguše smjerovi tih svojstvenih vektora, no za razliku od idealne teorije u praksi nemamo egzaktne vrijednosti već samo aproksimacije, koje nerijetko nisu ni zadovoljavajuće. Jedan od rješenja je da se uzmu aproksimacije svojstvenih vrijednosti koje se nalaze što dalje od željenih svojstvenih vrijednosti. Više o izboru pomaka i njihovoj prirodi se može naći u [6],[8]. Ukoliko prethodno nije došlo do deflacije, nakon pomaka se primjenjuje GSOAR procedura još jednom. Restartanje u slučaj kada se dogodila deflacija je malo kompliciranije i ono je detaljno opisano u [6].

Samo ažuriranje početnih vektora možemo eksplicitno definirati sljedećom formulom.

Teorem 5.2.5. *Vrijedi*

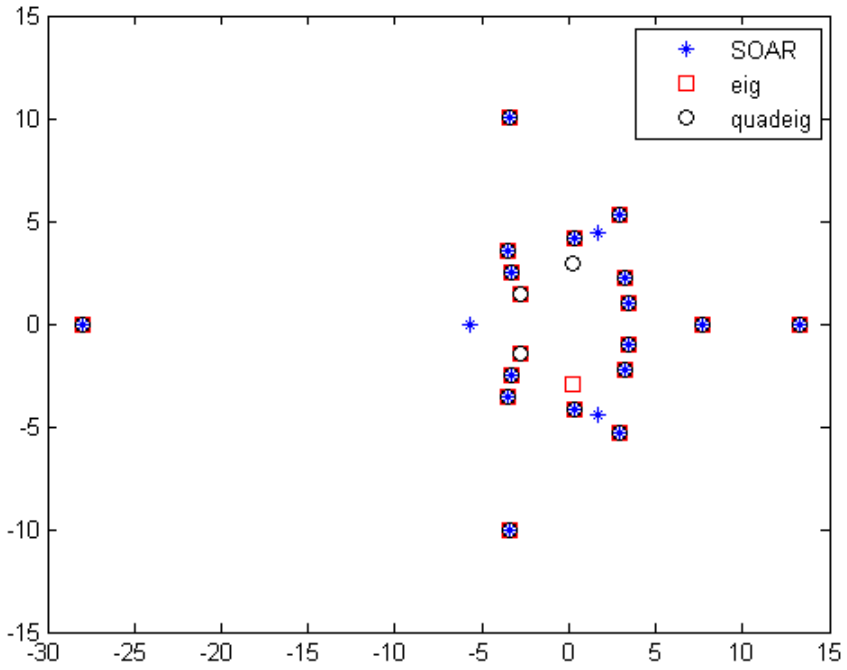
$$\begin{bmatrix} q_1^+ \\ p_1^+ \end{bmatrix} = \frac{1}{\tau} \psi(H) \begin{bmatrix} q_1 \\ p_1 \end{bmatrix}, \tag{5.21}$$

gdje je $\psi(\lambda) = \prod_{j=1}^{p} (\lambda - \mu_j)$ *i* τ *faktor normiranja.*

5.2.6 Eksperimenti

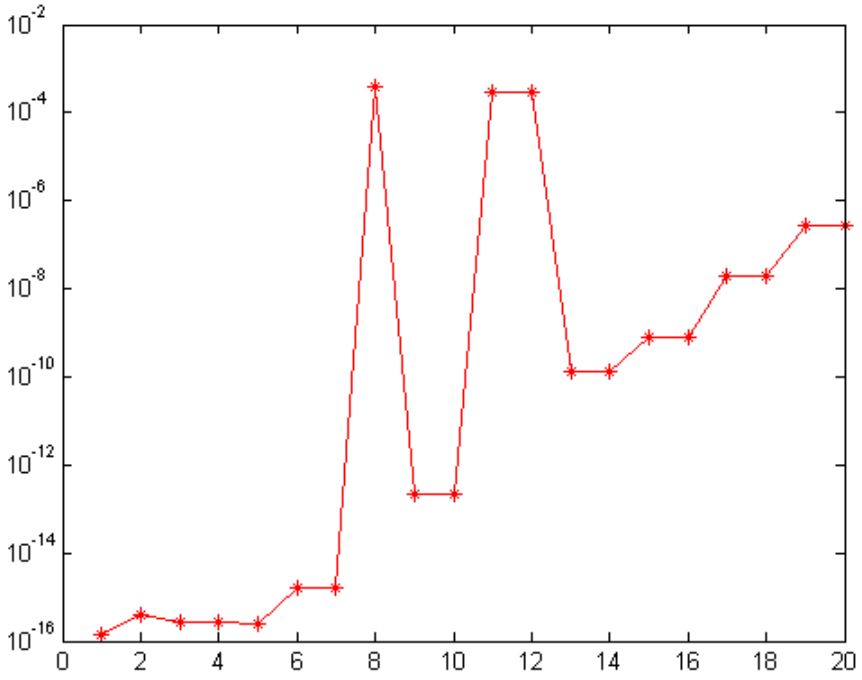
U sljedećim primjerima ćemo usporediti Arnoldijeve metode s potpunom metodom *quadeig* iz poglavlja 4. Jedna Arnoldijeva metoda je SOAR, koja je prethodno detaljno opisana, a druga je Arnoldijeva metoda za linearizirani problem definiran sa (5.11), koja je implicitno implementirana u MATLAB-ovoj funkciji *eigs*. U sljedećim primjerima su egzaktna rješenja nepoznata (kao što je obično slučaj u praksi) te ćemo kao referetne vrijednosti uzeti svojstvene vrijednosti izračunate direktnom metodom *quadeig*. Što se tiče SOAR metode; u prvom koraku je korištena osnovna SOAR procedura koja ne provjerava deflaciju ni prekid, a u trećem koraku reducirani QEP riješavamo metodom *quadeig*.

Primjer 1. Neka su M, C i K nasumične matrice veličine 200×200 ($n = 200$). Elementi tih matrica su normalno distribuirani sa srednjom vrijednosti 0, standardnom devijacijom 1. Za odgovarajući kvadratični problem tražimo $r = 20$ dominantnih svojstvenih vrijednosti (*atribut*=najveće po modulu). Aproksimacije tih svojstvenih vrijednosti smo tražili u Krilovljevom potprostoru dimenzije $m = 50$. Rezultati SOAR metode, *eigs* metode i *quadeig* metode su ilustrirani na slici 5.5



Slika 5.5: Aproksimacije svojstvenih vrijednosti

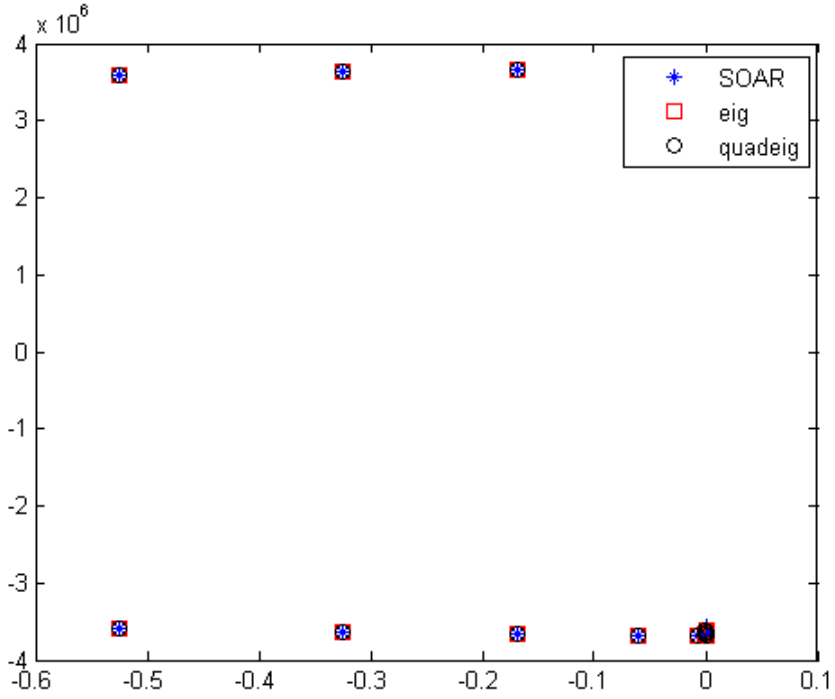
Slika 5.5 potvrđuje tvrdnju sa početka ovog poglavlja, a to je da Arnoldije metode pogađaju puno bolje svojstvene vrijednosti koje se nalaze na samom rubu spektra nego iz unutrašnjosti spektra.



Slika 5.6: Relativne rezidualne norme od aproksimiranih svojstvenih vektora

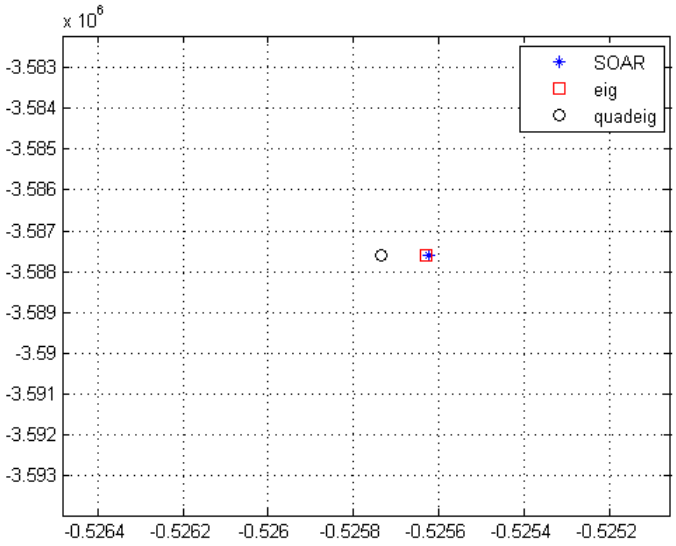
Primjetimo da je prvih 7 svojstvenih vrijednosti (njih 7 najvećih po modulu) iskonvergiralo, što se poklapa sa slikom 5.5 gdje se vidi da su svojstvene vrijednosti sa ruba spektra pogođene tj. u ovom primjeru su to upravo one svojstvene vrijednosti najveće po modulu.

Primjer 2. Primjenimo SOAR metodu na već poznati primjer prigušene grede koji smo definirali u prethodnom poglavlju. Matrice M,C i K nalazimo u NLEVP kolekciji. a svaka je dimenzije 200×200 ($n = 200$). U želji da izračunamo $r = 20$ dominatnih svojstvenih vrijednosti primjenjujemo SOAR metodu s $m = 100$ iteracija koja računa aproksimacije svojstvenih vrijednosti tj. tražimo aproksimacije tih 20 dominantnih svojstvenih vrijednosti u Krilovljevom potprostoru dimenzije 100. Aproksimacije svojstvenih vrijednosti su prikazane na slici 5.7

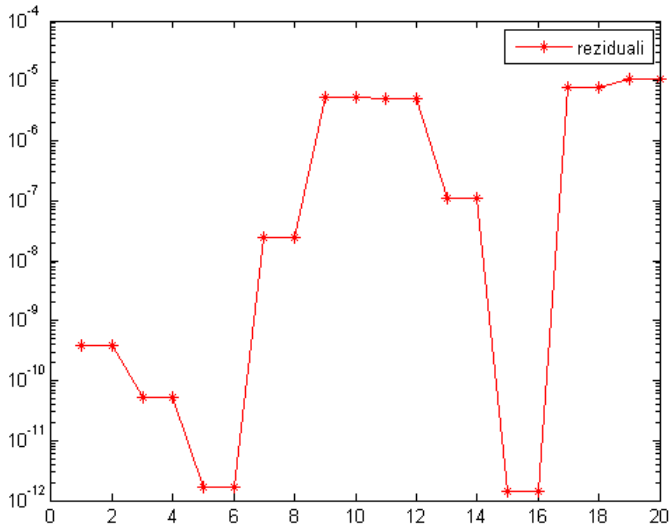


Slika 5.7: Aproksimacije svojstvenih vrijednosti

Iako na prvi pogled izgleda da aproksimacije različitih metoda daju iste rezultate, povećamo li sitaciju oko dominantne svojstvene vrijednosti, primjetit ćemo da se različite aproksimacije različitih metoda polagano razilaze. Ta situacija je prikazana na sljedećoj slici 5.8. *Eig* i *SOAR* se lagano počinju udaljavati od referentne vrijednosti koju smo izračunali *quadeig* metodom.



Slika 5.8: Aproksimacije dominantne svojstvene vrijednosti



Slika 5.9: Relativne rezidualne norme od aproksimiranih svojstvenih vektora

Poglavlje 6

Zaključak

Inžinjeri modeliranjem ponašanja raznih nebodera,mostova i sl. dolaze do kvadratičnih problema svojstvenih vrijednosti s matricama M, C i K iznimno velikih dimenzija, a traže samo par svojstvenih vrijednosti koje odgovaraju nekom atributu (npr. kod istraživanja stabilnosti sustava od interesa su one svojstvene vrijednosti kojima je realni dio blizu 0). Jedan od načina da se takvi problemi riješe je primijeniti *quadeig* metodu definiranu u poglavlju 4. koja daje sve svojstvene vrijednosti zadanog kvadratičnog problem. Jednom kad su one poznate možemo odabrati samo onih par koje odgovaraju nekom unaprijed zadanom atributu. Takav način bi bio nepraktičan, a ponekad i nemoguć, jer bi račun bio iznimno skup, što vremenski, što memorijski. Upravo zbog tog nedostatka efikasnosti okrećemo se iterativnim metodama. Metode koje se baziraju na Krilovljevim prostorima su samo neke od tih iterativnih metoda. SOAR metoda opisana u poglavlju 5. reducira početni kvadratični problem, a samim time i dimenziju prostora u kojem tražimo rješenje, čime možemo puno uštediti na memoriji, pogotovo kada je m (dimenzija Krilovljevog potprostora u kojem tražimo rješenje) puno manji od n. Naravno, nije dobro ni odabrati premali m jer time smanjujemo vjerojatnost dobre aproksimacije, u smislu da je s malim m šansa da smo pogodili baš m-dimenzionalan prostor koji sadrži tražene svojstvene vektore manja. Taj problem se rješava restartanjem metode u kombinaciji sa čišćenjem i zaključavanjem svojstvenih vrijednosti([8],[6]). S druge strane, nedostatak Arnoldijevih metoda je taj da sporo konveriraju pogotovo kada su svojstvene vrijednosti jako blizu jedna drugoj. Također, Arnoldijeve aproksimacije unutrašnosti spektra daju loše rezultate (vidi sliku 5.3). Pitanje unutrašnjosti spektra možemo riješiti s varijantom pomaka i invertiranja. Iterativna metoda koja je alternativa Arnoldijevim metodama je Jacobi-Davidsonova metoda koja daje najbližu svojstvenu vrijednost prethodno zadanom τ, te odgovarajući svojstveni vektor. Jacobi-Davidsonova metoda za razliku od Arnoldijeve računa jednu po jednu svojstvenu vrijednost zbog čega je puno sporija. Zasada ne postoji najbolja metoda, jer različiti kvadratični problemi svojstvenih vrijednsoti imaju različita svojstva, te zahtjevaju drugačiji

tretman. Taj tretman uključuje različite metode, različite linearizacije, različita skaliranja i slično. Uz strukturu samog problema bitno je i dobro definirati kakve svojstvene vrijednosti i koliko njih se traži jer ako nas zanimaju sve svojstvene vrijednosti onda bi *quadeig* metoda bila pravi izbor, u suprotnom, ako nas zanima samo nekoliko svojstvenih vrijednosti preporučuju se iterativne metode poput SOAR metode. Dakle, svakom kvadratičnom problemu svojstvenih vrijednosti bi trebalo pristupiti individualno i odabrati metodu koja baš njemu odgovara.

Bibliografija

[1] Zh. Bai i Y. Su, *SOAR: A second-order Arnoldi method for the solution of the quadratic eigenvalue problem*, (2005), 89–93.

[2] Durán R.G. Rodríguez R. Solomin J. Bermúdez, A., *A finite element method to compute damped vibration modes in dissipative acoustics*, (2000).

[3] T. Betcke, N.J Higham, V. Mehrmann, C. Schröder i F. Tisseur, *NLEVP: A Collection of Nonlinear Eigenvalue Problems*, (2011).

[4] Z. Drmač, *Krilovljevi potprostori*, Numerička analiza (2010).

[5] S. Hammarling, C.J Munro i F. Tisseur, *An Algorithm for the complete Solution of Quadratic Eigenvalue Problems*, The University of Manchester MIMS report (2013).

[6] Zh. Jia i Y. Sun, *Implicitly Restarted Generalized Second-order Arnoldi type algorithms for the quadratic eigenvalue problem*, (2000).

[7] M. Jurak, *Prostori konačnih elemenata*, Numeričko rješavanje parcijalnih jednadžbi (2015).

[8] R.B. Lehoucq i D.C Sorensen, *Deflation Techniques within an Implicitly Re-started Arnoldi Iteration*, (1995).

[9] Y. Saad, *Chebyshev acceleration techniques for solving nonsymmetric eigenvalue problems*, Mathematics of Computation (1984).

[10] F. Tisseur, *Backward error and condition of polynomial eigenvalue problems*, The University of Manchester MIMS report (2000).

[11] F. Tisseur i K. Meerbergen, *The quadaticc eigenvalue problem*, The University of Manchester MIMS report (2005).

[12] David S. Watkins, *The matrix eigenvalue problem: GR and Krylov subspace methods*, 1., Society for Industrial and Applied Mathematics, 2007.

Sažetak

U ovom radu smo se bavili kvadratičnim problemom svojstvenih vrijednosti čiji je zadatak naći skalare $\lambda \in \mathbb{C}$ i nenul vektore $x \in \mathbb{C}^n$ tako da vrijedi

$$(M\lambda^2 + C\lambda + K)x = 0,$$

gdje su $M, C, K \in \mathbb{C}^{n\times n}$. Ovako definiran problem ima $2n$ svojstvenih vrijednosti kao rješenje, te svojstvene vrijednosti mogu biti i konačne i beskonačne.
Prezentirali smo dvije različite numeričke metode koje rješavaju spomenuti problem; metoda *quadeig* i SOAR *(eng. Second Order ARnoldi method)* metoda. Metoda *quadeig* pripada klasi potpunih (direktnih) metoda, što znači da računa svih $2n$ svojstvenih vrijednosti, a pogodna je za korištenje kada je dimenzija početnog problema relativno mala. Ta metoda se detaljnije bavi skaliranjem zadanih matrica M,C i K početnog problema, kao i problemom beskonačnih svojstvenih vrijednosti. Numeričkim eksperimentima smo usporedili *quadeig* s metodom *polyeig* koja je implementirana u paketu MATLAB-a. S druge strane, prezentirali smo SOAR metodu koja pripada klasi iterativnih metoda koje se primijenjuju kada je potrebno izračunati samo dio spektra koji je od interesa. Iterativne metode su iznimno efikasne u slučaju velikih dimenzija početnog problema, kada je primjena potpunih metoda iznimno skupa (memorijski i vremenski) ili čak nemoguća. SOAR metoda, koja je u ovom radu predložena kao efikasno rješenje kvadratičnih problema, je poopćenje Arnoldijeve metode za standardni svojstveni problem, a osnovna ideja joj je pronalazak pogodno odabranog potprostora koji će dobro aproksimirati prostor koji razapinju traženi svojstveni vektori. Jedan takav pogodan potprostor na kojem je bazirana SOAR metoda je Krilovljev potprostor drugog reda. Glavni problem iterativnih metoda je pitanje konvergencije. Na ubrzavanju konvergencije se i dalje aktivno radi. Svi numerički eksperimenti u ovom radu su napravljeni u MATLABu.

Summary

In this thesis we study the quadratic eigenvalue problem (QEP), that is for given M, C, $K \in \mathbb{C}^{n\times n}$, task is to compute an eigenvalue $\lambda \in \mathbb{C}$ and an eigenvector $x \in \mathbb{C}^n$, $x \neq 0$

$$(M\lambda^2 + C\lambda + K)x = 0.$$

This QEP has $2n$ eigenvaluess, some of them can even be infinite.

We present two different numerical methods for solving QEP; *quadeig* method and SOAR *(Second Order ARnoldi)* method. *Quadeig* method is a complete (direct) method, meaning that it computes all $2n$ eigenvalues, and we recommend applying this method when the dimension of the initial QEP is not too big. One of the improvements that *quadeig* offers is scaling of the given matrices M,C and K. The way that *quadeig* method deals with infinite eigenvalues and zero eigenvalues improves the accuracy of the approximations for the finite eigenvalues. Our numerical experiments compare *quadeig* method with MATLAB's built-in function *polyeig* that solves the same QEP problem. Also, we investigate the SOAR method which belongs to the class of iterative methods. Unlike the complete methods, iterative methods are applied when we are interested only in a subset of the spectrum (only few eigenvalues). Those iterative methods are effective way of dealing with the problem of large dimension. Applying a complete method to such big problems would cause great memory cost and immense time consumption. SOAR method, that we suggest in this thesis for effectively solving QEP, is a generalization od the Arnoldi method for the standard eigenvalue problem and the basic idea is to find suitable subspace that is very close to the subspace that is spanned by the wanted eigenvectors. One of the good subspaces which is also the basis for the SOAR method is a second order Krylov subspace. The main problem of iterative methods is the convergence question. Enhancing the convergence is still one of the main topics in many researches and improvements are being made. All the numerical experiments in this thesis are made in MATLAB.

Životopis

Maša Avakumović rodila se 22.3.1992. u Zagrebu. Pohađala je Osnovnu školu Miroslava Krleže u Zagrebu. Po završetku osnovne škole upisala je XVI. gimnaziju u Zagrebu. 2010. upisuje Prirodoslovno-matematički fakultet na Sveučilištu u Zagrebu. Nakon završenog prediplomskog studija matematike, 2013. upisuje diplomski studij Primijenjena matematika. 2015 odlazi studirati jedan semestar na belgijskom Sveučilištu u Ghentu gdje se upoznala s modernim metodama analize podataka, što ju je pripremilo za stručnu praksu u Londonu gdje je radila kao *Data Scientist* u MyDrive Solutions krajem 2015 i početkom 2016. U Zagreb se vraća 2016. gdje završava diplomski studij Primijenjena matematika.