SVEUČILIŠTE U ZAGREBU

PRIRODOSLOVNO-MATEMATIČKI FAKULTET – MATEMATIČKI ODJEL

Zagreb, Bijenička cesta 30

Iterativne metode

Autori: Vjeran Hari, Sanja Singer, Saša Singer

U Zagrebu, 2007.

Poglavlje 1

ITERATIVNE METODE ZA LINEARNE SUSTAVE

1.1 Teorija perturbacije linearnih sustava

U ovom odjeljku prezentirat ćemo rezultate klasične teorije perturbacije po normi, linearnih sustava, ali i modernije perturbacije po komponentama. Pitanje na koje odgovaraju takve teorije perturbacije je koliko se rješenje linearnog sustava (1.1) promijeni po normi/po komponentama ako se po normi/po komponentama malo promijene A i b.

Da bismo izbjegli pisanje indeksa normi, sve norme koje ćemo u ovom poglavlju koristiti bit će konzistentne matrične norme i njima odgovarajuće vektorske norme (na primjer, Hölderove p-norme).

Pretpostavimo da, umjesto sustava

$$Ax = b, \tag{1.1}$$

egzaktno rješavamo sustav

$$(A + \Delta A)(x + \Delta x) = b + \Delta b, \tag{1.2}$$

Oduzmimo jednadžbu (1.1) od jednadžbe (1.2),

$$\Delta A\, x + (A + \Delta A)\Delta x = \Delta b. \tag{1.3}$$

Vektor $\hat{x} = x + \Delta x$ će u primjenama ove teorije najčešće biti izračunato rješenje. Stoga će ono kao izlazni rezultat nakon računanja na stroju biti poznato, pa je zgodno uvesti ga u račun smetnje. Jednadžbu (1.3) možemo zapisati kao $A\Delta x + \Delta A\hat{x} = \Delta b$ ili

$$\Delta x = A^{-1}(-\Delta A\hat{x} + \Delta b). \tag{1.4}$$

Uzimanjem normi dobivamo

$$\|\Delta x\| \leq \|A^{-1}\|(\|\Delta A\|\,\|\hat{x}\| + \|\Delta b\|). \tag{1.5}$$

Jednadžbu (1.5) možemo zapisati kao

$$\frac{\|\Delta x\|}{\|\hat{x}\|} \leq \|A^{-1}\|\|A\|\left(\frac{\|\Delta A\|}{\|A\|} + \frac{\|\Delta b\|}{\|A\|\|\hat{x}\|}\right). \tag{1.6}$$

Veličinu $\kappa(A) = \|A^{-1}\|\|A\|$ nazivamo *broj uvjetovanosti* matrice A jer mjeri relativnu promjenu rješenja po normi $\|\Delta x\|/\|\hat{x}\|$ pomoću relativnih promjena u polaznim podacima. Veliki $\kappa(A)$ će upućivati da male promjene u matrici A i/ili vektoru b mogu dati veliku relativnu promjenu u rješenju sustava. Važno je napomenuti da se nejednakost u (1.6) dostiže za određene ΔA i Δb, pa $\kappa(A)$ nije tek neka gornja ograda.

Ocjena (1.6) je pogodna u praksi jer koristi $\hat{x}$ koji je na raspolaganju za razliku od x koji je obično nepoznat. Matematički gledano, izraz $\|\Delta x\|/\|\hat{x}\|$ nije "čist" jer se Δx nalazi i u brojniku i u nazivniku (kao dio od $\hat{x}$). Morali bi zapravo ocijeniti $\|\Delta x\|/\|x\|$. Za to će nam trebati sljedeći pomoćni rezultat

Lema 1.1 *Neka za matricu S i matričnu normu $\|\cdot\|$ induciranu vektorskom normom, vrijedi $\|S\| < 1$. Tada je $I - S$ invertibilna i vrijedi*

$$(I-S)^{-1} = \sum_{k=1}^{\infty} S^k, \qquad \textit{pa je} \quad \|(I-S)^{-1}\| \leq \frac{1}{1-\|S\|}\,. \tag{1.7}$$

Dokaz: U izrazu (1.7) pojavljuje se suma reda $\sum_{k=1}^{\infty} S^k$. Za red matrica (ili vektora) kažemo da je konvergentan ako je konvergentan svaki red brojeva koji odgovara poziciji bilo kojeg elementa u matrici (vektoru). Uočimo da za element x_{ij} matrice X i za svaku vektorsku normu u $\mathbf{R}^n$, za koju u slučaju $n = 1$ vrijedi $|\alpha| = \|\alpha\|$, $\alpha \in \mathbf{R}$, imamo

$$|x_{ij}| = |e_i^T X e_j| = \|e_i^T(Xe_j)\| \leq \|e_i^T\|\|Xe_j\| \leq \|e_i^T\|\|X\|\|e_j\| = c_{ij}\|X\|, \quad 1 \leq i,j \leq n,$$

gdje je $c_{ij} = \|e_i^T\|\|e_j\|$, $1 \leq i,j \leq n$. Npr., za Hölderovu vektorsku p-normu $\|x\|_p = [\sum_{j=1}^n |x_j|^p]^{1/p}$ vrijedi $c_{ij} = 1$. Za bilo koju vektorsku normu će (zbog ekvivalentnosti normi u $\mathbf{R}^n$) vrijediti $c_{ij} > 0$ za sve i, j.

Kako je $\|S\| < 1$, red $\sum_{k=1}^{\infty} S^k$ konvergira jer je za svako i, j red $\sum_{k=1}^{\infty}(S^k)_{ij}$ majoriziran konvergentnim geometrijskim redom $c_{ij}\sum_{k=1}^{\infty}\|S\|^k$. Najme, vrijedi $|(S^k)_{ij}| \leq c_{ij}\|S^k\| \leq c_{ij}\|S\|^k$. Pritom smo iskoristili da je svaka inducirana matrična norma konzistentna, tj. vrijedi $\|AB\| \leq \|A\|\,\|B\|$.

Dakle je red matrica $\sum_{k=1}^{\infty} S^k$ konvergentan. Označimo sa Z limes tog reda.

Neka je S_r r-ta parcijalna suma tog konvergentnog reda. Vrijedi

$$\begin{aligned}(I-S)S_r &= (I-S)(I+S+S^2+\cdots+S^r) = I+S+S^2+\cdots+S^r - \\ &\quad (S+S^2+\cdots+S^r+S^{r+1}) = I - S^{r+1} \\ S_r(I-S) &= (I+S+S^2+\cdots+S^r)(I-S) = I+S+S^2+\cdots+S^r - \\ &\quad (S+S^2+\cdots+S^r+S^{r+1}) = I - S^{r+1}, paje\end{aligned}$$

Jer $\|S^r\| \leq \|S\|^r \downarrow 0$ kad $r \to \infty$, imamo $\lim_{r\to\infty} S^r = 0$

$$\begin{aligned}(I-S)S_r &\to I \quad \text{kad } r \to \infty\,,\\ S_r(I-S) &\to I \quad \text{kad } r \to \infty.\end{aligned}$$

U limesu S_r prelazi u Z, množenje matrica je neprekidna funkcija faktora, pa vrijedi

$$(I-S)Z = I \qquad \text{i} \qquad Z(I-S) = I. \tag{1.8}$$

Matrica $I-S$ je regularna. Doista, kad bi bila singularna, postojao bi vektor $x \neq 0$, takav da je $(I-S)x = 0$. Sada bi $x = Sx$ i kozistentnost normi dala $\|x\| = \|Sx\| \leq \|S\|\|x\|$. To je isto što i $(1-\|S\|)\|x\| \leq 0$. No, $\|x\| > 0$, pa bi bilo $1-\|S\| \leq 0$ odnosno $\|S\| \geq 1$. Dobiveno proturječje s pretpostavkom $\|S\| < 1$ pokazuje da $I-S$ ne može biti singularna. Sada iz relacije (1.8) slijedi $Z = (I-S)^{-1}$.

Druga tvrdnja slijedi iz činjenice da je za svako $r \geq 0$

$$\begin{aligned}\|S_r\| &= \|I+S+S^2+\cdots+S^r\| \leq \|I\|+\|S\|+\|S^2\|+\cdots+\|S^r\|\\ &\leq 1+\|S\|+\|S\|^2+\cdots+\|S\|^r = \frac{1-\|S\|^{r+1}}{1-\|S\|}.\end{aligned}$$

Ovdje je $\|I\| = 1$ jer se radi o induciranoj matričnoj normi. Najme,

$$\|I\| = \max_{\|x\|=1}\|Ix\| = \max_{\|x\|=1}\|x\| = 1.$$

Puštajući $r \to \infty$ i koristeći $\|S\|^{r+1} \downarrow 0$, neprekidnost norme i $S_r \to Z$, odmah dobivamo

$$\|Z\| \leq \frac{1}{1-\|S\|}.$$

■

Vratimo se na jednadžbu (1.3), $\Delta A\, x + (A+\Delta A)\Delta x = \Delta b$. Ako ju riješimo po Δx dobijemo

$$\begin{aligned}\Delta x &= (A+\Delta A)^{-1}(-\Delta A x + \Delta b)\\ &= \left[A(I+A^{-1}\Delta A)\right]^{-1}(-\Delta A x + \Delta b)\\ &= \left(I+A^{-1}\Delta A\right)^{-1}A^{-1}(-\Delta A x + \Delta b)\end{aligned}$$

Uzimajući normu lijeve i desne strane, dijeleći sa $\|x\|$, te pretpostavljajući da je $\|A^{-1}\Delta A\| \leq$

$\|A^{-1}\|\,\|\Delta A\| < 1$, dobijemo

$$\begin{aligned}
\frac{\|\Delta x\|}{\|x\|} &\leq \|\left(I + A^{-1}\Delta A\right)^{-1}\|\,\|A^{-1}\|\left(\|\Delta A\| + \frac{\|\Delta b\|}{\|x\|}\right) \\
&\leq \frac{\|A^{-1}\|}{1 - \|A^{-1}\Delta A\|}\left(\|\Delta A\| + \frac{\|\Delta b\|}{\|x\|}\right) \\
&\leq \frac{\|A^{-1}\|}{1 - \|A^{-1}\|\|\Delta A\|\frac{\|A\|}{\|A\|}}\frac{\|A\|}{\|A\|}\left(\|\Delta A\| + \frac{\|\Delta b\|}{\|x\|}\right) \\
&= \frac{\|A^{-1}\|\|A\|}{1 - \|A^{-1}\|\|A\|\frac{\|\Delta A\|}{\|A\|}}\left(\frac{\|\Delta A\|}{\|A\|} + \frac{\|\Delta b\|}{\|A\|\,\|x\|}\right) \\
&\leq \frac{\kappa(A)}{1 - \kappa(A)\frac{\|\Delta A\|}{\|A\|}}\left(\frac{\|\Delta A\|}{\|A\|} + \frac{\|\Delta b\|}{\|b\|}\right).
\end{aligned}$$

Ovdje smo iskoristili drugu tvrdnju leme 1.1, a u zadnjem redu i činjenicu da $b = Ax$ povlači $\|b\| \leq \|Ax\| \leq \|A\|\,\|x\|$. Ako uvedemo relativne greške po normi

$$\varepsilon(A) = \frac{\|\Delta A\|}{\|A\|}, \quad \varepsilon(b) = \frac{\|\Delta b\|}{\|b\|}, \quad \text{i} \quad \varepsilon(x) = \frac{\|\Delta x\|}{\|x\|} \tag{1.9}$$

onda zadnju relaciju možemo zapisati u kompaktnom obliku

$$\varepsilon(x) \leq \frac{\kappa(A)}{1 - \kappa(A)\varepsilon(A)}\left(\varepsilon(A) + \varepsilon(b)\right). \tag{1.10}$$

Vidimo da relacija (1.10) vrijedi ako je A regularna i ako je

$$\varepsilon(A) < \frac{1}{\kappa(A)}. \tag{1.11}$$

Uvjet (1.11) garantira da je i matrica $A + \Delta A$ regularna. Doista, vrijedi

$$A + \Delta A = A(I + A^{-1}\Delta A)$$

i pritom je A regularna, a izraz u zagradi je regularna matrica jer je

$$\|A^{-1}\Delta A\| \leq \|A^{-1}\|\,\|\Delta A\| = \|A^{-1}\|\,\|A\|\,\frac{\|\Delta A\|}{\|A\|} = \kappa(A)\varepsilon(A) < 1.$$

Sjetimo se da je gore pokazano da $\|S\| < 1$ za konzistentnu normu povlači regularnost matrice $I - S$. Ovdje je $S = A^{-1}\Delta A$. Time je dokazan

Teorem 1.2 *Neka je A regularna matrica i neka je ΔA takva da vrijedi uvjet (1.11) uz oznake kao u relaciji (1.9) pri čemu je matrična norma inducirana vektorskom normom. Tada je $A + \Delta A$ regularna pa su za proizvoljne vektore b i Δb relacijama (1.1) i (1.2) dobro definirani vektori x i Δx. Za njih tada vrijedi ocjena (1.10).*

Inverz broja uvjetovanosti ima geometrijsko značenje relativne udaljenosti po normi matrice od skupa singularnih matrica. To osigurava sljedeći teorem kojeg navodimo bez dokaza

Teorem 1.3 *Ako je A regularna, tada je*

$$\min\left\{\frac{\|\Delta A\|}{\|A\|} : \ A+\Delta A \textit{ singularna}\right\} = \frac{1}{\|A^{-1}\|\,\|A\|} = \frac{1}{\kappa(A)}.$$

Broj uvjetovanosti ima sljedeća osnovna svojstva.

- Za svaku induciranu matričnu normu vrijedi $\kappa(A) \geq 1$.
- Za svaku matričnu normu i svaki skalar $\alpha \neq 0$ vrijedi $\kappa(aA) = \kappa(A)$.
- Za spektralnu normu vrijedi $\kappa_2(A) = 1$ ako i samo ako je A multipl unitarne matrice.

Doista, jer je matrična norma inducirana vektorskom, imamo $\|I\| = 1$, pa za regularnu matricu A vrijedi

$$1 = \|I\| = \|AA^{-1}\| \leq \|A\|\,\|A^{-1}\| = \kappa(A).$$

Ako je $\alpha \neq 0$ i A regularna, tada je $(\alpha A)^{-1} = \alpha^{-1}A^{-1}$, pa je

$$\kappa(\alpha A) = \|\alpha A\|\,\|\alpha^{-1}A^{-1}\| = |\alpha|\,|\alpha^{-1}|\|A\|\,\|A^{-1}\| = \kappa(A).$$

Konačno, neka je $A = U\Sigma V^*$ singularna dekompozicija od A sa singularnim vrijednostima $\sigma_{max}(A) = \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n = \sigma_{min}(A) > 0$. Tada je zbog unitarne invarijantnosti spektralne norme

$$\|A\|_2 = \|U\Sigma V^*\|_2 = \|\Sigma\|_2 = \sigma_{max}(A), \quad \|A^{-1}\|_2 = \|V\Sigma^{-1}U^*\|_2 = \|\Sigma^{-1}\|_2 = \frac{1}{\sigma_{min}(A)}.$$

Dakle je

$$\kappa_2(A) = \frac{\sigma_{max}(A)}{\sigma_{min}(A)},$$

pa uvjet $\kappa_2(A) = 1$ povlači $\sigma_{max}(A) = \sigma_{min}(A)$, tj. $\Sigma = \sigma_{max}(A)I$. Dobili smo da je $A = \sigma_{max}(A)UV^*$, multipl unitarne matrice UV^*.

Drugi smjer ide još lakše, jer ako je $A = \alpha W$ za neku unitarnu matricu, tada je $A^{-1} = \alpha^{-1}U^*$, pa je

$$\kappa_2(A) = \|\alpha W\|\,\|\alpha^{-1}W^*\| = |\alpha|\,|\alpha^{-1}| = 1.$$

Sljedeći pristup teoriji perturbacije pogodan je i za iterativne metode jer koristi u ocjeni reziduale. Ako je $\hat{x}$ bilo koji vektor, tada možemo ocijeniti razliku rješenja x sustava $Ax = b$ i $\hat{x}$. Neka je $\delta x = \hat{x} - x$. Tada je

$$\delta x = \hat{x} - x = \hat{x} - A^{-1}b = -A^{-1}(b - A\hat{x}) = A^{-1}r,$$

gdje je $r = A\hat{x} - b$ rezidual od $\hat{x}$. Dakle je

$$\|\delta x\| = \|A^{-1}r\| \leq \|A^{-1}\|\|r\|. \tag{1.12}$$

Ova jednostavna ocjena vrlo je upotrebljiva u praksi jer se r jednostavno i brzo računa iz $\hat{x}$. Pritom nema potrebe ocjenjivati ΔA i Δb. Zapravo su dva opisana pristupa vrlo povezana kako pokazuje sljedeci teorem.

Teorem 1.4 *Neka je $r = A\hat{x} - b$. Tada postoji ΔA takav da je $\|\Delta A\| = \|r\|/\|x\|$ i $(A + \Delta A)\hat{x} = b$. Ne postoji ΔA manje norme koji bi zadovoljavao $(A + \Delta A)\hat{x} = b$. To znači da je ΔA najmanja moguća povratna greška mjerena normom. Tvrdnja vrijedi za svaku vektorsku normu i pripadnu induciranu matričnu normu (takoder za Euklidsku vektorsku i Frobeniusovu matričnu normu).*

Dakle, najmanja $\|\Delta A\|$ za koju će vrijediti $(A + \Delta A)\hat{x} = b$ dana je izrazom $\|r\|/\|x\|$. Sada možemo tu perturbaciju ΔA iskoristiti u relaciji (1.5). Uočimo da je sada pretpostavljeno da je $\Delta b = 0$. Imamo

$$\|\Delta x\| \leq \|A^{-1}\|(\|\Delta A\|\|\hat{x}\|) = \|A^{-1}\|(\frac{\|r\|}{\|\hat{x}\|}\|\hat{x}\|) = \|A^{-1}\|\|r\|,$$

pa smo dobili istu ogradu kao u relaciji (1.12).

1.2 Iterativne metode za rješavanje linearnih sustava

1.2.1 Općenito o iterativnim metodama

Umjesto direktnih metoda za rješavanje linearnih sustava, u praksi se često koriste iterativne metode, posebno za šuplje sustave vrlo velikih redova.

Pretpostavimo da je A regularna matrica reda n. Iterativna metoda koja pronalazi približno rješenje sustava $Ax = b$ zadana je početnim vektorom $x^{(0)}$ i generirana je nizom iteracija $x^{(m)}$, $m \in \mathbf{N}$, koji (nadamo se) konvergira prema rješenju linearnog sustava x.

U praksi se gotovo isključivo koriste iterativne metode prvog reda, koje iz jednog prethodnog vektora $x^{(m)}$ nalaze sljedeću aproksimaciju $x^{(m+1)}$.

Ideja iterativnih metoda je brzo računanje $x^{(m+1)}$ iz $x^{(m)}$. Kriterij zaustavljanja sličan je kao kod svih metoda "limes" tipa – tj. kad je $x^{(m)}$ dovoljno dobra aproksimacija za pravo rješenje x. Naravno, i tu postoji problem, jer pravi x ne znamo. Zbog toga, koristimo svojstvo da je konvergentan niz Cauchyjev, tj. da susjedni članovi niza moraju postati po volji bliski. Standardno, uzima se da su $x^{(m+1)}$ i $x^{(m)}$ dovoljno bliski ako je

$$\|x^{(m+1)} - x^{(m)}\| \leq \varepsilon,$$

gdje je ε neka unaprijed zadana točnost (reda veličine u, odnosno $n \cdot u$), a $\| \ \|$ neka vektorska norma.

Da bismo definirali iterativnu metodu, potrebno je pažljivo rastaviti matricu A.

Definicija 1.2.1. Rastav *matrice A je par matrica (M, K) (obje reda n) za koje vrijedi*

(a) $A = M - K$,

(b) M *je regularna.*

Bilo koji rastav matrice A generira iterativnu metodu na sljedeći način:

$$Ax = Mx - Kx = b \implies Mx = Kx + b,$$

pa zbog regularnosti od M izlazi

$$x = M^{-1}Kx + M^{-1}b.$$

Ako označimo $R = M^{-1}K$, $c = M^{-1}b$ onda je prethodna relacija ekvivalentna s

$$x = Rx + c.$$

Time smo definirali iterativnu metodu

$$x^{(m+1)} = Rx^{(m)} + c, \quad m \in \mathbf{N}_0. \tag{1.13}$$

Pravo rješenje x je fiksna točka iteracione funkcije (1.13),

$$f(x) = Rx + c.$$

To znači da u analizi konvergencije možemo koristiti poznate teoreme o fiksnoj točki.

Kriterij konvergencije ovakvih metoda je jednostavan.

Lema 1.5 *Niz iteracija* $\left(x^{(m)}\right)$, $m \in \mathbf{N}_0$, *generiran relacijom* (1.13) *konvergira prema rješenju linearnog sustava* $Ax = b$ *za* sve početne vektore $x^{(0)}$ *i* sve desne strane b, *ako je*

$$\|R\| < 1,$$

za neku induciranu normu.

Dokaz: Oduzmimo $x = Rx + c$ od $x^{(m+1)} = Rx^{(m)} + c$. Dobivamo

$$x^{(m+1)} - x = R\big(x^{(m)} - x\big),$$

pa uzimanjem norme dobivamo

$$\left\|x^{(m+1)} - x\right\| \leq \|R\| \left\|x^{(m)} - x\right\| \leq \|R\|^{m+1} \left\|x^{(0)} - x\right\|.$$

No, zbog $\|R\| < 1$ slijedi $\|R\|^{m+1} \to 0$ za $m \to \infty$, odakle izlazi $\left\|x - x^{(m+1)}\right\| \to 0$. Drugim riječima, za svaki $\varepsilon > 0$, postoji $m_0 \in \mathbf{N} : 0$ takav da je za svaki $m \geq m_0$ $\left\|x - x^{(m+1)}\right\| < \varepsilon$. To znači da kako god malu okolinu (otvorenu kuglu) oko x uzeli, svi osim konačno članova niza bit će u njoj. To upravo znači da $x^{(m+1)} \to x$ kad $k \to \infty$. Tvrdnja vrijedi za svaki polazni $x^{(0)}$. ■

No, može se dobiti i nešto bolji rezultat, korištenjem veze spektralnog radijusa i inducirane (operatorske) norme matrice.

Spektralni radijus $\rho(X)$ definiran je kao najveća udaljenost neke vlastite vrijednosti matrice X od ishodišta

$$\rho(X) = \max_{\lambda \in \sigma(X)} |\lambda|, \qquad \sigma(X) = \{\lambda\,;\, Xx = \lambda x,\ x \neq 0\}.$$

Dakle, zatvoren krug radijusa $\rho(X)$ oko ishodišta sadrži sve vlastite vrijednosti matrice X. To je najmanji takav krug koji obuhvaća cijeli spektar $\sigma(X)$ matrice X.

Lema 1.6 *Za svaku operatorsku normu* $\|\ \|$ *vrijedi*

$$\rho(R) \leq \|R\|.$$

Za svaki R *i svaki* $\varepsilon > 0$ *postoji operatorska norma* $\|\ \|_*$ *takva da je*

$$\|R\|_* \leq \rho(R) + \varepsilon.$$

Norma $\|\ \|_*$ *ovisi i o* R, *i o* ε.

Dokaz: Da bismo dokazali da je $\rho(R) \leq \|R\|$ za bilo koju operatorsku normu, neka je λ ona vlastita vrijednost od R za koju vrijedi $\rho(R) = |\lambda|$. Neka je x pripadni vlastiti vektor od λ. Tada za svaku vektorsku i pripadnu operatorsku normu vrijedi

$$\|R\| = \max_{y \neq 0} \frac{\|Ry\|}{\|y\|} \geq \frac{\|Rx\|}{\|x\|} = \frac{\|\lambda x\|}{\|x\|} = \frac{|\lambda| \, \|x\|}{\|x\|} = |\lambda| = \rho(R).$$

S druge strane, da bismo konstruirali operatorsku normu $\| \ \|_*$ takvu da je $\|R\|_* \leq \rho(R) + \varepsilon$, svedimo matricu R na njenu Jordanovu formu. Znamo da postoji regularna matrica S za koju je $J = S^{-1}RS$ Jordanova forma matrice R. Jedinice iznad dijagonale u J općenito "udaljuju" matričnu normu od spektralnog radijusa od J, pa je ideja da na njihovom mjestu pokušamo dobiti zadani $\varepsilon > 0$. To se postiže korištenjem transformacije sličnosti s dijagonalnom matricom. Neka je

$$D_\varepsilon = \operatorname{diag}(1, \varepsilon, \varepsilon^2, \ldots, \varepsilon^{n-1}).$$

Tada je

$$J_\varepsilon = (SD_\varepsilon)^{-1} R (SD_\varepsilon) = D_\varepsilon^{-1} J D_\varepsilon = \left[\begin{array}{cccc|ccc|c} \lambda_1 & \varepsilon & & & & & & \\ & \ddots & \ddots & & & & & \\ & & \ddots & \varepsilon & & & & \\ & & & \lambda_1 & & & & \\ \hline & & & & \lambda_2 & \varepsilon & & \\ & & & & & \ddots & \ddots & \\ & & & & & & \ddots & \varepsilon \\ & & & & & & & \lambda_2 \\ \hline & & & & & & & \ddots \end{array}\right].$$

Uzmimo kao vektorsku normu

$$\|x\|_* = \|(SD_\varepsilon)^{-1} x\|_\infty$$

(pokažite da je to vektorska norma!). Sa $\| \ \|_*$ označimo operatorsku (matričnu) normu generiranu tom vektorskom normom. U toj operatorskoj normi vrijedi

$$\begin{aligned} \|R\|_* &= \max_{x \neq 0} \frac{\|Rx\|_*}{\|x\|_*} = \max_{x \neq 0} \frac{\|(SD_\varepsilon)^{-1} Rx\|_\infty}{\|(SD_\varepsilon)^{-1} x\|_\infty} = \max_{y \neq 0} \frac{\|(SD_\varepsilon)^{-1} R (SD_\varepsilon) y\|_\infty}{\|y\|_\infty} \\ &= \|(SD_\varepsilon)^{-1} R (SD_\varepsilon)\|_\infty = \|J_\varepsilon\|_\infty \leq \max_i |\lambda_i| + \varepsilon = \rho(R) + \varepsilon. \end{aligned}$$

■

Konačno, prethodne dvije leme daju potpunu karakterizaciju konvergencije iterativnih metoda.

Teorem 1.7 *Niz iteracija* $\left(x^{(m)}\right)$, $m \in \mathbf{N}_0$, *generiran relacijom* (1.13) *konvergira prema rješenju linearnog sustava* $Ax = b$ *za* sve početne vektore $x^{(0)}$ *i* sve desne strane b, *ako i samo ako je*

$$\rho(R) < 1,$$

pri čemu je $\rho(R)$ *spektralni radijus matrice* $R = M^{-1}K$. *Uočite da* $\rho(R)$ *ovisi samo o* A *i njenom rastavu, a ne o* b.

Dokaz: Kako bi pokazali da je $\rho(R) < 1$ nužan uvjet konvergencije koristimo tzv. obrat po kontrapoziciji (umjesto da dokažemo implikaciju $A \Rightarrow B$ pokazujemo ekvivalentnu implikaciju $\neg B \Rightarrow \neg A$). To znači da pretpostavka kako nije $\rho(R) < 1$ povlači da postoje vektori b i x_0 takvi da niz $\left(x^{(m)}\right)$, $m \in \mathbf{N}_0$ ne konvergira.

Ako je $\rho(R) \geq 1$, izaberimo proizvoljni vektor b i startnu aproksimaciju $x^{(0)}$ takvu da je $x^{(0)} - x$ svojstveni vektor koji pripada najvećoj po modulu vlastitoj vrijednosti od R, $|\lambda| = \rho(R)$. Ovdje je x jedinstveno rješenje sustava $Ax = b$.

Tada vrijedi

$$\left(x^{(m+1)} - x\right) = R\left(x^{(m)} - x\right) = \cdots = R^{m+1}\left(x^{(0)} - x\right) = \lambda^{m+1}(x^{(0)} - x\big).$$

Dakle, $\|x^{(m+1)} - x\| = |\lambda|^{m+1}\|x^{(0)} - x\|$ raste u ∞ kad $m \to \infty$ pa niz $\left(x^{(m)}\right)$, $m \in \mathbf{N}_0$ nije konvergentan. Uočite da je $x^{(0)} - x$ različit od nul-vektora jer je svojstven vektor.

Pokažimo da je uvjet $\rho(R) < 1$ dovoljan za konvergenciju niza $\left(x^{(m)}\right)$, $m \in \mathbf{N}_0$ za svaki $x^{(0)}$ i svaki b. Neka je $\rho(R) < 1$. Tada možemo izabrati $\varepsilon > 0$ takav da je

$$\rho(R) + \varepsilon < 1,$$

a zatim po lemi 1.6 i operatorsku normu $\| \ \|_*$ takvu da vrijedi

$$\|R\|_* \leq \rho(R) + \varepsilon < 1. \tag{1.14}$$

Relacija (1.14) i lema 1.5 odmah daju traženi rezultat. ■

Lema 1.5 i teorem 1.7, zapravo nam daju i brzinu konvergencije. Prisjetimo se što je brzina konvergencije.

Definicija 1.8 *Za niz aproksimacija* $x^{(m)}$, $m \in \mathbf{N}_0$, *reći ćemo da konvergira prema* x *s redom* p *ako je*

$$\|x^{(m+1)} - x\| \leq c\|x^{(m)} - x\|^p, \quad c \in \mathbf{R}_0^+.$$

Ako je $p = 1$, *imamo linearnu konvergenciju i tada mora biti* $c < 1$ *(tzv. geometrijska konvergencija s faktorom* c*).*

U slučaju iterativnih metoda, konvergencija je linearna, a faktor je $\rho(R) < 1$, tj. vrijedi

$$\|x^{(m+1)} - x\|_* \leq \rho(R)\|x^{(m)} - x\|_*.$$

Podijelimo li prethodnu relaciju s $\rho(R) \cdot \|x^{(m+1)} - x\|_*$, a zatim logaritmiramo, dobivamo

$$-\log \rho(R) \leq \log \|x^{(m)} - x\|_* - \log \|x^{(m+1)} - x\|_*, \tag{1.15}$$

pa zbog toga broj

$$r(R) = -\log \rho(R)$$

možemo definirati kao brzinu konvergencije iteracija. Što nam kaže relacija (1.15)? Broj $r(R)$ je porast broja korektnih decimalnih znamenki u rješenju po iteraciji. Dakle, što je manji $\rho(R)$, to je veća brzina konvergencije iteracija.

Naravno, sljedeći cilj nam je odgovoriti na pitanje kako odrediti rastav matrice $A = M - K$ koji zadovoljava dva uvjeta:

(1) da se $Rx = M^{-1}Kx$ i $c = M^{-1}b$ lako računaju,

(2) da je $\rho(R)$ malen ?

Odmah nam se nameću neka jednostavna rješenja za ova dva suprotna cilja. Na primjer, izaberemo li $M = I$, M je regularna, i Rx i c se lako računaju, ali nije jasno da smo zadovoljili da je $\rho(R) < 1$. S druge strane, izbor $K = 0$ je izvrsan za drugi cilj ($\rho(R) = 0$), ali nije dobar za prvi cilj, jer je $c = A^{-1}b$, tj. dobivamo polazni problem kojeg treba riješiti ($x = c$).

Dakle, rastav koji bi uvijek dobro radio nije lako konstruirati. Međutim tu će nam pomoći praksa. Matrice koje se javljaju u praksi su ili pozitivno definitne ili (strogo) dijagonalno dominantne, pa će za takve tipove matrica biti mnogo lakše konstruirati iterativne metode i pokazati da one konvergiraju.

Uvedimo sljedeću notaciju. Pretpostavimo da A nema nula na dijagonali. Tada A možemo zapisati kao

$$A = D - \widetilde{L} - \widetilde{U} = D(I - L - U),$$

pri čemu je D dijagonala od A, $-\widetilde{L}$ striktno donji trokut od A, a $-\widetilde{U}$ striktno gornji trokut od A. Pritom je $DL = \widetilde{L}$, $DU = \widetilde{U}$.

1.3 Jacobijeva metoda

Općenito govoreći Jacobijeva metoda u petlji (npr. za $j = 1, 2, \ldots, n$) prolazi kroz jednadžbe linearnog sustava, mijenjajući j-tu varijablu, tako da j-ta jednadžba bude ispunjena. Dakle, u $(m+1)$-om koraku vrijednost varijable x_j, u oznaci $x_j^{(m+1)}$, računamo iz j-te jednadžbe korištenjem aproksimacija iz m-tog koraka za preostale varijable, tj. vrijedi

$$a_{jj}x_j^{(m+1)} + \sum_{\substack{k=1 \\ k\neq j}}^{n} a_{jk}x_k^{(m)} = b_j, \quad j = 1, \ldots, n. \tag{1.16}$$

Naravno, to ide ako i samo ako je $a_{jj} \neq 0$ za sve j.

Vidimo da na nove komponente "djeluju" samo dijagonalni elementi matrice A, dok svi ostali djeluju na stare (prethodne ili prošle) komponente. Skupimo li sve jednadžbe iz (1.16) za jedan korak, onda ih zajedno možemo zapisati kao

$$Dx^{(m+1)} = (\widetilde{L} + \widetilde{U})x^{(m)} + b,$$

ili

$$x^{(m+1)} = D^{-1}(\widetilde{L} + \widetilde{U})x^{(m)} + D^{-1}b := R_{Jac}x^{(m)} + c_{Jac},$$

uz

$$R_{Jac} = D^{-1}(\widetilde{L} + \widetilde{U}) = L + U, \quad c_{Jac} = D^{-1}b.$$

Pripadni rastav matrice A je

$$A = D - (\widetilde{L} + \widetilde{U}),$$

tj. $M = D$, $K = \widetilde{L} + \widetilde{U}$.

Algoritam 1.3.1. (Jedan korak Jacobijeve metode)

for $j := 1$ **to** n **do**

$$x_j^{(m+1)} := \left(b_j - \sum_{\substack{k=1 \\ k \neq j}}^{n} a_{jk} x_k^{(m)} \right) \Big/ a_{jj};$$

Uočimo da petlju po j u ovom algoritmu, odnosno prolaz kroz jednadžbe sustava u (1.16), možemo napraviti **bilo kojim** redom po j — ne nužno sekvencijalnim. Komponente novog vektora $x^{(m+1)}$ ovise samo o komponentama starog vektora $x^{(m)}$, a ne i o nekim novim komponentama. Zbog toga je Jacobijeva metoda idealna za paralelno računanje, jer pojedine komponente novog vektora možemo računati potpuno nezavisno.

1.4 Gauss–Seidelova metoda

Ako komponente novog vektora u Jacobijevoj metodi zaista računamo sekvencijalno, od prve prema zadnjoj, odmah se nameće ideja za poboljšanje. Naime, u aproksimaciji j-te varijable u $(m+1)$-om koraku koristimo aproksimaciju svih ostalih varijabli iz prethodnog m-tog koraka, iako već imamo izračunate poboljšane varijable $x_i^{(m+1)}$, za $i < j$, iz novog koraka. Iskoristimo li sve poznate nove komponente umjesto starih, onda (1.16) glasi

$$a_{jj}x_j^{(m+1)} + \sum_{k=1}^{j-1} a_{jk}x_k^{(m+1)} + \sum_{k=j+1}^{n} a_{jk}x_k^{(m)} = b_j, \quad j = 1, \ldots, n. \tag{1.17}$$

Ovu metodu zovemo Gauss–Seidelova metoda. Poredak prolaska kroz jednadžbe sustava postaje potpuno određen i strogo sekvencijalan, od prve prema zadnjoj. Opet, to ide ako i samo ako je $a_{jj} \neq 0$ za sve j.

Na nove komponente djeluju, osim dijagonalnih, i svi elementi donjeg trokuta matrice A, dok samo strogo gornji trokut djeluje na stare komponente. Sve jednadžbe iz (1.17) za jedan korak iteracija možemo zajedno zapisati u obliku

$$(D - \widetilde{L})x^{(m+1)} = \widetilde{U}x^{(m)} + b,$$

ili

$$x^{(m+1)} = (D - \widetilde{L})^{-1}\widetilde{U}x^{(m)} + (D - \widetilde{L})^{-1}b := R_{GS}x^{(m)} + c_{GS},$$

uz

$$R_{GS} = (D - \widetilde{L})^{-1}\widetilde{U} = (I - L)^{-1}U, \quad c_{GS} = (D - \widetilde{L})^{-1}b = (I - L)^{-1}D^{-1}b.$$

Matrica R_{GS} je singularna, jer je U strogo gornja trokutasta ($\det U = 0$). Pripadni rastav matrice A je

$$A = (D - \widetilde{L}) - \widetilde{U},$$

tj. $M = D - \widetilde{L}$, $K = \widetilde{U}$.

Algoritam 1.4.1. (Jedan korak Gauss–Seidelove metode)

for $j := 1$ **to** n **do**

$$x_j^{(m+1)} := \left(b_j - \sum_{k=1}^{j-1} a_{jk}x_k^{(m+1)} - \sum_{k=j+1}^{n} a_{jk}x_k^{(m)} \right) \Big/ a_{jj};$$

Zgodna je stvar da u implementaciji Gauss–Seidelove metode ne moramo pamtiti dva vektora, već samo jedan, tako da nove izračunate komponente prepisujemo preko starih u istom polju x.

S druge strane, redosljed "popravaka" varijabli u Jacobijevom algoritma ne igra nikakvu ulogu, u smislu da bilo koji poredak izvršavanja petlje po j daje isti rezultat i, naravno, istu iterativnu metodu. Nasuprot tome, redosljed "popravaka" kod Gauss–Seidelovog algoritma je bitan. U (1.17) i u algoritmu 1.4.1. uzeli smo prirodni redosljed — od prve prema zadnjoj jednadžbi, odnosno, varijabli (petlja po j od 1 do n).

To **ne** znači da je to i jedini mogući redosljed. U principu, možemo uzeti i bilo koji drugi redosljed, tj. bilo koju drugu od mogućih $n!$ permutacija jednadžbi. No, zbog sekvencijalnosti popravaka, rezultat nije isti, tj. dobivamo drugačiju iterativnu metodu. U praksi se katkad i koriste drugačiji redosljedi, ali za sasvim posebne linearne sustave koji nastaju diskretizacijom parcijalnih diferencijalnih jednadžbi. Na primjer, za Laplaceovu jednadžbu u dvije dimenzije koristi se tzv. "crveno–crni" poredak (engl. "red–black" ordering), koji naliči šahovskoj ploči s crvenim i crnim poljima, s tim da se prvo računaju sva polja crvene boje, a zatim sva polja crne boje. Zbog posebne strukture sustava, ovaj poredak dozvoljava i efikasnu paralelizaciju.

1.5 Uvođenje relaksacijskog parametra

Kad jednom znamo konstruirati iterativni proces, nameće se vrlo jednostavna ideja za njegovo poboljšanje, uvođenjem jednog realnog parametra. Nove aproksimacije možemo računati u dva koraka. Prvo iz $x^{(m)}$ nađemo (jednostavnu) pomoćnu sljedeću aproksimaciju $x_*^{(m+1)}$, a zatim za "pravu" novu aproksimaciju $x^{(m+1)}$ uzmemo težinsku sredinu prethodne aproksimacije $x^{(m)}$ i pomoćne nove aproksimacije $x_*^{(m+1)}$

$$x^{(m+1)} = (1-\omega)x^{(m)} + \omega x_*^{(m+1)} = x^{(m)} + \omega(x_*^{(m+1)} - x^{(m)}), \tag{1.18}$$

gdje je ω težinski parametar kojeg možemo birati. Očito, za $\omega = 1$ dobivamo $x^{(m+1)} = x_*^{(m+1)}$, pa je ova metoda proširenje metode za nalaženje pomoćnih aproksimacija. Obično se uzima $\omega \in \mathbf{R}$ i $\omega \neq 0$, da ne dobijemo stacionaran niz.

Ideja za ovakav postupak dolazi iz općih metoda za rješavanje jednadžbi i minimizaciju funkcionala. Pomoćna aproksimacija $x_*^{(m+1)}$ daje **smjer** korekcije prethodne aproksimacije $x^{(m)}$ u kojem treba ići da bismo se približili pravom rješenju sustava, dakle smanjili rezidual $r(x) = Ax - b$ u nekoj normi. No, ako je $x_*^{(m+1)} - x^{(m)}$ dobar smjer korekcije, onda moramo odrediti i izbor duljine koraka ω u smjeru te korekcije, tako da dobijemo što bolji $x^{(m+1)}$. Općenito očekujemo da je $\omega > 0$, tj. da idemo u smjeru vektora korekcije, a ne suprotno od njega, a stvarno želimo dobiti $\omega > 1$, tako da se još više maknemo u dobrom smjeru i približimo pravom rješenju ili točki minimuma.

Naziv "relaksacija" dolazi upravo iz minimizacijskih metoda, a odnosi se na sve iterativne metode koje koriste neki oblik minimizacije ili pokušaja minimizacije reziduala. U tom smislu, često se koristi i tradicionalni naziv "relaksacijski parametar" za ω.

Obzirom na vrijednost parametra ω, imamo tri različita slučaja. Ako je $\omega = 1$, onda se metoda svodi na pomoćnu metodu i to je tzv. obična ili standardna relaksacija. Ako je $\omega < 1$, onda takvu metodu zovemo podrelaksacija (engl. "underrelaxation"), a ako je $\omega > 1$ onda metodu zovemo nad- ili pre-relaksacija (engl. "overrelaxation").

U općem slučaju, ω se posebno računa u svakoj pojedinoj iteraciji, tako da dobijemo što bolji $x^{(m+1)}$. Postupak se svodi na jednodimenzionalnu optimizaciju (kao i određivanje koraka u višedimenzionalnoj optimizaciji), a ovisi o kriteriju optimalnosti za mjerenje "kvalitete" aproksimacija.

Srećom, za neke klase linearnih sustava, koje su izrazito bitne u praksi, unaprijed se može dobro odabrati optimalni ili skoro optimalni parametar ω za maksimalno ubrzanje konvergencije iterativnih metoda i to tako da isti ω vrijedi za sve iteracije. U većini slučajeva dobivamo $\omega > 1$ za optimalni ω, pa se takve metode standardno zovu "OverRelaxation" i skraćeno označavaju s OR. Obzirom na to da se ω zadaje ili bira unaprijed, a zatim koristi za sve iteracije, metoda je ovisna o jednom parametru i standardno koristimo oznaku OR(ω).

1.5.1 JOR metoda (Jacobi overrelaxation)

Ako se pomoćna nova aproksimacija $x_*^{(m+1)}$ iz (1.18) računa po Jacobijevoj metodi, dobivamo Jacobijevu nadrelaksaciju ili JOR metodu. Iz (1.16) za komponente pomoćne aproksimacije $x_{Jac}^{(m+1)}$ vrijedi

$$x_{j,Jac}^{(m+1)} = \left(b_j - \sum_{\substack{k=1 \\ k \neq j}}^{n} a_{jk} x_k^{(m)} \right) \Big/ a_{jj}, \quad j = 1, \ldots, n.$$

Kad to uvrstimo u (1.18) dobivamo sljedeći algoritam.

Algoritam 1.5.1. (Jedan korak JOR(ω) metode)

for $j := 1$ **to** n **do**

$$x_j^{(m+1)} := (1-\omega) x_j^{(m)} + \frac{\omega}{a_{jj}} \left(b_j - \sum_{\substack{k=1 \\ k \neq j}}^{n} a_{jk} x_k^{(m)} \right);$$

Kao i kod obične Jacobijeve metode, petlju po j možemo izvršiti bilo kojim redom po j, uz isti rezultat, pa se i ovdje komponente novog vektora $x^{(m+1)}$ mogu paralelno računati.

Vektorski oblik iteracija u JOR(ω) metodi je

$$x^{(m+1)} = (1-\omega) x^{(m)} + \omega (R_{Jac} x^{(m)} + c_{Jac}) := R_{JOR(\omega)} x^{(m)} + c_{JOR(\omega)}$$

pa je

$$\begin{aligned} R_{JOR(\omega)} &= (1-\omega) I + \omega R_{Jac} = (1-\omega) I + \omega (L+U), \\ c_{JOR(\omega)} &= \omega c_{Jac} = \omega D^{-1} b. \end{aligned}$$

Pripadni rastav matrice A je

$$A = \frac{1}{\omega} D - \left(\frac{1-\omega}{\omega} D + \widetilde{L} + \widetilde{U} \right),$$

tj.

$$M = \frac{1}{\omega} D, \qquad K = \frac{1-\omega}{\omega} D + \widetilde{L} + \widetilde{U}.$$

Naravno, za $\omega = 1$ dobivamo Jacobijevu metodu.

Iz oblika matrice $R_{JOR(\omega)}$ vidimo da se optimalni parametar ω koji minimizira njen spektralni radijus može odrediti unaprijed, prije početka iteracija. Drugim riječima, treba koristiti isti ω u svim iteracijama, naravno, pod uvjetom da imamo konvergenciju.

1.6 SOR metoda (Successive overrelaxation)

Relacija (1.18) koristi ideju težinske sredine ili duljine koraka na nivou vektorskih aproksimacija $x^{(m)}$. To prirodno odgovara Jacobijevoj metodi i paralelnom računanju. Međutim, potpuno istu ideju možemo koristiti i za poboljšanje svake pojedine varijable $x_j^{(m)}$, tj. pojedinačnih komponenti vektora $x^{(m)}$, što odgovara "Gauss–Seidelovskom" pristupu. Dakle, nova aproksimacija j-te varijable ima oblik

$$x_j^{(m+1)} = (1-\omega)x_j^{(m)} + \omega x_{j,*}^{(m+1)} = x_j^{(m)} + \omega(x_{j,*}^{(m+1)} - x_j^{(m)}), \quad j = 1, \ldots, n, \tag{1.19}$$

gdje je $x_{j,*}^{(m+1)}$ neka pomoćna nova aproksimacija j-te varijable, koju računamo tog trenutka kad nam treba, za svaki pojedini j.

SOR metoda (engl. "Successive OverRelaxation" ili ponovljena nad- ili prerelaksacija) je proširenje ili poboljšanje Gauss–Seidelove metode u smislu da se pomoćna nova aproksimacija $x_{j,*}^{(m+1)}$ iz (1.19) računa po Gauss–Seidelovoj metodi, pa ju označavamo s $x_{j,GS}^{(m+1)}$. Relacija (1.19) za j-tu komponentu nove aproksimacije u SOR(ω) metodi ima oblik

$$x_j^{(m+1)} = (1-\omega)x_j^{(m)} + \omega x_{j,GS}^{(m+1)}, \quad j = 1, \ldots, n. \tag{1.20}$$

Iz (1.17), trenutna pomoćna Gauss–Seidelova aproksimacija $x_{j,GS}^{(m+1)}$ koju možemo izračunati iz poznatih prvih $j-1$ komponenti novog vektora $x^{(m+1)}$ i preostalih komponenti iz prethodne aproksimacije $x^{(m)}$ je

$$x_{j,GS}^{(m+1)} = \left(b_j - \sum_{k=1}^{j-1} a_{jk} x_k^{(m+1)} - \sum_{k=j+1}^{n} a_{jk} x_k^{(m)} \right) \Big/ a_{jj}.$$

Kad to uvrstimo u (1.20), možemo izračunati j-tu komponentu $x_j^{(m+1)}$ nove aproksimacije po SOR(ω) metodi.

Algoritam 1.6.1. (Jedan korak SOR(ω) metode)

for $j := 1$ **to** n **do**

$$x_j^{(m+1)} := (1-\omega)x_j^{(m)} + \frac{\omega}{a_{jj}} \left(b_j - \sum_{k=1}^{j-1} a_{jk} x_k^{(m+1)} - \sum_{k=j+1}^{n} a_{jk} x_k^{(m)} \right);$$

Iz algoritma odmah vidimo da je

$$a_{jj} x_j^{(m+1)} + \omega \sum_{k=1}^{j-1} a_{jk} x_k^{(m+1)} = (1-\omega) a_{jj} x_j^{(m)} - \omega \sum_{k=j+1}^{n} a_{jk} x_k^{(m)} + \omega b_j, \quad j = 1, \ldots, n.$$

Ove jednadžbe možemo zapisati u vektorskom obliku

$$(D - \omega\widetilde{L})x^{(m+1)} = \big((1-\omega)D + \omega\widetilde{U}\big)x^{(m)} + \omega b,$$

odakle slijedi (npr. prvo množenjem s lijeva s D^{-1}, a zatim množenjem s lijeva s $(I - \omega L)^{-1}$)

$$\begin{aligned} R_{SOR(\omega)} &= (I - \omega L)^{-1}((1-\omega)I + \omega U), \\ c_{SOR(\omega)} &= \omega(I - \omega L)^{-1}D^{-1}b. \end{aligned}$$

I ovdje vidimo da se dobra vrijednost za ω, ako postoji, može odrediti unaprijed za sve iteracije.

Nakon analize konvergencije ovih iterativnih metoda, pokazat ćemo da za neke klase matrica možemo naći optimalni izbor parametra ω koji ubrzava konvergenciju, i da vrijedi $\omega > 1$, što opravdava naziv OR u ovim metodama.

Zaustavljanje procesa

Kako zaustavljamo iteracije? Najlakši način je tzv. heuristička konvergencija. Unaprijed zadamo traženu točnost ε i prekidamo iteracije čim vrijedi

$$\|x^{(m+1)} - x^{(m)}\| \leq \varepsilon,$$

u nekoj pogodno odabranoj vektorskoj normi, na primjer, ∞-normi. To znači da u svakom koraku moramo računati i ovu normu, ali to obično nije pretjerano skupo, a može se isplatiti, ako "slučajno" greška naglo padne u nekoj iteraciji.

S druge strane, ako znamo vrijednost neke operatorske norme $\|R\|$ matrice iteracija (bez pretjerano računanja), s tim da je $\|R\| < 1$, onda unaprijed možemo izračunati potreban broj iteracija. Naime, u pripadnoj vektorskoj normi vrijedi ocjena

$$\|x^{(m+1)} - x\| \leq \frac{\|R\|}{1 - \|R\|}\,\|x^{(m+1)} - x^{(m)}\|, \tag{1.21}$$

kao u Banachovom teoremu o fiksnoj točki. Dokaz ove relacije je jednostavan:

$$\begin{aligned} x^{(m+1)} - x &= R(x^{(m)} - x) = R(x^{(m)} - x^{(m+1)}) + R(x^{(m+1)} - x) \\ (I - R)(x^{(m+1)} - x) &= -R(x^{(m+1)} - x^{(m)}) \\ x^{(m+1)} - x &= -(I - R)^{-1}R(x^{(m+1)} - x^{(m)}), \end{aligned}$$

jer znamo da je $I - R$ regularna matrica. Primjenom norme i ocjenama izlazi (1.50).

Relacija (1.50) pokazuje da je $\|x^{(m+1)} - x\| \leq \varepsilon$ osigurano ako je

$$\frac{\|R\|}{1 - \|R\|}\|x^{(m+1)} - x^{(m)}\| \leq \varepsilon\,.$$

Stoga ako je $\|R\|$ poznata i $\|R\| \lesssim 1$, dakle blizu je jedinici, onda zaustavljanje procesa uz uvjet $\|x^{(m+1)} - x^{(m)}\| \leq \varepsilon$ sigurno nije dobro.

Iz ocjene (1.50) dobivamo i

$$\|x^{(m+1)} - x\| \leq \frac{\|R\|^{m+1}}{1 - \|R\|} \|x^{(1)} - x^{(0)}\|,$$

a iz ove relacije možemo, nakon prve iteracije, izračunati potreban broj iteracija $m + 1$, tako da, do na greške zaokruživanja, vrijedi

$$\|x^{(m+1)} - x\| \leq \varepsilon.$$

Ipak, bolje je načiniti još nekoliko iteracija tako da $x^{(k)}$ bude bliže rješenju nego $x^{(1)}$, i da $\|x^{(k)} - x^{(k-1)}\|$ bude malo, pa onda ocijeniti još potreban broj iteracija $m - k$ pomoću formule

$$\|x^{(m)} - x\| \leq \frac{\|R\|^{m-k+1}}{1 - \|R\|} \|x^{(k)} - x^{(k-1)}\| \leq \varepsilon.$$

1.7 Konvergencija Jacobijeve i Gauss–Seidelove metode

U ovom odjeljku nabrojit ćemo, a u većini slučajeva i dokazati, koji su dovoljni uvjeti za konvergenciju pojedine metode. Prvi rezultat koristi pojam strogo dijagonalne matrice po recima.

Definicija 1.7.1. *Matrica A je strogo dijagonalno dominantna po recima ako vrijedi*

$$|a_{jj}| > \sum_{\substack{k=1 \\ k \neq j}}^{n} |a_{jk}| \qquad \textit{za svaki} \qquad 1 \leq j \leq n.$$

Matrica je A je strogo dijagonalno dominantna po stupcima ako je A^T strogo dijagonalno dominantna po recima.

Teorem 1.7.1. *Ako je A strogo dijagonalno dominantna matrica po recima, onda i Jacobijeva i Gauss–Seidelova metoda konvergiraju i vrijedi*

$$\|R_{GS}\|_\infty \leq \|R_{Jac}\|_\infty < 1.$$

Dokaz:

Prije dokaza, uočimo da relacija $\|R_{GS}\|_\infty \leq \|R_{Jac}\|_\infty$ znači da u jednom koraku Gauss–Seidelove metode možemo očekivati bolje približavanje prema rješenju x nego kod Jacobijeve metode. To ne znači nužno da će Gauss–Seidelova metoda konvergirati brže nego Jacobijeva

za bilo koju polaznu iteraciju. Naprosto, može se dogoditi da Jacobijeva metoda brže konvergira jer se je polazna (ili neka kasnija) iteracija našla u invarijantnom potprostoru za R_{Jac} koji odgovara malim vlastitim vrijednostima (vidjeti poglavlje o invarijantnim potprostorima).

Prvo pokažimo da je $\|R_{Jac}\|_\infty < 1$. Zbog stroge dijagonalne dominantnosti po recima, vrijedi

$$|a_{jj}| > \sum_{\substack{k=1 \\ k\neq j}}^{n} |a_{jk}|, \quad 1 \leq j \leq n\,.$$

Dijeljenjem sa $|a_{jj}|$, zaključujemo da je

$$1 > \max_j \frac{1}{|a_{jj}|} \sum_{\substack{k=1 \\ k\neq j}}^{n} |a_{jk}| = \max_j \sum_{\substack{k=1 \\ k\neq j}}^{n} \frac{|a_{jk}|}{|a_{jj}|} = \max_j \sum_{\substack{k=1 \\ k=1}}^{n} \left|[R_{Jac}]_{jk}\right| = \|R_{Jac}\|_\infty.$$

Sada nam preostaje dokazati da je $\|R_{GS}\|_\infty \leq \|R_{Jac}\|_\infty$.

Matrice iteracija možemo napisati u obliku

$$R_{Jac} = L + U \quad \text{i} \quad R_{GS} = (I - L)^{-1}U.$$

Želimo dokazati da vrijedi $\|R_{GS}\|_\infty \leq \|R_{Jac}\|_\infty$. Sjetimo se da je $\|X\|_\infty = \| \, |X| \, e\|_\infty$ pri čemu je e vektor sa svim komponentama jednakim 1, tj. $e = (1, \ldots, 1)^T$. Stoga trebamo pokazati da je

$$\| \, |R_{GS}| \, e\|_\infty \leq \| \, |R_{Jac}| \, e\|_\infty. \tag{1.22}$$

Ovu relaciju možemo lako pokazati ako dokažemo jaču tvrdnju, nejednakost na nivou komponenenata, $|R_{GS}| \cdot e \leq |R_{Jac}| \cdot e$ ili u drugom zapisu

$$|(I - L)^{-1}U| \cdot e \leq (|L| + |U|) \cdot e. \tag{1.23}$$

Doista, ako je svaka komponenta vektora x po modulu manja ili jednaka od odgovarajuće komponente vektora y, dakle $|x_i| \leq |y_i|$, za sve $1 \leq i \leq n$, tada je nužno $\|x\|_\infty \leq \|y\|_\infty$.

Krenimo slijeva i iskoristimo nejednakost $|X \cdot Y| \leq |X| \cdot |Y|$ koja slijedi iz činjenice da je apsolutna vrijednost skalarnog produkta dvaju vektora manja ili jednaka skalarnom produktu modula tih vektora, $|(x|y)| \leq (|x| \, | \, |y|)$. Ovdje nejednakost dolazi od primjene nejednakosti trokuta za apsolutnu vrijednost. Zato za lijevu stranu nejednakosti (1.23) vrijedi

$$|(I - L)^{-1}U| \leq |(I - L)^{-1}| \, |U| \, .$$

Sumiranjem unutar redaka lijeve i desne strane, dobivamo

$$|(I - L)^{-1}U| \cdot e \leq |(I - L)^{-1}| \, |U| \cdot e. \tag{1.24}$$

Matrica L je strogo donje trokutasta pa su sve njene vlastite vrijednosti nula. Isto vrijedi i za $-L$. Stoga matrica $I - L$ ima vlastite vrijednosti 1 pa je regularna. Pokažimo da možemo $(I - L)^{-1}$ razviti u konačni red

$$(I - L)^{-1} = I + L + L^2 + \cdots + L^{n-1}. \tag{1.25}$$

Da bi to dokazali dovoljno je pomnožiti lijevu i desnu stranu te jedakosti s $I - L$ i iskoristiti činjenicu da je $L^n = 0$.

Tvrdnju da je $L^n = 0$ je najlakše pokazati tako da se dokaže (npr. indukcijom po k) da je $L^k e_j = 0$ za sve $n - k + 1 \leq j \leq n$, gdje su e_j stupci od I_n. Nakon toga odmah slijedi da je $L^n e_j = 0$ za sve $1 \leq j \leq n$, a to znači da su svi stupci od L^n nula.

Iskoristimo li izraz za $(I - L)^{-1}$ iz relacije (1.25), lako dobijemo

$$|(I - L)^{-1}| = \left| \sum_{i=0}^{n-1} L^i \right| \leq \sum_{i=0}^{n-1} |L^i| \leq \sum_{i=0}^{n-1} |L|^i = (I - |L|)^{-1}.$$

Ovdje smo osim nejednakosti $|X^k| \leq |X|^k$ koja općenito vrijedi za kvadratnu matricu X i svako $k \geq 2$, koristili i $|X_1 + X_2 + \cdots + X_k| \leq |X_1| + |X_2| + \cdots + |X_k|$ koja vrijedi za bilo koje matrice X_1, X_2,...,X_k istog tipa i $k \geq 2$. Sada imamo

$$|(I - L)^{-1}U| \leq |(I - L)^{-1}|\,|U| \leq (I - |L|)^{-1}\,|U|,$$

pa opet sumiranjem unutar svakog retka matrice na lijevoj i desnoj strani, dobijemo

$$|(I - L)^{-1}U| \cdot e \leq (I - |L|)^{-1}\,|U| \cdot e.$$

Relacija (1.23) će vrijediti ako dokažemo nejednakost

$$(I - |L|)^{-1}\,|U| \cdot e \leq (|L| + |U|) \cdot e. \tag{1.26}$$

Uočimo da je i $|L|$ strogo donje trokutasta, pa su svi članovi u redu za $(I - |L|)^{-1}$ matrice s nenegativnim elementima. Stoga je $(I - |L|)^{-1}$ matrica s nenegativnim elementima. Da bi vrijedilo (1.26), dovoljno je pokazati da je

$$|U| \cdot e \leq (I - |L|)\,(|L| + |U|) \cdot e = (|L| + |U| - |L|^2 - |L|\,|U|) \cdot e,$$

odnosno

$$0 \leq (|L| - |L|^2 - |L|\,|U|) \cdot e = |L|\,(I - |L| - |U|) \cdot e. \tag{1.27}$$

Najme, kad pomnožimo matričnu nejednakost oblika $X \leq Y$ s matricom Z s lijeva (ili s desna) gdje su X, Y i Z matrice s nenegativnim elementima, tada će vrijediti $ZX \leq ZY$ ($XZ \leq YZ$). Mi smo pomnožili matričnu nejednakost $|U| \cdot e \leq (I - |L|)\,(|L| + |U|) \cdot e$ s lijeva sa $(I - |L|)^{-1}$. Stoga, ako vrijedi (1.27), vrijedit će i (1.26).

Ponovno, budući da su svi elementi $|L|$ nenegativni, prethodna nejednakost bit će ispunjena ako je

$$0 \leq (I - |L| - |U|) \cdot e,$$

odnosno

$$(|L| + |U|) \cdot e \leq e.$$

Budući da je $|R_{Jac}| = |L + U| = |L| + |U|$, jer se elementi L i U nigdje ne zbrajaju, onda je posljednja nejednakost ekvivalentna s

$$|R_{Jac}| \cdot e \leq e.$$

No, ova posljednja nejednakost vrijedi (čak sa strogom nejdnakosti u svakoj komponenti) jer je

$$\| \, |R_{Jac}| \cdot e\|_\infty = \|R_{Jac}\|_\infty < 1.$$

Time je dokaz gotov. Još se možemo zapitati da li $\|R_{Jac}\|_\infty < 1$ povlači strogu nejednakost u (1.23) pa zato i u (1.22). Odgovor je ne, jer nejednakost u (1.27) više ne mora biti stroga u svim komponentama. Najme, $|L|$ je donje trokutasta pa će sigurno prva komponenta od $|L| \, (I - |L| - |U|) \cdot e$ biti nula. ■

Analogni se rezultat, tj. da Jacobijeva i Gauss–Seidelova metoda konvergiraju, može se dobiti i za strogo dijagonalno dominantne matrice po stupcima, samo onda ulogu norme ∞ igra norma 1.

Uvjeti prethodnog teorema mogu se još malo oslabiti, do ireducibilnosti i slabe dijagonalne dominantnosti matrice A.

Definicija 1.7.2. *Matrica A je reducibilna ako i samo ako postoji matrica permutacije P takva da je*

$$PAP^T = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix},$$

pri čemu su A_{11} i A_{22} kvadratni blokovi reda manjeg od n, tj. PAP^T je blok gornjetrokutasta matrica. Matrica A je ireducibilna, ako nije reducibilna, tj. ako ne postoji matrica permutacije P za koju je PAP^T blok gornjetrokutasta matrica.

Da bismo lakše prepoznali ireducibilnu matricu, koristit ćemo vezu između matrica i grafova.

Definicija 1.7.3. *Matrici A odgovara usmjereni graf $G(A)$ s čvorovima $1, \ldots, n$. Usmjereni brid tog grafa od čvora i do čvora j postoji ako i samo ako je $a_{ij} \neq 0$.*

Definicija 1.7.4. *Usmjereni graf je jako povezan ako postoji put iz svakog čvora u svaki čvor. Komponenta jake povezanosti usmjerenog grafa je podgraf koji je jako povezan i ne može se povećati, a da ostane jako povezan.*

Najjednostavnija karakterizacija ireducibilnih matrica je preko jako povezanih pripadnih grafova.

Lema 1.7.1. *Matrica A je ireducibilna ako i samo ako je $G(A)$ jako povezan.*

Dokaz:

Ako je A reducibilna

$$PAP^T = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}$$

onda ne postoji "povratak" iz čvorova koji odgovaraju A_{22} u čvorove koji odgovaraju A_{11}, pa $G(A)$ nije jako povezan. Obratno, ako $G(A)$ nije jako povezan, postoji komponenta jake povezanosti koja ne sadrži sve čvorove grafa. Ako matricu prepermutiramo tako da čvorovi iz te komponente dođu na početak (u blok A_{11}), dobit ćemo traženi blok gornjetrokutasti oblik.■

Definicija 1.7.5. *Matrica A je slabo dijagonalno dominantna po recima ako za svaki j vrijedi*

$$|a_{jj}| \geq \sum_{\substack{k=1 \\ k \neq j}}^{n} |a_{kj}|,$$

a stroga nejednakost se javlja barem jednom.

Sada možemo oslabiti uvjete teorema 1.7.1., s tim da ovaj rezultat navodimo bez dokaza.

Teorem 1.7.2. *Ako je A ireducibilna i slabo dijagonalno dominantna matrica po recima, onda i Jacobijeva i Gauss–Seidelova metoda konvergiraju i vrijedi*

$$\|R_{GS}\|_\infty \leq \|R_{Jac}\|_\infty < 1.$$

Unatoč navedenim rezultatima da je pod nekim uvjetima Gauss–Seidelova metoda brža nego Jacobijeva, ne postoji nikakav generalni rezultat te vrste. Dapače, postoje nesimetrične matrice za koje Jacobijeva metoda konvergira, a Gauss–Seidelova ne, kao i matrice za koje Gauss–Seidelova metoda konvergira, a Jacobijeva divergira.

1.8 Konvergencija JOR i SOR metode*

Promotrimo sad konvergenciju relaksacijskih metoda JOR(ω) i SOR(ω) u ovisnosti o parametru ω. Obzirom na to da se ove metode za $\omega = 1$ svode na Jacobijevu, odnosno, Gauss–Seidelovu metodu, usput ćemo dobiti i neke rezultate o konvergenciji ovih osnovnih metoda.

Za početak, promatramo JOR(ω) metode, jer su bitno jednostavnije za analizu.

Teorem 1.8.1. *Ako Jacobijeva metoda za rješenje linearnog sustava $Ax = b$ konvergira za svaku početnu iteraciju $x^{(0)}$, onda za bilo koji $\omega \in (0, 1]$ konvergira i JOR(ω) metoda za svaku početnu iteraciju.*

Dokaz:

Prema teoremu 1.7, pretpostavka o konvergenciji Jacobijeve metode za svaku početnu iteraciju ekvivalentna je s činjenicom da vrijedi $\rho(R_{Jac}) < 1$. Neka su μ_j, $j = 1, \ldots, n$, svojstvene vrijednosti matrice R_{Jac}. Onda je $|\mu_j| < 1$ za sve j. Ako μ_j napišemo kao kompleksne brojeve u obliku $\mu_j = \alpha_j + i\beta_j$, uz $\alpha_j, \beta_j \in \mathbf{R}$, onda iz $|\mu_j| < 1$ slijedi $\alpha_j^2 + \beta_j^2 < 1$ i $|\alpha_j| < 1$.

Matrica iteracije u JOR(ω) metodi je

$$R_{JOR(\omega)} = (1-\omega)I + \omega R_{Jac} = (1-\omega)I + \omega(L+U),$$

što je polinom (stupnja 1) u funkciji od R_{Jac}, pa za svojstvene vrijednosti λ_j matrice $R_{JOR(\omega)}$ vrijedi

$$\lambda_j = 1 - \omega + \omega\mu_j, \quad j = 1, \ldots, n,$$

odakle izlazi

$$|\lambda_j|^2 = |(1-\omega+\omega\alpha_j) + i\omega\beta_j|^2 = (1-\omega+\omega\alpha_j)^2 + \omega^2\beta_j^2.$$

Ako je $0 < \omega \leq 1$, onda vrijedi ocjena

$$\begin{aligned} |\lambda_j|^2 &\leq (1-\omega)^2 + 2\omega(1-\omega)|\alpha_j| + \omega^2(\alpha_j^2 + \beta_j^2) \\ &< (1-\omega)^2 + 2\omega(1-\omega) + \omega^2 = (1-\omega+\omega)^2 = 1, \end{aligned}$$

odakle slijedi $\rho(R_{JOR(\omega)}) < 1$, a to znači da JOR($\omega$) metoda konvergira za svaku početnu iteraciju. ■

Prethodni rezultat daje samo uvjetnu konvergenciju, u smislu da ako metoda konvergira za $\omega = 1$, onda konvergira i za sve ω iz skupa $(0, 1]$. Precizniju, ali negativnu informaciju daje sljedeći rezultat.

Teorem 1.8.2. *Vrijedi $\rho(R_{JOR(\omega)}) \geq |\omega - 1|$, pa je $0 < \omega < 2$ nužan uvjet za konvergenciju JOR metode.*

Dokaz:

Znamo da je trag matrice jednak zbroju svih njezinih svojstvenih vrijednosti. Promotrimo trag matrice

$$R_{JOR(\omega)} = (1-\omega)I + \omega R_{Jac} = (1-\omega)I + \omega(L+U).$$

Drugi član $\omega(L+U)$ ima nul-dijagonalu, pa samo prvi član daje trag

$$\text{trag}\,(R_{JOR(\omega)}) = n(1-\omega) = \sum_{j=1}^{n} \lambda_j.$$

Dobivamo da je

$$n|1-\omega| \leq \sum_{j=1}^{n} |\lambda_j| \leq n\rho(R_{JOR(\omega)}),$$

odakle slijedi $\rho(R_{JOR(\omega)}) \geq |\omega - 1|$. Dakle, za konvergenciju JOR(ω) metode mora vrijediti $|\omega - 1| < 1$, ili $0 < \omega < 2$. U protivnom, za neke $x^{(0)}$ metoda divergira. ■

Za simetrične (hermitske) i pozitivno definitne matrice A, možemo dobiti i pozitivnu informaciju o garantiranoj konvergenciji JOR metode.

Teorem 1.8.3. *Neka je A simetrična (hermitska) i pozitivno definitna matrica i neka za svojstvene vrijednosti μ_j matrice R_{Jac} Jacobijeve metode vrijedi*

$$\mu_j < 1, \quad j = 1, \ldots, n.$$

Onda JOR(ω) metoda konvergira za sve parametre ω za koje vrijedi

$$0 < \omega < \frac{2}{1 - \mu} \leq 2, \tag{1.28}$$

gdje je $\mu := \min_j \mu_j$ najmanja svojstvena vrijednost matrice R_{Jac}.

Dokaz:

Prvo uočimo da je $R_{Jac} = D^{-1}(\widetilde{L} + \widetilde{L}^*) = D^{-1/2}(D^{-1/2}(\widetilde{L} + \widetilde{L}^*)D^{-1/2})D^{1/2}$ slična simetričnoj (hermitskoj) matrici $D^{-1/2}(\widetilde{L} + \widetilde{L}^*)D^{-1/2}$, pa su joj svojstvene vrijednosti μ_j realne. Već smo vidjeli da za svojstvene vrijednosti λ_j matrice $R_{JOR(\omega)}$ vrijedi

$$\lambda_j = 1 - \omega + \omega\mu_j, \quad j = 1, \ldots, n,$$

pa je $1 - \lambda_j = \omega(1 - \mu_j)$, odakle slijedi da je $|\lambda_j| < 1$ za sve j, ako i samo ako vrijedi

$$0 < \omega(1 - \mu_j) < 2, \quad j = 1, \ldots, n.$$

Iz $\mu_j < 1$ slijedi $1 - \mu_j > 0$, pa mora biti $\omega > 0$. S druge strane, iz $\omega(1 - \mu_j) \leq \omega(1 - \mu)$ izlazi uvjet $\omega(1 - \mu) < 2$, pa je $\omega < 2/(1 - \mu)$. Na kraju, iz trag $(R_{Jac}) = 0$ slijedi da najmanja svojstvena vrijednost μ zadovoljava $\mu \leq 0$, što dokazuje (1.28). ■

Ovaj rezultat **ne** znači da Jacobijeva metoda konvergira jer se može dogoditi da je $\mu \leq -1$, pa u (1.28) dobivamo $\omega < 1$. Međutim, ako Jacobijeva metoda konvergira, onda konvergira i JOR, s tim da u (1.28) možemo uzeti i $\omega > 1$.

Korolar 1.8.1. *Neka je A simetrična (hermitska) pozitivno definitna matrica i pretpostavimo da Jacobijeva metoda konvergira. Onda konvergira i JOR(ω) metoda za sve parametre ω za koje vrijedi* (1.28) *i u toj relaciji je $2/(1 - \mu) > 1$.*

Dokaz:

Iz konvergencije Jacobijeve metode slijedi $-1 < \mu_j < 1$, za sve j, pa vrijedi zaključak prethodnog teorema. Osim toga, vrijedi i $\mu > -1$, pa je $1 - \mu < 2$, što pokazuje da je gornja granica za ω u (1.28) veća od 1. ■

Ovo je pojačanje teorema 1.8.1. za simetrične (hermitske) pozitivno definitne matrice. Nažalost, gornju granicu za dozvoljeni $\omega > 1$ nije lako naći. Ako znamo $\rho(R_{Jac})$, možemo koristiti ocjenu $-1 < -\rho(R_{Jac}) \leq \mu$ i birati ω tako da je

$$1 < \omega < \frac{2}{1+\rho(R_{Jac})} \leq \frac{2}{1-\mu} \leq 2.$$

Međutim, nije jasno da ćemo takvim izborom parametra ω ubrzati konvergenciju iterativne metode. Obzirom na to da za SOR metodu možemo dobiti jače rezultate, JOR metoda se relativno rijetko koristi u praksi.

Prvi rezultat za SOR metodu je isti nužni uvjet konvergencije kao i kod JOR metode.

Teorem 1.8.4. *Vrijedi $\rho(R_{SOR(\omega)}) \geq |\omega - 1|$, pa je $0 < \omega < 2$ nužan uvjet za konvergenciju SOR metode.*

Dokaz:

Znamo da je determinanta matrice jednaka produktu svih njezinih svojstvenih vrijednosti. Izračunajmo determinantu matrice

$$R_{SOR(\omega)} = (I - \omega L)^{-1}((1-\omega)I + \omega U)$$

koja je produkt trokutastih matrica. Iskoristimo Binet–Cauchyjev teorem i činjenicu da su L i U strogo trokutaste matrice. Zbog toga, samo dijagonale, tj. članovi s I ulaze u determinante, pa je

$$\begin{aligned} \det R_{SOR(\omega)} &= \det(I - \omega L)^{-1} \cdot \det((1-\omega)I + \omega U) \\ &= \det I \cdot \det((1-\omega)I) = (1-\omega)^n. \end{aligned}$$

S druge strane je

$$\det R_{SOR(\omega)} = \prod_{j=1}^{n} \lambda_j(R_{SOR(\omega)}),$$

pa iz $|\lambda_j(R_{SOR(\omega)})| \leq \rho(R_{SOR(\omega)})$ dobivamo

$$|1-\omega|^n = \prod_{j=1}^{n} |\lambda_j(R_{SOR(\omega)})| \leq (\rho(R_{SOR(\omega)}))^n,$$

odakle slijedi $\rho(R_{SOR(\omega)}) \geq |\omega - 1|$. Dakle, za konvergenciju SOR(ω) metode mora vrijediti $|\omega - 1| < 1$, ili $0 < \omega < 2$. U protivnom, metoda sigurno divergira, bar za neke početne vektore $x^{(0)}$. ■

Ako je A simetrična (hermitska) i pozitivno definitna, uvjet $0 < \omega < 2$ je i dovoljan za konvergenciju.

Teorem 1.8.5. *Ako je A simetrična (hermitska) i pozitivno definitna matrica tada je*

$$\rho(R_{SOR(\omega)}) < 1 \quad za \quad 0 < \omega < 2,$$

pa SOR(ω) konvergira. Posebno, uzimajući $\omega = 1$, slijedi da i Gauss–Seidelova metoda konvergira.

Dokaz:

Da bismo skratili pisanje, označimo s $R := R_{SOR(\omega)}$. Trebamo dokazati da za sve svojstvene vrijednosti matrice R vrijedi $|\lambda_j(R)| < 1$, tj. da one leže unutar jediničnog kruga u kompleksnoj ravnini. Da bismo to dokazali, trebamo iskoristiti da su svojstvene vrijednosti od A na pozitivnoj realnoj osi. Dokaz se sastoji od dva koraka. U prvom, prebacujemo problem iz jediničnog kruga u desnu otvorenu poluravninu $\operatorname{Re} z > 0$, gdje je lakše iskoristiti pozitivnu definitnost matrice A.

Za prvi korak koristimo razlomljene linearne transformacije. Lako se provjerava da takva (tzv. Möbiusova) transformacija oblika

$$\zeta(z) := \frac{1+z}{1-z}$$

bijektivno preslikava unutrašnjost jediničnog kruga $|z| < 1$ na desnu otvorenu poluravninu $\operatorname{Re}\zeta > 0$. Na isti način želimo transformirati svojstvene vrijednosti matrice R. Dakle, trebamo gledati matricu $\zeta(R)$. Obzirom na to da množenje matrica ne mora biti komutativno, pokazat će se da je zgodnije inverz pisati kao lijevi faktor (sami pogledajte put kroz dokaz ako inverz pišemo s desne strane). Definiramo matricu

$$S := (I-R)^{-1}(I+R). \tag{1.29}$$

Tada za svojstvene vrijednosti vrijedi $\lambda_j(S) = \zeta(\lambda_j(R))$, pa S ima svojstvene vrijednosti u desnoj otvorenoj poluravnini ako i samo ako R ima svojstvene vrijednosti unutar jediničnog kruga. Nažalost, to vrijedi samo ako je S korektno definirana, tj. ako je $I-R$ regularna. To još ne znamo, pa pokušajmo doći do relacije za S koja je uvijek korektna.

Polazna matrica A je po pretpostavci regularna. Da bismo dobili iterativnu metodu s matricom iteracije R (sasvim općenito), koristimo rastav ili cijepanje matrice $A = M - K$, pa ako je M regularna, onda je

$$R = M^{-1}K = M^{-1}(M-A) = I - M^{-1}A.$$

Tada je $I - R = M^{-1}A$ očito regularna (produkt regularnih), a $I + R = 2I - M^{-1}A$. Za S dobivamo

$$S = (I-R)^{-1}(I+R) = A^{-1}M(2I - M^{-1}A) = 2A^{-1}M - I.$$

Dakle, ako definiramo

$$S := 2A^{-1}M - I = A^{-1}(2M - A), \tag{1.30}$$

onda je S korektno definirana za bilo koju regularnu matricu A, čak i kad M nije regularna. Za nastavak dokaza treba iskoristiti ostale pretpostavke na A i pogledati kad svojstvene vrijednosti matrice S leže u desnoj otvorenoj poluravnini, u ovisnosti o ω u SOR metodi.

Neka je (λ, x) bilo koji svojstveni par od S, tj. $Sx = \lambda x$. Iz (1.30), množenjem s A, onda vrijedi i

$$(2M - A)x = ASx = \lambda Ax.$$

Množenjem s x^* slijeva dobivamo

$$x^*(2M - A)x = \lambda x^* Ax.$$

Napišimo adjungiranu (ozvjezdičenu) jednadžbu, iskoristivši simetričnost (hermitičnost) matrice $A = A^*$

$$x^*(2M^* - A)x = \bar{\lambda} x^* Ax,$$

i zbrojimo ih. Dijeljenjem s 2 dobivamo

$$x^*(M + M^* - A)x = \frac{\lambda + \bar{\lambda}}{2}\, x^* Ax = (\operatorname{Re}\lambda) x^* Ax.$$

Po pretpostavci je $x \neq 0$ (svojstveni vektor od S), pa iz pozitivne definitnosti od A slijedi $x^*Ax > 0$. Dijeljenjem dobivamo

$$\operatorname{Re}\lambda = \frac{x^*(M + M^* - A)x}{x^*Ax}, \tag{1.31}$$

pa je $\operatorname{Re}\lambda > 0$ ako i samo ako je brojnik pozitivan.

Sad iskoristimo da matrica M kod rastava matrice A u SOR(ω) metodi ima oblik

$$M = \omega^{-1}(D - \omega\widetilde{L}) = \omega^{-1}D - \widetilde{L},$$

pa je M korektno definirana za $\omega \neq 0$. Osim toga, iz $A = A^*$ slijedi $\widetilde{U} = \widetilde{L}^*$, ili $A = D - \widetilde{L} - \widetilde{L}^*$. Koristeći $D = D^*$, dobivamo da je

$$M + M^* - A = (\omega^{-1}D - \widetilde{L}) + (\omega^{-1}D - \widetilde{L}^*) - (D - \widetilde{L} - \widetilde{L}^*) = (2\omega^{-1} - 1)D.$$

Dijagonala D simetrične (hermitske) pozitivno definitne matrice A je i sama simetrična (hermitska) pozitivno definitna matrica. Na kraju, iz

$$x^*(M + M^* - A)x = (2\omega^{-1} - 1)x^* Dx$$

dobivamo da je brojnik u (1.31) pozitivan, ako i samo ako je $2\omega^{-1} - 1 > 0$ ili $0 < \omega < 2$.

Dakle, za matricu S u SOR metodi vrijedi $\operatorname{Re}\lambda_j(S) > 0$ za sve j, ako (i samo ako) je $0 < \omega < 2$. Nažalost, još uvijek ne možemo iskoristiti (1.29). No, inverz funkcije ζ

$$z(\zeta) = \frac{\zeta - 1}{\zeta + 1}$$

preslikava desnu otvorenu poluravninu na unutrašnjost jediničnog kruga. Rješavanjem jednadžbe (1.29) po S, očekujemo da je

$$R = (S - I)(S + I)^{-1}.$$

Zaista, ako je $\operatorname{Re} \lambda_j(S) > 0$ za sve j, onda je $S + I$ regularna, pa iz (1.30) slijedi

$$(S - I)(S + I)^{-1} = (2A^{-1}M - 2I)\,(2A^{-1}M)^{-1} = I - M^{-1}A = R,$$

a onda iz veze spektara vrijedi $|\lambda_j(R)| < 1$ za sve j

$$\begin{aligned} |\lambda_j(R)| &= \left|\frac{\lambda_j(S) - 1}{\lambda_j(S) + 1}\right| = \left|\frac{(\operatorname{Re}\lambda_j(S) - 1)^2 + (\operatorname{Im}\lambda_j(S))^2}{(\operatorname{Re}\lambda_j(S) + 1)^2 + (\operatorname{Im}\lambda_j(S))^2}\right|^{1/2} \\ &= \left|\frac{(\operatorname{Re}\lambda_j(S))^2 - 2\operatorname{Re}\lambda_j(S) + 1 + (\operatorname{Im}\lambda_j(S))^2}{(\operatorname{Re}\lambda_j(S))^2 + 2\operatorname{Re}\lambda_j(S) + 1 + (\operatorname{Im}\lambda_j(S))^2}\right|^{1/2} < 1. \end{aligned}$$

Lako se vidi da vrijedi i obrat, tj. iz $|\lambda_j(R)| < 1$ za sve j, koristeći (1.29), dobivamo $\operatorname{Re}\lambda_j(S) > 0$ za sve j. ■

Očito je da smo prethodni dokaz mogli provesti i mnogo brže ili kraće. Njegov deduktivni dio ima samo dva bitna koraka:

(1) definiramo $S = A^{-1}(2M - A)$ i pokažemo da je $\operatorname{Re}\lambda_j(S) > 0$ za sve j,

(2) pokažemo da je $R = (S - I)(S + I)^{-1}$, a zatim da je $|\lambda_j(R)| < 1$ za sve j.

Međutim, oblik matrice S i njezina veza s R nisu pali "s neba", niti su rezultat pogađanja. U pozadini cijele konstrukcije je lijepo matematičko opravdanje koje smo posebno željeli naglasiti.

Iz posljednja dva teorema odmah izlazi sljedeći rezultat.

Korolar 1.8.2. *Ako je A simetrična (hermitska) pozitivno definitna matrica, onda SOR(ω) konvergira ako i samo ako je $0 < \omega < 2$.*

U usporedbi s korolarom 1.8.1. za JOR metodu, ovo je bitno proširenje i pojačanje. Ovdje dobivamo bezuvjetnu konvergenciju, veći raspon za ω i maksimalnu moguću gornju granicu 2, koja ne ovisi o λ.

1.9 Optimalni izbor relaksacijskog parametra*

Pokažimo još da se dobrim izborom relaksacijskog parametra ω može postići bitno ubrzanje konvergencije iteracija u SOR(ω) metodi, bar za neke klase matrica.

Dokazat ćemo klasični rezultat koji daje optimalni izbor parametra ω za klasu tzv. konzistentno poredanih matrica. Dokazao ga je D. M. Young, još 1950. godine, a ima veliku praktičnu vrijednost jer pokriva matrice koje se javljaju u diskretizaciji nekih parcijalnih diferencijalnih jednadžbi, poput Poissonove.

Za početak, moramo definirati potrebne pojmove, koji se opet oslanjaju na vezu između matrica i pripadnih grafova.

Definicija 1.9.1. *Matrica A ima svojstvo (A) ako postoji matrica permutacije P takva da vrijedi*

$$PAP^T = \begin{bmatrix} D_1 & A_{12} \\ A_{21} & D_2 \end{bmatrix},$$

gdje su D_1 i D_2 dijagonalne matrice. Drugim riječima, u pripadnom grafu $G(A)$ čvorovi su podijeljeni u dva disjunktna skupa S_1 i S_2, s tim da ne postoji brid koji veže dva različita čvora iz istog skupa S_i. Ako ignoriramo bridove koji vežu čvor sa samim sobom, onda bridovi vežu samo čvorove iz različitih skupova. Takav se graf zove bipartitni graf.

Ako matrica A ima svojstvo (A), onda rastav ili cijepanje matrice PAP^T za iterativnu metodu ima posebnu strukturu. Tada je

$$PAP^T = \begin{bmatrix} D_1 & A_{12} \\ A_{21} & D_2 \end{bmatrix} = \begin{bmatrix} D_1 & \\ & D_2 \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ -A_{21} & 0 \end{bmatrix} - \begin{bmatrix} 0 & -A_{12} \\ 0 & 0 \end{bmatrix} := D - \widetilde{L} - \widetilde{U}.$$

Ako je D regularna, onda možemo korektno definirati sve opisane iterativne metode za matricu PAP^T, a iz prethodne strukture dokazat ćemo posebna svojstva tih iterativnih metoda. U praksi se obično pretpostavlja da je A sa svojstvom (A) već dovedena u ovaj oblik odgovarajućom permutacijom P, tj. da polazna matrica A ima ovu strukturu.

U nastavku pretpostavljamo da je D regularna i koristimo standardne oznake $L = D^{-1}\widetilde{L}$ i $U = D^{-1}\widetilde{U}$.

Definicija 1.9.2. *Neka je $\alpha \neq 0$. Definirajmo familiju matrica*

$$R_{Jac}(\alpha) = \alpha D^{-1}\widetilde{L} + \frac{1}{\alpha} D^{-1}\widetilde{U} = \alpha L + \frac{1}{\alpha} U. \tag{1.32}$$

Vidimo da je $R_{Jac}(1) = R_{Jac}$ matrica iteracije u Jacobijevoj metodi.

Matrice $R_{Jac}(\alpha)$ za $\alpha \neq 1$ nemaju direktnu interpretaciju kao matrice iteracije u nekoj od standardnih iterativnih metoda koje smo opisali.

Propozicija 1.9.1. *Za matrice A sa svojstvom (A), svojstvene vrijednosti matrica $R_{Jac}(\alpha)$ ne ovise o α, s tim da D, L i U dobivamo iz rastava matrice PAP^T.*

Dokaz:

Po definiciji je

$$\begin{aligned} R_{Jac}(\alpha) &= \alpha L + \frac{1}{\alpha}U = D^{-1}\left(\alpha\widetilde{L} + \frac{1}{\alpha}\widetilde{U}\right) \\ &= \begin{bmatrix} D_1^{-1} & \\ & D_2^{-1} \end{bmatrix}\left(\alpha\begin{bmatrix} 0 & 0 \\ -A_{21} & 0 \end{bmatrix} + \frac{1}{\alpha}\begin{bmatrix} 0 & -A_{12} \\ 0 & 0 \end{bmatrix}\right) \\ &= -\begin{bmatrix} 0 & \frac{1}{\alpha}D_1^{-1}A_{12} \\ \alpha D_2^{-1}A_{21} & 0 \end{bmatrix}. \end{aligned}$$

Ovu relaciju možemo napisati i u obliku

$$R_{Jac}(\alpha) = \begin{bmatrix} I & \\ & \alpha I \end{bmatrix}\left(-\begin{bmatrix} 0 & D_1^{-1}A_{12} \\ D_2^{-1}A_{21} & 0 \end{bmatrix}\right)\begin{bmatrix} I & \\ & \alpha I \end{bmatrix}^{-1} = \Delta_\alpha R_{Jac}\Delta_\alpha^{-1},$$

gdje je $\Delta_\alpha = \mathrm{diag}(I, \alpha I)$ dijagonalna matrica u kojoj dimenzije blokova odgovaraju dimenzijama blokova D_1 i D_2. Zbog $\alpha \neq 0$, očito je Δ_α regularna. Zaključujemo da su $R_{Jac}(\alpha)$ i R_{Jac} slične matrice, a to povlači da imaju iste svojstvene vrijednosti, tj. svojstvene vrijednosti $R_{Jac}(\alpha)$ ne ovise o α. ■

Svojstvo (A) iz definicije 1.9.1. je geometrijsko ili grafovsko svojstvo matrice. Posljedica toga je ovo algebarsko svojstvo invarijantnosti, kojeg ćemo bitno koristiti u nastvaku, pa mu dajemo posebno ime.

Definicija 1.9.3. *Neka je A proizvoljna matrica takva da je*

$$A = D - \widetilde{L} - \widetilde{U}$$

s tim da je D regularna i

$$R_{Jac}(\alpha) = \alpha D^{-1}\widetilde{L} + \frac{1}{\alpha}D^{-1}\widetilde{U} = \alpha L + \frac{1}{\alpha}U.$$

Ako svojstvene vrijednosti matrice $R_{Jac}(\alpha)$ ne ovise o α, onda kažemo da A ima konzistentan poredak (engl. consistent ordering).

Ako A ima svojstvo (A), onda PAP^T ima konzistentan poredak. Obratno ne vrijedi, tj. ako matrica ima konzistentan poredak, ne mora imati svojstvo (A).

Primjer 1.9.1. *Blok trodijagonalne matrice oblika*

$$\begin{bmatrix} D_1 & U_1 & & \\ L_1 & \ddots & \ddots & \\ & \ddots & \ddots & U_{m-1} \\ & & L_{m-1} & D_m \end{bmatrix}$$

imaju konzistentan poredak kad su D_i regularne dijagonalne matrice bilo kojih redova, za bilo koji blok red m. Pokažite to!

Međutim, za $m > 2$, ako su vandijagonalni blokovi netrivijalni, ove matrice nemaju svojstvo (A). Takvim matricama odgovaraju tzv. slojeviti grafovi, u kojima čvorove možemo podijeliti u m disjunktnih skupova — slojeva, tako da bridovi idu samo između čvorova koji su u različitim, ali susjednim slojevima. Kao i prije, ignoriramo bridove iz čvora u samog sebe. U ovom kontekstu, bipartitni graf ima samo 2 sloja.

Za matrice koje imaju konzistentan poredak postoje jednostavne formule koje vežu svojstvene vrijednosti matrica R_{Jac}, R_{GS} i $R_{SOR(\omega)}$.

Teorem 1.9.1. *Ako matrica A ima konzistentan poredak i ako je $\omega \neq 0$, onda vrijedi:*

(*a*) *Svojstvene vrijednosti matrice $R_{Jac}(\alpha)$ dolaze u $\pm$ parovima. Preciznije, ako je $\mu \neq 0$ svojstvena vrijednost od R_{Jac} multipliciteta ν, onda je i $-\mu$ svojstvena vrijednost od R_{Jac} multipliciteta ν.*

(*b*) *Ako je μ svojstvena vrijednost od R_{Jac} i ako λ zadovoljava jednadžbu*

$$(\lambda + \omega - 1)^2 = \lambda\omega^2\mu^2, \tag{1.33}$$

onda je λ svojstvena vrijednost od $R_{SOR(\omega)}$.

(*c*) *Obratno, ako je $\lambda \neq 0$ svojstvena vrijednost od $R_{SOR(\omega)}$, onda je μ iz* (1.33) *svojstvena vrijednost od R_{Jac}.*

Dokaz:

Prije početka dokaza uočimo da predznak od μ ne igra ulogu u (1.33), što je u suglasnosti s prvom tvrdnjom. Također, relacija (1.33) generira isti broj vrijednosti za λ i μ, uz uvjet da je $\lambda \neq 0$.

(*a*) Po definiciji 1.9.3., ako A ima konzistentan poredak, onda svojstvene vrijednosti od $R_{Jac}(\alpha)$ ne ovise o α. Posebno, to znači da matrice $R_{Jac} = R_{Jac}(1)$ i $R_{Jac}(-1)$ imaju iste svojstvene vrijednosti.

S druge strane, prema definiciji (1.32) za $R_{Jac}(-1)$ izlazi

$$R_{Jac}(-1) = -D^{-1}\widetilde{L} - D^{-1}\widetilde{U} = -(L + U) = -R_{Jac}(1) = -R_{Jac},$$

pa matrice R_{Jac} i $-R_{Jac}$ istovremeno moraju imati iste i suprotne svojstvene vrijednosti. To je moguće ako i samo ako svojstvene vrijednosti μ dolaze u $\pm$ parovima s istim multiplicitetom, čim je $\mu \neq 0$.

(*b*) Neka je μ svojstvena vrijednost od R_{Jac} i pretpostavimo da λ zadovoljava jednadžbu (1.33). Ako je $\lambda = 0$, onda u (1.33) mora biti $\omega = 1$, tj. SOR metoda se svodi na Gauss–Seidelovu metodu. Tada je $R_{SOR(1)} = R_{GS} = (I - L)^{-1}U$, a znamo da je R_{GS} singularna, jer je U strogo gornja trokutasta. Dakle, $\lambda = 0$ je tada svojstvena vrijednost za $R_{SOR(1)}$.

Pretpostavimo sad da je $\lambda \neq 0$. Onda, zbog $\omega \neq 0$, iz jednadžbe (1.33) možemo izračunati μ

$$\mu = \frac{\lambda + \omega - 1}{\sqrt{\lambda}\omega}. \tag{1.34}$$

Matrica iteracije u SOR(ω) metodi ima oblik

$$R_{SOR(\omega)} = (I - \omega L)^{-1}((1-\omega)I + \omega U).$$

Pogledajmo vrijednost karakterističnog polinoma $p(\lambda) = \det(\lambda I - R_{SOR(\omega)})$ ove matrice u točki λ. Da bismo se riješili inverza u $R_{SOR(\omega)}$, uočimo da je $\det(I - \omega L) = 1$, jer je L strogo donja trokutasta. Zbog toga, po Binet–Cauchyjevom teoremu vrijedi

$$\begin{aligned} \det(\lambda I - R_{SOR(\omega)}) &= \det((I - \omega L)(\lambda I - R_{SOR(\omega)})) \\ &= \det(\lambda(I - \omega L) - ((1-\omega)I + \omega U)) \\ &= \det((\lambda + \omega - 1)I - \omega\lambda L - \omega U). \end{aligned} \tag{1.35}$$

Zadnja dva člana želimo svesti na oblik $R_{Jac}(\alpha)$ za neki α, izlučivanjem odgovarajućeg faktora. Imamo

$$\omega\lambda L + \omega U = \sqrt{\lambda}\omega\left(\sqrt{\lambda}L - \frac{1}{\sqrt{\lambda}}U\right) = \sqrt{\lambda}\omega\, R_{Jac}(\sqrt{\lambda}),$$

pa je $\alpha = \sqrt{\lambda}$, s tim da smo iskoristili $\lambda \neq 0$. Kad u (1.35) izlučimo isti faktor $\sqrt{\lambda}\omega$, uvrstimo ovu relaciju i uvažimo da je $R_{Jac}(\sqrt{\lambda}) = R_{Jac}(1) = R_{Jac}$, dobivamo

$$\begin{aligned} \det(\lambda I - R_{SOR(\omega)}) &= \det\left(\sqrt{\lambda}\omega\left(\left(\frac{\lambda + \omega - 1}{\sqrt{\lambda}\omega}\right)I - R_{Jac}\right)\right) \\ &= \left(\sqrt{\lambda}\omega\right)^n \det\left(\left(\frac{\lambda + \omega - 1}{\sqrt{\lambda}\omega}\right)I - R_{Jac}\right). \end{aligned}$$

Vidimo da je faktor uz I upravo jednak μ iz (1.34), pa prethodna relacija postaje

$$\det(\lambda I - R_{SOR(\omega)}) = \left(\sqrt{\lambda}\omega\right)^n \det(\mu I - R_{Jac}). \tag{1.36}$$

Ako je μ svojstvena vrijednost od R_{Jac}, onda je desna strana jednaka nuli, pa to vrijedi i za lijevu stranu, tj. λ je svojstvena vrijednost od $R_{SOR(\omega)}$.

(c) Relaciju (1.36) smo izveli baš uz pretpostavku da je $\lambda \neq 0$, pa odmah slijedi tvrdnja. ■

Korolar 1.9.1. *Ako A ima konzistentan poredak, tada je*

$$\rho(R_{GS}) = (\rho(R_{Jac}))^2,$$

što znači da Gauss–Seidelova metoda konvergira dvostruko brže nego Jacobijeva metoda (ako barem jedna od njih kovergira).

Dokaz:

Izaberemo li u prethodnom teoremu $\omega = 1$, onda je SOR metoda baš Gauss–Seidelova metoda, pa za taj ω relacija (1.33) glasi

$$\lambda^2 = \lambda\mu^2,$$

ili $\lambda = \mu^2$. Budući da to vrijedi za svaku svojstvenu vrijednost, onda to vrijedi i za spektralni radijus. ■

Odavde odmah slijedi da za konzistentno poredane matrice Gauss–Seidelova metoda konvergira ako i samo ako konvergira i Jacobijeva metoda. Vidjet ćemo da slično vrijedi i za SOR metodu. Međutim, dobit ćemo i puno jači rezultat koji nam kaže kako treba izabrati parametar ω za ubrzanje konvergencije u SOR metodi.

Na početku ovog poglavlja vidjeli smo da brzina konvergencije iterativne metode ovisi o spektralnom radijusu $\rho(R)$ matrice iteracija R. Manji $\rho(R)$, općenito, osigurava i bržu konvergenciju, jer vrijedi ocjena

$$\|x^{(m+1)} - x\|_* \leq \rho(R)\|x^{(m)} - x\|_*.$$

Za neke početne vektore i u nekim iteracijama možemo dobiti i manju grešku, ali znamo da je ova ocjena dostižna. Dakle, da bismo globalno ubrzali konvergenciju metode treba dobiti što manji spektralni radijus $\rho(R)$. U tom smislu, kod SOR(ω) metode, one vrijednosti parametra ω za koje je $\rho(R_{SOR(\omega)})$ globalno najmanji, zovemo **optimalnim** relaksacijskim parametrima i označavamo s ω_{opt}.

Uz blage dodatne uvjete i malo truda, iz relacije (1.33) možemo dobiti i taj optimalni izbor parametra ω_{opt} koji minimizira $R_{SOR(\omega)}$, tako da on ovisi samo o spektralnom radijusu $\rho(R_{Jac})$ u Jacobijevoj metodi.

Teorem 1.9.2. *Pretpostavimo da matrica A ima konzistentan poredak i da matrica R_{Jac} u Jacobijevoj metodi ima samo realne svojstvene vrijednosti. Onda SOR(ω) metoda konvergira za bilo koji početni vektor ako i samo ako je $\mu := \rho(R_{Jac}) < 1$ (tj. Jacobijeva metoda konvergira) i vrijedi $0 < \omega < 2$. Dodatno, za $\mu < 1$ onda vrijedi i*

$$\begin{aligned} \omega_{\text{opt}} &= \frac{2}{1+\sqrt{1-\mu^2}}, \\ \rho(R_{SOR(\omega_{\text{opt}})}) &= \omega_{\text{opt}} - 1 = \frac{\mu^2}{(1+\sqrt{1-\mu^2})^2} = \frac{1-\sqrt{1-\mu^2}}{1+\sqrt{1-\mu^2}}, \end{aligned}$$

a za sve $\omega \in (0,2)$ vrijedi

$$\rho(R_{SOR(\omega)}) = \begin{cases} 1-\omega+\frac{1}{2}\omega^2\mu^2 + \omega\mu\sqrt{1-\omega+\frac{1}{4}\omega^2\mu^2}, \text{ za } 0 < \omega \leq \omega_{\text{opt}}, \\ \omega - 1, \text{ za } \omega_{\text{opt}} \leq \omega \leq 2. \end{cases}$$

Dokaz:

Matrica A ima konzistentan poredak, pa možemo iskoristiti teorem 1.9.1.(*b*) da iz svojstvenih vrijednosti μ_j matrice R_{Jac} izračunamo svojstvene vrijednosti λ_j matrice $R_{SOR(\omega)}$. Relacija (1.33) daje vezu

$$(\lambda_j + \omega - 1)^2 = \lambda_j \omega^2 \mu_j^2,$$

što možemo napisati kao kvadratnu jednadžbu za λ_j

$$\lambda_j^2 - 2\left(1 - \omega + \frac{1}{2}(\omega\mu_j)^2\right)\lambda_j + (\omega - 1)^2 = 0. \tag{1.37}$$

Prema tvrdnji (*a*) teorema 1.9.1., ako $\mu_j = 0$ ima multiplicitet m, onda pripadni $\lambda_j = 1 - \omega$ ima isti multiplicitet m. Osim toga, svojstvene vrijednosti $\mu_j \neq 0$ dolaze u $\pm$ parovima, pa u rješavanju jednadžbe (1.37) možemo ignorirati predznak od μ_j, jer svaki $\pm$ par daje dvije svojstvene vrijednosti λ_j.

Sad iskoristimo pretpostavku da R_{Jac} ima samo realne svojstvene vrijednosti μ_j, pa jednadžba (1.37) ima **realne** koeficijente i kod rješavanja možemo gledati samo $\mu_j > 0$.

Znamo da SOR(ω) metoda konvergira za bilo koji početni vektor ako i samo ako je $|\lambda_j| < 1$ za sve j. Ako je pripadni $\mu_j = 0$, onda je $\lambda_j = 1 - \omega$, pa $|\lambda_j| < 1$ vrijedi ako i samo ako je $0 < \omega < 2$. Ako su to i jedine svojstvene vrijednosti od R_{Jac}, tj. $R_{Jac} = 0$, onda je prva tvrdnja dokazana.

U protivnom, postoje $\mu_j > 0$ i treba analizirati rješenja jednadžbe (1.37). Da bismo pojednostavnili zapis, promatramo kvadratnu jednadžbu

$$\lambda^2 + b\lambda + c = 0 \tag{1.38}$$

s realnim koeficijentima b i c. Tražimo kriterij (ako i samo ako uvjet) da korijeni ove jednadžbe leže unutar jediničnog kruga. Rješenja jednadžbe su

$$\lambda_{1,2} = \frac{1}{2}(-b \pm \sqrt{b^2 - 4c}).$$

Znamo da je $|c| = |\lambda_1|\,|\lambda_2|$, pa je $|c| < 1$ očiti nužni uvjet da bi oba korijena bila u jediničnom krugu.

Ako je diskriminanta negativna $b^2 - 4c < 0$, onda su rješenja konjugirano kompleksna

$$\lambda_{1,2} = \frac{1}{2}(-b \pm i\sqrt{4c - b^2}),$$

pa je

$$|\lambda_{1,2}|^2 = \frac{1}{4}(b^2 + 4c - b^2) = c. \tag{1.39}$$

Dakle, ako je $b^2 < 4c$, onda je $|\lambda_{1,2}| < 1$, ako i samo ako je $c < 1$. Dodatno, uočimo da je tada $c > 0$ i $|b| \leq 2\sqrt{c} < 1 + c$, zbog $(1 - \sqrt{c})^2 > 0$.

Ako je diskriminanta nenegativna $b^2 - 4c \geq 0$, onda imamo par realnih rješenja

$$\lambda_{1,2} = \frac{1}{2}(-b \pm \sqrt{b^2 - 4c}),$$

pa je

$$\max|\lambda_{1,2}| = \frac{1}{2}\max|-b \pm \sqrt{b^2 - 4c}| = \frac{1}{2}(|b| + \sqrt{b^2 - 4c}). \tag{1.40}$$

Iz $\max|\lambda_{1,2}| < 1$ dobivamo redom

$$\begin{aligned} |b| + \sqrt{b^2 - 4c} &< 2 \\ \sqrt{b^2 - 4c} &< 2 - |b| \\ b^2 - 4c &< (2 - |b|)^2 = 4 - 4|b| + b^2 \\ |b| &< 1 + c. \end{aligned}$$

Dakle, ako je $b^2 \geq 4c$, onda je $|\lambda_{1,2}| < 1$, ako i samo ako je $|b| < 1 + c$. Dodatno, iz $2 - |b| > 0$ slijedi $b^2 < 4$, pa mora biti i $c < 1$.

Zaključujemo da u oba slučaja iz $|\lambda_{1,2}| < 1$ slijedi $c < 1$ i $|b| < 1 + c$. Međutim, očito vrijedi i obrat, jer ovisno o odnosu b^2 i $4c$, iskoristimo pravu (jaču) od ove dvije pretpostavke. Na kraju, iz $|b| < 1 + c$ slijedi $1 + c > 0$ ili $c > -1$, što zajedno s $c < 1$, osigurava i raniji nužni uvjet $|c| < 1$.

Dokazali smo da rješenja jednadžbe (1.38) leže unutar jediničnog kruga, ako i samo ako vrijedi $c < 1$ i $|b| < 1 + c$.

Usporedbom (1.37) i (1.38) vidimo da je

$$b = -2\left(1 - \omega + \frac{1}{2}(\omega\mu_j)^2\right), \quad c = (\omega - 1)^2. \tag{1.41}$$

Uvjet $c < 1$ daje $(\omega - 1)^2 < 1$, ili $0 < \omega < 2$. Drugi uvjet $|b| < 1 + c$ daje

$$2\left|1 - \omega + \frac{1}{2}(\omega\mu_j)^2\right| < 1 + (\omega - 1)^2.$$

Ako lijevu stranu napišemo u obliku

$$|2 - 2\omega + \omega^2 + \omega^2(\mu_j^2 - 1)| = |1 + (\omega - 1)^2 + \omega^2(\mu_j^2 - 1)|,$$

uz oznaku $a := 1 + (\omega - 1)^2 > 0$ dobivamo uvjet

$$|a + \omega^2(\mu_j^2 - 1)| < a,$$

što je moguće ako i samo ako je drugi član negativan, tj. za $\mu_j^2 < 1$. Dakle, sve vrijednosti λ_j leže u jediničnom krugu, ako i samo ako je $0 < \omega < 2$ i vrijedi $\mu_j^2 < 1$ za sve j, što je ekvivalentno s $\mu := \rho(R_{Jac}) < 1$.

Time smo dokazali prvi dio tvrdnje. Drugi dio dobivamo tako da nađemo spektralni radijus $\rho(R_{SOR(\omega)})$ za svaki $\omega \in (0,2)$, a zatim ga minimiziramo po ω.

Neka je $\omega \in (0,2)$. Svojstvene vrijednosti λ_j možemo podijeliti u tri grupe (skupa). Ako je $\mu_j = 0$ za neki j, onda je pripadni $\lambda_j = 1 - \omega$ i

$$|\lambda_j| = |1 - \omega|. \tag{1.42}$$

Označimo skup svih takvih λ_j s S_0. Ako je $\mu = 0$, tj. $R_{Jac} = 0$, onda je očito $\rho(R_{SOR(\omega)}) = |1-\omega|$, pa je optimalna vrijednost parametra $\omega_{\text{opt}} = 1$. Dobivamo $\rho(R_{SOR(\omega_{\text{opt}})}) = 0$, što odgovara Gauss–Seidelovoj metodi s $R_{GS} = 0$. Dakle, i drugi dio tvrdnje vrijedi ako je $\mu = 0$.

Zbog toga možemo pretpostaviti da je $\mu > 0$, tj. da postoji barem jedna svojstvena vrijednost $\mu_j > 0$. Za $\mu_j > 0$, rješavanjem (1.37) dobivamo

$$\begin{aligned}(\lambda_j)_{1,2} &= 1 - \omega + \frac{1}{2}(\mu_j\omega)^2 \pm \mu_j\omega\sqrt{1 - \omega + \left(\frac{1}{2}\mu_j\omega\right)^2} \\ &= \left(\frac{1}{2}\mu_j\omega \pm \sqrt{1 - \omega + \left(\frac{1}{2}\mu_j\omega\right)^2}\right)^2. \end{aligned} \tag{1.43}$$

Ponašanje ovih rješenja ovisi o vrijednostima diskriminante

$$\Delta(\mu_j) = 1 - \omega + \left(\frac{1}{2}\mu_j\omega\right)^2.$$

Primijetimo da je $\Delta(\mu_j)$ rastuća funkcija od μ_j. Ovisno o predznaku $\Delta(\mu_j)$ dobivamo dvije grupe svojstvenih vrijednosti λ_j. Prva grupa S_- odgovara negativnim, a druga S_+ nenegativnim diskriminantama.

Za sve $\mu_j > 0$ za koje je $\Delta(\mu_j) < 0$, ako takvih ima, dobivamo kompleksno konjugirani par $(\lambda_j)_{1,2}$. Tada mora biti $1 - \omega < 0$, što pokazuje da je S_- neprazan samo ako je $\omega > 1$, tj. ova grupa je sigurno prazna za $\omega \leq 1$. Ako uvrstimo b i c iz (1.41) u (1.39), izlazi da je

$$|(\lambda_j)_{1,2}| = \sqrt{c} = \omega - 1, \tag{1.44}$$

pa vidimo da ove vrijednosti **ne ovise** o μ_j.

S druge strane, za one $\mu_j > 0$ za koje je $\Delta(\mu_j) \geq 0$, ako takvih ima, dobivamo par realnih svojstvenih vrijednosti $(\lambda_j)_{1,2}$. Analogno, iz (1.41) i (1.40), izlazi da je

$$\begin{aligned}\max|(\lambda_j)_{1,2}| &= \frac{1}{2}(|b| + \sqrt{b^2 - 4c}) \\ &= 1 - \omega + \frac{1}{2}(\mu_j\omega)^2 + \mu_j\omega\sqrt{1 - \omega + \left(\frac{1}{2}\mu_j\omega\right)^2} \\ &= \left(\frac{1}{2}\mu_j\omega + \sqrt{1 - \omega + \left(\frac{1}{2}\mu_j\omega\right)^2}\right)^2, \end{aligned}$$

jer iz $\Delta(\mu_j) \geq 0$ slijedi $1 - \omega + (\mu_j\omega/2)^2 > 0$. Očito je $\max|(\lambda_j)_{1,2}|$ rastuća funkcija po μ_j, pa najveću vrijednost dobivamo za najveći μ_j, tj. za $\mu_j = \mu = \rho(R_{Jac})$, naravno, pod uvjetom da je $\Delta(\mu) \geq 0$.

Obzirom na to da i $\Delta(\mu_j)$ raste s μ_j, druga grupa je neprazna ako i samo ako je $\Delta(\mu) \geq 0$ ili

$$\mu^2 > 4\,\frac{\omega - 1}{\omega^2},$$

pa je ova grupa sigurno neprazna za $\omega \leq 1$. Precizni kriterij nepraznosti S_+ dobivamo rješavanjem $\Delta(\mu) \geq 0$ po ω. Iz

$$0 = \Delta(\mu) = 1 - \omega + \left(\frac{1}{2}\mu\omega\right)^2$$

dobivamo granične vrijednosti za ω

$$\omega_{1,2} = 2\frac{1 \pm \sqrt{1-\mu^2}}{\mu^2} = \frac{2}{1 \mp \sqrt{1-\mu^2}}.$$

Intervalu $(0, 2)$ pripada samo manja od te dvije vrijednosti koju (zasad bez opravdanja) označavamo s ω_{opt}

$$\omega_{\text{opt}} = 2\frac{1 - \sqrt{1-\mu^2}}{\mu^2} = \frac{2}{1 + \sqrt{1-\mu^2}}.$$

Na kraju, zaključujemo da je S_+ neprazan ako i samo ako je $\omega \leq \omega_{\text{opt}}$. Uočimo da je $\omega_{\text{opt}} > 1$, zbog $\mu > 0$. Maksimum apsolutnih vrijednosti za $\lambda_j \in S_+ \neq \emptyset$ je

$$\lambda := 1 - \omega + \frac{1}{2}(\mu\omega)^2 + \mu\omega\sqrt{1 - \omega + \left(\frac{1}{2}\mu\omega\right)^2}. \tag{1.45}$$

Sad konačno možemo izračunati $\rho(R_{SOR(\omega)})$ uspoređujući maksimalne apsolutne vrijednosti iz (1.42), (1.44) i (1.45).

Za $\omega \leq 1$ je $S_- = \emptyset$. Za $\mu_j = 0$ imamo $|\lambda_j| = 1 - \omega$, pa je očito $|\lambda_j| < \lambda$, jer su u (1.45) svi članovi desne strane od trećeg nadalje pozitivni, a prva dva su upravo $|\lambda_j|$. Dakle, za $\omega \leq 1$ je $\rho(R_{SOR(\omega)}) = \lambda$.

Ako je $1 < \omega \leq \omega_{\text{opt}}$, onda je S_+ sigurno neprazan. Ako S_+ sadrži sve svojstvene vrijednosti, onda je očito $\rho(R_{SOR(\omega)}) = \lambda$. U protivnom, za $S_0 \cup S_- \neq \emptyset$, u (1.42) i (1.44) dobivamo isti maksimum $\omega - 1$. Usporedimo ga s λ iz (1.45). Zbog $\Delta(\mu) \geq 0$ dobivamo

$$\begin{aligned} \lambda - (\omega - 1) &= 2 - 2\omega + \frac{1}{2}(\mu\omega)^2 + \mu\omega\sqrt{1 - \omega + \left(\frac{1}{2}\mu\omega\right)^2} && (1.46)\\ &= 2\Delta(\mu) + \mu\omega\sqrt{\Delta(\mu)} \geq 0. && (1.47) \end{aligned}$$

Jednakost se dostiže samo za $\Delta(\mu) = 0$, tj. za $\omega = \omega_{\text{opt}}$, pa je opet $\rho(R_{SOR(\omega)}) = \lambda$.

Na kraju, za $\omega_{\text{opt}} < \omega < 2$, dobivamo da je $S_+ = \emptyset$. Barem jedan od skupova S_0, S_- nije prazan. U (1.42) i (1.44) dobivamo isti maksimum $\omega - 1$, pa je $\rho(R_{SOR(\omega)}) = \omega - 1$.

Dokazali smo da je

$$\rho(R_{SOR(\omega)}) = \{\lambda, \text{ za } 0 < \omega \leq \omega_{\text{opt}}, \omega - 1, \text{ za } \omega_{\text{opt}} \leq \omega \leq 2.$$

Time je zadnja relacija iz tvrdnje teorema u potpunosti dokazana.

Ostaje još naći najmanju moguću vrijednost od $\rho(R_{SOR(\omega)})$ za $\omega \in (0, 2)$. Kad λ iz (1.45) promatramo kao funkciju od ω

$$\begin{aligned} \lambda(\omega) &= 1 - \omega + \tfrac{1}{2}(\mu\omega)^2 + \mu\omega\sqrt{1 - \omega + \left(\tfrac{1}{2}\mu\omega\right)^2} \\ &= \left(\frac{1}{2}\mu\omega + \sqrt{1 - \omega + \left(\frac{1}{2}\mu\omega\right)^2}\right)^2, \end{aligned}$$

derviranjem drugog oblika po ω dobivamo

$$\begin{aligned} \frac{d\lambda}{d\omega} &= 2\sqrt{\lambda}\left(\frac{1}{2}\mu + \frac{\frac{1}{2}\mu^2\omega - 1}{2\sqrt{1 - \omega + (\frac{1}{2}\mu\omega)^2}}\right) \\ &= \sqrt{\lambda}\,\frac{\mu\sqrt{1 - \omega + (\frac{1}{2}\mu\omega)^2} + \frac{1}{2}\mu^2\omega - 1}{\sqrt{1 - \omega + (\frac{1}{2}\mu\omega)^2}} \end{aligned}$$

Brojnik možemo napisati u obliku $(\lambda - 1)/\omega$, pa je

$$\frac{d\lambda}{d\omega} = \frac{\sqrt{\lambda}\,(\lambda - 1)}{\omega\sqrt{\Delta(\mu)}} < 0, \quad \omega \in (0, \omega_{\text{opt}}),$$

jer je $0 < \lambda < 1$, $\omega > 0$ i $\Delta(\mu) > 0$. Dakle, $\rho(R_{SOR(\omega)}) = \lambda(\omega)$ monotono pada na $(0, \omega_{\text{opt}}]$, s tim da u ω_{opt} derivacija nije definirana (teži u $-\infty$), jer je $\Delta(\mu) = 0$.

Za $\omega \in (\omega_{\text{opt}}, 2)$ je $\rho(R_{SOR(\omega)}) = \omega - 1$, što je očito rastuća funkcija. U točki ω_{opt}, iz (1.47), jer je desna strana jednaka 0, slijedi

$$\rho(R_{SOR(\omega_{\text{opt}})}) = \lambda(\omega_{\text{opt}}) = \omega_{\text{opt}} - 1.$$

Dakle, $\rho(R_{SOR(\omega)})$ dostiže jedinstveni globalni minimum za $\omega = \omega_{\text{opt}}$ na $(0, 2)$, pa je i oznaka opravdana. To dokazuje i zadnji dio tvrdnje.

Vidimo da se minimum dostiže kad je $\Delta(\mu) = 0$, tj. kad je izraz pod korijenom u (1.43) jednak 0. Iz $\mu_j = \mu$ tada dobivamo dvostruki korijen $(\lambda_j)_{1,2} = \lambda(\omega_{\text{opt}})$. ■

Dobili smo da za optimalni parametar vrijedi $\omega_{\text{opt}} > 1$ pa je optimalni SOR zaista nadrelaksacija. Zgodno je primijetiti da što je μ bliže 1, tj. što sporije konvergira Jacobijeva metoda, to je ω_{opt} bliže 2, tj. SOR tada "produljuje" korak. I obratno, ako je μ blizu 0 (Jacobi brz), onda je ω_{opt} blizu 1, pa je optimalni SOR blizak Gauss–Seidelovoj metodi.

Ovo su slični uvjetni rezultati kao i za JOR metodu, u smislu da se pretpostavlja da Jacobijeva metoda konvergira. Srećom, za neke matrice nije teško pronaći spektralni radijus matrice u Jacobijevoj metodi i ustanoviti da Jacobijeva metoda konvergira. A tada imamo i optimalni izbor parametra za SOR, dok za JOR to nemamo. Ako usporedimo relacije koje vežu svojstvene vrijednosti matrica u JOR i SOR metodi s onima iz Jacobijeve metode

$$\begin{aligned} \lambda_j \in \sigma(R_{JOR(\omega)}): \quad & \lambda_j + \omega - 1 &= \omega\mu_j, \\ \lambda_j \in \sigma(R_{SOR(\omega)}): \quad & \lambda_j + \omega - 1 &= \omega\mu_j\sqrt{\lambda_j}, \end{aligned}$$

izlazi da ne bi bilo preteško dobiti neki rezultat o optimalnosti za JOR metodu. Probajte ga dobiti. Međutim, to se ne isplati. SOR metoda s optimalnim parametrom je bitno brža.

Teorem 1.9.2. o optimalnom izboru parametra u SOR metodi, osim konvergencije Jacobijeve metode, ima još i pretpostavku da su sve svojstvene vrijednosti matrice R_{Jac} realne. Kad bismo znali da je R_{Jac} simetrična (ili hermitska), onda je ta pretpostavka ispunjena. Međutim, to obično nije slučaj.

U principu znamo da je polazna matrica A simetrična ili hermitska. Tada je $R_{Jac} = D^{-1}(\widetilde{L} + \widetilde{L}^*)$, što ne mora biti simetrična matrica, osim ako D nije skalarna matrica, tj. ima konstantnu dijagonalu. Ako je A još i pozitivno definitna, onda je (vidjeti dokaz teorema 1.8.3.)

$$R_{Jac} = D^{-1}(\widetilde{L} + \widetilde{L}^*) = D^{-1/2}(D^{-1/2}(\widetilde{L} + \widetilde{L}^*)D^{-1/2})D^{1/2},$$

jer je i D pozitivno definitna (dijagonalna s pozitivnim elementima), pa je R_{Jac} slična simetričnoj (hermitskoj) matrici $D^{-1/2}(\widetilde{L} + \widetilde{L}^*)D^{-1/2}$, odakle slijedi da ima realne svojstvene vrijednosti μ_j.

To još uvijek ne garantira konvergenciju Jacobijeve metode, kao što ćemo pokazati na primjeru u sljedećem odjeljku. Međutim, ako je A još i konzistentno poredana, onda i Jacobijeva metoda mora konvergirati.

Teorem 1.9.3. *Neka je A simetrična (hermitska) i pozitivno definitna matrica i pretpostavimo da A ima konzistentan poredak. Onda matrica iteracije R_{Jac} u Jacobijevoj metodi ima samo realne svojstvene vrijednosti i vrijedi $\rho(R_{Jac}) < 1$, tj. Jacobijeva metoda konvergira.*

Dokaz:

Prvi dio tvrdnje da je $\mu_j \in \mathbf{R}$ smo već dokazali. Iz pozitvne definitnosti od $A = D - \widetilde{L} - \widetilde{L}^*$ slijedi da za bilo koji vektor $x \neq 0$ vrijedi

$$0 < x^*Ax = x^*(D - \widetilde{L} - \widetilde{L}^*)x$$

pa je

$$x^*(\widetilde{L} + \widetilde{L}^*)x < x^* D x.$$

Ako definiramo $y = D^{1/2}x$, onda je $y \neq 0$ ako i samo ako je $x \neq 0$, i za sve $y \neq 0$ vrijedi

$$y^* D^{-1/2}(\widetilde{L} + \widetilde{L}^*)D^{-1/2} y < y^* y.$$

što znači da je matrica $I - D^{-1/2}(\widetilde{L} + \widetilde{L}^*)D^{-1/2}$ pozitivno definitna. Zbog toga su sve svojstvene vrijednosti matrice $D^{-1/2}(\widetilde{L}+\widetilde{L}^*)D^{-1/2}$ strogo manje od 1, pa to zbog sličnosti vrijedi i za R_{Jac}.

Sad iskoristimo da A ima konzistentan poredak, pa iz teorema 1.9.1.(*a*) slijedi da svojstvene vrijednosti od R_{Jac} dolaze u $\pm$ parovima, tj. $\mu_j < 1$ povlači i $\mu_j > -1$. Dobivamo da je $\rho(R_{Jac}) < 1$, pa Jacobijeva metoda konvergira. ■

Naravno da tada konvergiraju i Gauss–Seidelova metoda i SOR(ω) metode za $\omega \in (0, 2)$. To vrijedi i bez pretpostavke o konzistentnom poretku, ali tada nemamo optimalni izbor parametra za SOR metodu.

Ako imamo zadanu matricu A, onda nije jednostavno provjeriti da li ona ima konzistentan poredak koristeći definiciju 1.9.3., odnosno algebarsko svojstvo (1.32). Mnogo je lakše provjeriti "grafovska" svojstva, poput svojstva (A), koja garantiraju konzistentan poredak za permutiranu matricu PAP^T. Zbog toga je vrlo korisno naći generalizacije svojstva (A).

Na tu temu postoje mnogi rezultati. U primjeru 1.9.1. imali smo blok trodijagonalu matricu s regularnim dijagonalnim matricama.

Definicija 1.9.4. *Matrica A ima svojstvo (A^π) ako postoji matrica permutacije P takva da se PAP^T može particionirati u blok trodijagonalnu matricu oblika*

$$PAP^T = \begin{bmatrix} D_1 & U_1 & & \\ L_1 & \ddots & \ddots & \\ & \ddots & \ddots & U_{n-1} \\ & & L_{n-1} & D_n \end{bmatrix},$$

gdje su D_i, $i = 1, \ldots, m$, regularne matrice.

Dokažite da matrice sa svojstvom (A^π) imaju konzistentan poredak. Dokaz ide slično kao u propoziciji 1.9.1., tako da se pogodnom dijagonalnom matricom sličnosti transformira $R_{Jac}(\alpha)$ i pokaže da njene svojstvene vrijednosti ne ovise o α. Ovaj rezultat je također dokazao Young, 1950. godine.

Na kraju, spomenimo da je R. S. Varga generalizirao i ovaj rezultat na tzv. p-cikličke matrice, uz $p \geq 2$, pri čemu su 2-cikličke matrice upravo one sa svojstvom (A^π).

1.10 Primjeri — akademski i praktični

Za početak, ilustrirajmo odnos između Jacobijeve i Gauss–Seidelove metode, u smislu da postoje primjeri kad jedna od njih konvergira, a druga ne.

Znamo da Gauss–Seidelova metoda konvergira za simetrične (hermitske) pozitivno definitne matrice. Jacobijeva metoda tad ne mora konvergirati. Naravno, treba uzeti matricu koja nije dijagonalno dominantna po recima ili stupcima, tj. vandijagonalni elementi trebaju biti relativno veliki.

Primjer 1.10.1. *Lako se provjerava da je matrica*

$$A = \begin{bmatrix} 1 & 0.9 & 0.9 \\ 0.9 & 1 & 0.9 \\ 0.9 & 0.9 & 1 \end{bmatrix}$$

simetrična i pozitivno definitna, jer su sve vodeće glavne minore pozitivne: $D_1 = 1$, $D_2 = 1 - (0.9)^2 = 0.19$ *i*

$$\begin{aligned} D_3 &= 1 + 2(0.9)^3 - 3(0.9)^2 = 1 + 2 \cdot 0.729 - 3 \cdot 0.81 = 1 + 1.458 - 2.43 \\ &= 2.458 - 2.43 = 0.028. \end{aligned}$$

Matrica iteracije u Jacobijevoj metodi je ($D = I$),

$$R_{Jac} = - \begin{bmatrix} 0 & 0.9 & 0.9 \\ 0.9 & 0 & 0.9 \\ 0.9 & 0.9 & 0 \end{bmatrix}$$

i ima svojstvene vrijednosti $\mu_1 = \mu_2 = 0.9$ *i* $\mu_3 = -1.8$. *Dakle,* $\rho(R_{Jac}) = 1.8$, *pa Jacobijeva metoda ne konvergira za sve početne iteracije.*

Pogledajmo usput što daje teorem 1.8.3. *o konvergenciji JOR metode. Najmanja svojstvena vrijednost od* R_{Jac} *je* $\mu := \min_j \mu_j = -1.8$. *Iz* (1.28) *dobivamo da JOR*(ω) *metoda konvergira za sve parametre* ω *za koje vrijedi*

$$0 < \omega < \frac{2}{1-\mu} = \frac{5}{7} < 1,$$

tj. imamo pravu podrelaksaciju.

Primjer 1.10.2. *Za matricu*

$$A = \begin{bmatrix} 1 & 1 & -1 \\ \frac{1}{2} & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

odmah vidimo da nije dijagonalno dominantna. Ipak, Jacobijeva metoda konvergira s matricom iteracije

$$R_{Jac} = \begin{bmatrix} 0 & -1 & 1 \\ -\frac{1}{2} & 0 & -1 \\ -1 & -1 & 0 \end{bmatrix}.$$

Pokažite da je $\rho(R_{Jac}) < 1$. *Izračunajte matricu* R_{GS} *i pokažite da je* $\rho(R_{GS}) > 1$, *tj. da Gauss–Seidelova metoda ne konvergira za sve početne iteracije.*

U nastavku ovog odjeljka ilustrirat ćemo iterativne metode na jednoj klasi matrica koja se javlja u praksi i ima tipična svojstva.

Iterativne metode za rješavanje linearnih sustava koriste se kod rješavanja rubnog problema za obične diferencijalne jednadžbe i kod rješavanja parcijalnih diferencijalnih jednadžbi.

Ideja metoda koje vode na linearne sustave je diskretizacija, tj. umjesto da tražimo funkciju koja bi bila rješenje problema, aproksimiramo rješenje u odabranim čvorovima, tako da aproksimiramo derivacije koje se javljaju u problemu.

Ako je problem jednodimenzionalan, obično se interval na kojem tražimo rješenje ekvidistantno podijeli. Kod dvodimenzionalnog problema, područje se obično podijeli u pravokutnu mrežu.

Na primjer, želimo riješiti tzv. Poissonovu (eliptičku parcijalnu diferencijalnu) jednadžbu u dvije dimenzije

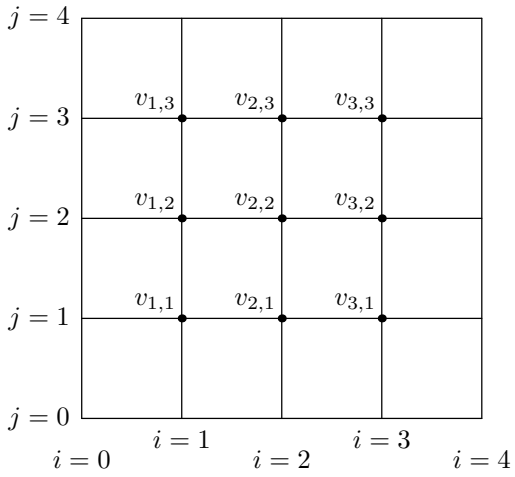
$$-\frac{\partial^2 v(x,y)}{\partial x^2} - \frac{\partial^2 v(x,y)}{\partial y^2} = f(x,y)$$

na kvadratu $\{(x,y) \mid 0 < x, y < 1\}$ uz rubni uvjet $v = 0$, tj. funkcija v je jednaka 0 na rubu kvadrata. Kvadrat podijelimo u mrežu čvorova, a da nam bude jednostavnije, pretpostavimo da je i ta mreža kvadratna, tj. korak u x i y smjeru je jednak

$$h = \frac{1}{N+1}.$$

Uz tako definirane korake, unutarnji čvorovi mreže su točke (x_i, y_j), gdje je $x_i = ih$, $y_j = jh$, za $i, j = 1, \ldots, N$. Dakle, imamo $n := N^2$ unutarnjih čvorova mreže.

Takva mreža za $N = 3$ izgleda ovako:



Vrijednost rješenja u čvoru (x_i, y_j) označavamo s $v_{i,j} := v(ih, jh)$, a funkcijsku vrijednost s $f_{i,j} := f(ih, jh)$.

Kako se aproksimiraju derivacije? Pretpostavimo da su točke x_{i-1}, x_i i x_{i+1} ekvidistantne i da je $x_{i+1} - x_i = x_i - x_{i-1} = h$. Ako za funkciju f postoji Taylorov red oko x_i, onda uvrštavanjem točaka x_{i-1} i x_{i+1} u taj red dobivamo

$$\begin{aligned} f(x_{i-1}) &= f(x_i) - \frac{f'(x_i)}{1!}h + \frac{f''(x_i)}{2!}h^2 - \frac{f'''(\xi_{i,i-1})}{3!}h^3 \\ f(x_{i+1}) &= f(x_i) + \frac{f'(x_i)}{1!}h + \frac{f''(x_i)}{2!}h^2 + \frac{f'''(\xi_{i,i+1})}{3!}h^3. \end{aligned}$$

Oduzimanjem prve jednakosti od druge, izlazi

$$f(x_{i+1}) - f(x_{i-1}) = 2\frac{f'(x_i)}{1!}h + O(h^3),$$

pa je dobra aproksimacija derivacije funkcije f u točki x_i

$$f'(x_i) \approx \frac{f(x_{i+1}) - f(x_{i-1})}{2h}.$$

Ta aproksimacija derivacije obično se naziva simetrična (centralna) razlika.

Prvo, izaberimo dva čvora (x_i^-, y_j) i (x_i^+, y_j) takva da je

$$x_i^- = \frac{x_i + x_{i-1}}{2}, \qquad x_i^+ = \frac{x_i + x_{i+1}}{2}.$$

Korištenjem simetrične razlike, aproksimirajmo prve parcijalne derivacije u ta dva čvora. Dobivamo

$$\begin{aligned} \left.\frac{\partial v}{\partial x}\right|_{x=x_i^-,\ y=y_j} &\approx \frac{v_{i,j} - v_{i-1,j}}{h}, \\ \left.\frac{\partial v}{\partial x}\right|_{x=x_i^+,\ y=y_j} &\approx \frac{v_{i+1,j} - v_{i,j}}{h}. \end{aligned}$$

Ponovno, primijenimo simetričnu razliku, ali ovaj puta za drugu parcijalnu derivaciju po x u (x_i, y_j), korištenjem derivacije u točkama (x_i^-, y_j) i (x_i^+, y_j). Odmah imamo

$$\left.\frac{\partial^2 v}{\partial x^2}\right|_{x=x_i,\ y=y_j} \approx \frac{1}{h}\left(\left.\frac{\partial v}{\partial x}\right|_{x=x_i^+,\ y=y_j} - \left.\frac{\partial v}{\partial x}\right|_{x=x_i^-,\ y=y_j}\right) = \frac{v_{i-1,j} - 2v_{i,j} + v_{i+1,j}}{h^2}.$$

Na isti način dobivamo i formulu za drugu parcijalnu derivaciju po y

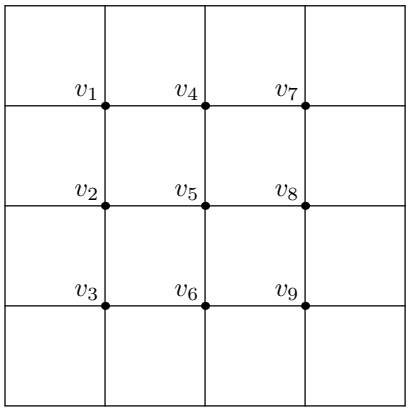
$$\left.\frac{\partial^2 v}{\partial y^2}\right|_{x=x_i,\ y=y_j} \approx \frac{1}{h}\left(\left.\frac{\partial v}{\partial y}\right|_{x=x_i,\ y=y_j^+} - \left.\frac{\partial v}{\partial y}\right|_{x=x_i,\ y=y_j^-}\right) = \frac{v_{i,j-1} - 2v_{i,j} + v_{i,j+1}}{h^2}.$$

Uvrstimo li te aproksimacije derivacija u diferencijalnu jednadžbu, dobivamo

$$4v_{i,j} - v_{i-1,j} - v_{i+1,j} - v_{i,j-1} - v_{i,j+1} = h^2 f_{i,j}, \quad 1 \le i, j \le N. \tag{1.48}$$

Pitanje je kako treba napisati ove jednadžbe, tako da se dobije linearni sustav s nekom strukturom. Postoje dva načina da bi se to napravilo. Jedan je sekvencijalno numeriranje $v_{i,j}$ po recima ili stupcima (slijeva nadesno, ili zdesna nalijevo, odozgo nadolje ili odozdo nagore), a drugi tzv. crveno–crni poredak čvorova.

Ako v_{ij} sekvencijalno numeriramo po stupcima odozgo nadolje, na primjer za $N = 3$, dobivamo ovakav poredak čvorova



Dakle, u (1.48) lako zamjenjujemo $v_{i,j}$ s v_k. Ako se na isti način transformiraju i $f_{i,j}$ u f_k, onda dobivamo linearni sustav

$$T_{N\times N} v = h^2 f,$$

gdje je $v = [v_1, v_2, \ldots, v_{N\times N}]^T$, $f = [f_1, f_2, \ldots, f_{N\times N}]^T$, a matrica $T_{N\times N}$ ima N blok-redaka i stupaca, svaki dimenzije N. Matrica

$$T_{N\times N} = \begin{bmatrix} T_N + 2I_N & -I_N & & \\ -I_N & \ddots & \ddots & \\ & \ddots & \ddots & -I_N \\ & & -I_N & T_N + 2I_N \end{bmatrix}, \tag{1.49}$$

pri čemu je I_N jedinična matrica reda N, a

$$T_N = \begin{bmatrix} 2 & -1 & & \\ -1 & \ddots & \ddots & \\ & \ddots & \ddots & -1 \\ & & -1 & 2 \end{bmatrix}.$$

Može se pokazati da je T_N matrica koja nastaje diskretizacijom odgovarajuće jednodimenzionalne Poissonove jednadžbe.

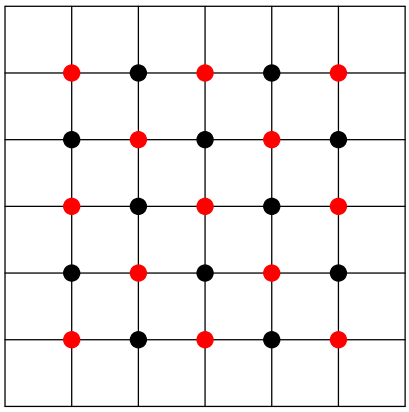
Na primjer, za $N = 3$, matrica linearnog sustava je

$$T_{3\times 3} = \left[\begin{array}{ccc|ccc|ccc} 4 & -1 & & -1 & & & & & \\ -1 & 4 & -1 & & -1 & & & & \\ & -1 & 4 & & & -1 & & & \\ \hline -1 & & & 4 & -1 & & -1 & & \\ & -1 & & -1 & 4 & -1 & & -1 & \\ & & -1 & & -1 & 4 & & & -1 \\ \hline & & & -1 & & & 4 & -1 & \\ & & & & -1 & & -1 & 4 & -1 \\ & & & & & -1 & & -1 & 4 \end{array}\right]$$

Uočite da je matrica $T_{N\times N}$ slabo dijagonalno dominantna i ireducibilna, pa će i Jacobijeva i Gauss–Seidelova metoda konvergirati. Dapače, pokazat ćemo da $T_{N\times N}$ ima svojstvo (A) i da je konzistentno poredana.

Ako čvorove $v_{i,j}$ poredamo u tzv. crveno–crni poredak, dobit ćemo konzistentno poredanu matricu. Crveno–crni poredak dobivamo tako da ih obojamo poput šahovske ploče: svaki crveni čvor (osim rubnog) je okružen s četiri crna susjeda i obratno.

Na primjer, za $N = 5$ takvo crveno–crno bojanje čvorova izgleda ovako:



Ako zatim sve čvorove koji su crveno obojani popišemo prije crnih (dodijelimo im indekse prije crnih), ili obratno, dobit ćemo blok matricu oblika

$$PT_{N\times N}P^T = \begin{bmatrix} D_1 & T_{12} \\ T_{21} & D_2 \end{bmatrix}.$$

Lako je vidjeti da su dijagonalni blokovi baš dijagonalne matrice, jer ne postoji veza između dva crvena ili dva crna čvora (osim čvora sa samim sobom).

Konkretno, crveno–crni poredak za matricu $T_{3\times 3}$ daje

$$PT_{3\times 3}P^T = \left[\begin{array}{ccccc|cccc} 4 & & & & & -1 & -1 & & \\ & 4 & & & & -1 & & -1 & \\ & & 4 & & & -1 & -1 & -1 & -1 \\ & & & 4 & & & -1 & & -1 \\ & & & & 4 & & & -1 & -1 \\ \hline -1 & -1 & -1 & & & 4 & & & \\ -1 & & -1 & -1 & & & 4 & & \\ & -1 & -1 & & -1 & & & 4 & \\ & & -1 & -1 & -1 & & & & 4 \end{array}\right].$$

Dakle, da bismo ispitali konvergenciju iterativnih metoda, dovoljno je naći spektralni radijus matrice R_{Jac}. Prvo, nađimo rastav (cijepanje) matrice $T_{N\times N}$

$$T_{N\times N} = 4I_{N\times N} - (4I_{N\times N} - T_{N\times N}),$$

pa je $M = 4I_{N\times N}$, $K = (4I_{N\times N} - T_{N\times N})$,

$$R_{Jac} = M^{-1}K = (4I_{N\times N})^{-1}\,(4I_{N\times N} - T_{N\times N}) = I_{N\times N} - \frac{1}{4}\,T_{N\times N}.$$

Drugim riječima, R_{Jac} je polinom od $T_{N\times N}$, pa ako je $\lambda_{i,j}$ svojstvena vrijednost od $T_{N\times N}$, onda je $1 - \lambda_{i,j}/4$ svojstvena vrijednost od R_{Jac}.

Pametnim raspisivanjem matrice $T_{N\times N}$, može se pokazati da su svojstvene vrijednosti matrice $T_{N\times N}$ jednake

$$\lambda_{i,j} = 4 - 2\left(\cos\frac{\pi i}{N+1} + \cos\frac{\pi j}{N+1}\right),$$

odakle slijedi da je

$$\rho(R_{Jac}) = \max_{i,j}\left|1 - \frac{\lambda_{i,j}}{4}\right| = \left|1 - \frac{\lambda_{1,1}}{4}\right| = \left|1 - \frac{\lambda_{N,N}}{4}\right| = \cos\frac{\pi}{N+1}.$$

Odmah je vidljivo da porastom N argument kosinusa ide prema nuli, pa će $\rho(R_{Jac})$ biti sve bliže 1, a iterativne će metode sve sporije konvergirati.

Čak štoviše, možemo procijeniti $\rho(R_{Jac})$ za velike N. Dovoljno dobra aproksimacija bit će prva dva člana u Taylorovom redu za funkciju kosinus

$$\rho(R_{Jac}) = \cos\frac{\pi}{N+1} \approx 1 - \frac{\pi^2}{2(N+1)^2}.$$

Spektralni radijus za Gauss–Seidelovu metodu lako je dobiti koristeći korolar 1.9.1.

$$\rho(R_{GS}) = \big(\rho(R_{Jac})\big)^2 = \cos^2\frac{\pi}{N+1}.$$

Približno, kvadriranjem prva dva člana u Taylorovom redu, vrijedi

$$\rho(R_{GS}) \approx 1 - \frac{\pi^2}{(N+1)^2}.$$

Konačno, koristeći teorem 1.9.2., za SOR(ω) dobivamo

$$\omega_{opt} = \frac{2}{1 + \sin\frac{\pi}{N+1}}, \qquad \rho(R_{SOR(\omega_{opt})}) = \frac{\cos^2\frac{\pi}{N+1}}{\left(1 + \sin\frac{\pi}{N+1}\right)^2}.$$

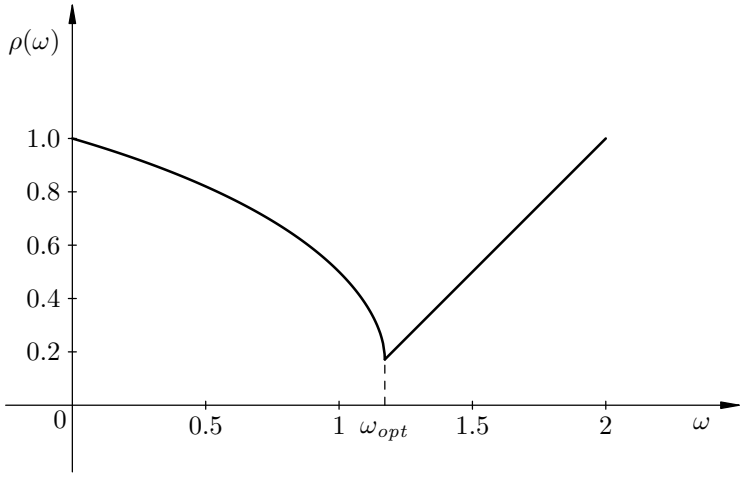
Veličinu $\rho(R_{SOR(\omega_{opt})})$ možemo približno ocijeniti za velike N. Vrijedi

$$\frac{\cos^2\frac{\pi}{N+1}}{\left(1 + \sin\frac{\pi}{N+1}\right)^2} = \frac{1 - \sin\frac{\pi}{N+1}}{1 + \sin\frac{\pi}{N+1}} = 1 - \frac{2\sin\frac{\pi}{N+1}}{1 + \sin\frac{\pi}{N+1}} \approx 1 - 2\sin\frac{\pi}{N+1} \approx 1 - \frac{2\pi}{N+1}.$$

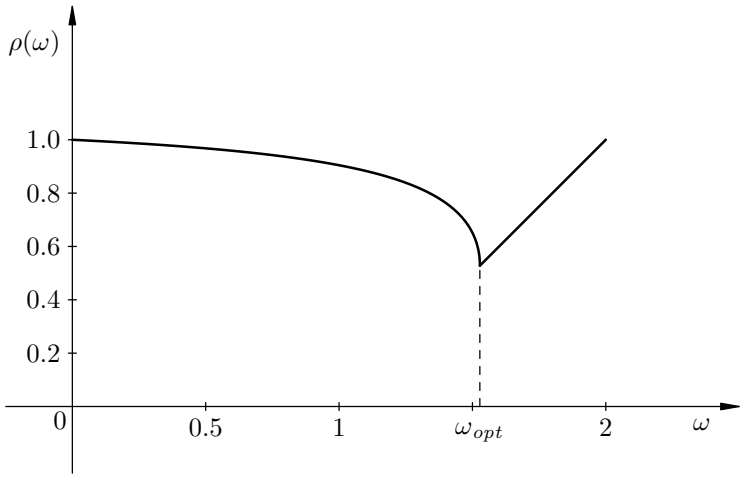
Primijetite da spektralni radijus i kod Jacobijeve metode i kod Gauss–Seidelove metode ima oblik $1 - O(1/N^2)$, dok kod SOR metode s optimalnim parametrom spektralni radijus ima oblik $1 - O(1/N)$, što pokazuje da bi SOR za optimalni izbor parametra morao biti reda veličine N puta brži i od Jacobijeve i od Gauss–Seidelove metode.

Pogledajmo kako se ponaša $\rho(R_{SOR(\omega)})$ kao funkcija od ω, te ovisno o N, kako se ponaša ω_{opt}. Redom, za $N = 3, 9, 16, 25$, dobivamo sljedeće grafove

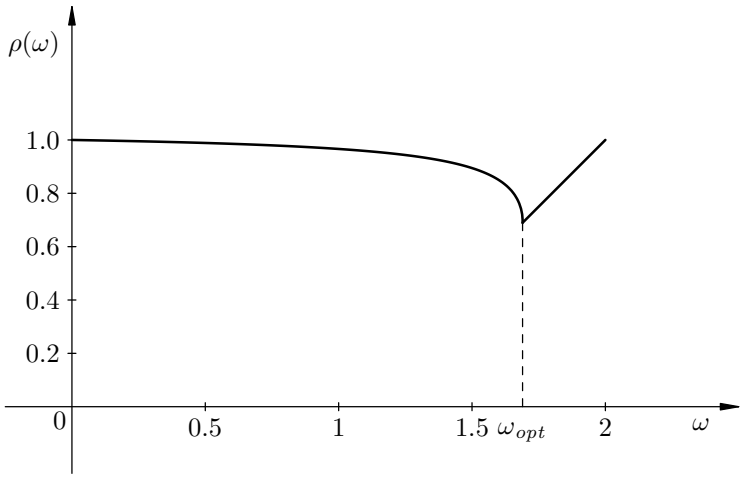
$N = 3$



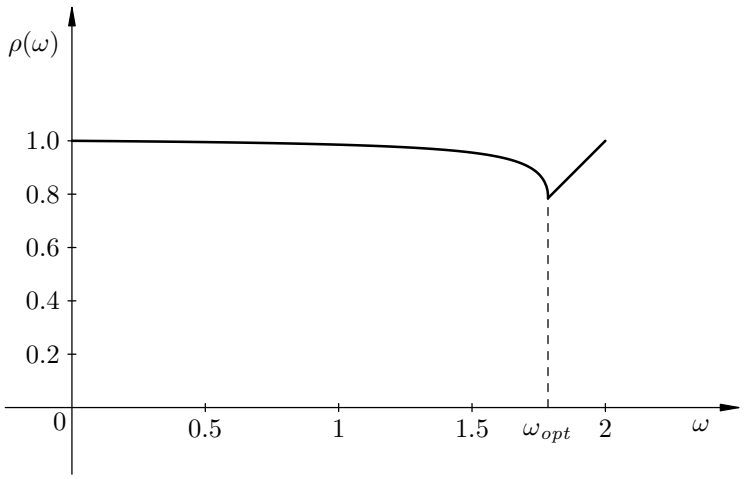
$N = 9$



$N = 16$



$N = 25$



Kao što smo očekivali, optimalni se parametar pomiče prema 2, a $\rho(R_{SOR(\omega_{opt})})$ postaje sve bliže 1.

U sljedećoj tablici dan je pregled spektralnih radijusa za različite iterativne metode za razne N.

N	$\rho(R_{Jac})$	$\rho(R_{GS})$	ω_{opt}	$\rho(R_{SOR(\omega_{opt})})$
4	0.8090169944	0.6545084972	1.2596161837	0.2596161837
9	0.9510565163	0.9045084972	1.5278640450	0.5278640450
16	0.9829730997	0.9662361147	1.6895466227	0.6895466227
25	0.9927088741	0.9854709087	1.7848590191	0.7848590191
36	0.9963974885	0.9928079552	1.8436477483	0.8436477483
49	0.9980267284	0.9960573507	1.8818383898	0.8818383898
64	0.9988322268	0.9976658174	1.9078264563	0.9078264563
81	0.9992661811	0.9985329006	1.9262204896	0.9262204896
100	0.9995162823	0.9990327986	1.9396763332	0.9396763332
121	0.9996684675	0.9993370449	1.9497968003	0.9497968003
144	0.9997652980	0.9995306511	1.9575898703	0.9575898703
169	0.9998292505	0.9996585301	1.9637127389	0.9637127389
196	0.9998728466	0.9997457093	1.9686076088	0.9686076088
225	0.9999033847	0.9998067787	1.9725803349	0.9725803349
256	0.9999252867	0.9998505789	1.9758476503	0.9758476503

Sad kad imamo sve informacije o ovim iterativnim metodama, trebalo bi još samo izabrati neki početni vektor i računati rješenje za zadani f i podjelu N.

Kako se bira početni vektor $x^{(0)}$? Ako znamo matricu iteracija R i iterativnu metodu realiziramo u obliku (1.13)

$$x^{(m+1)} = Rx^{(m)} + c, \quad m \in \mathbf{N}_0,$$

onda se obično bira $x^{(0)} = c$, što bi odgovaralo tome da je prethodna iteracija bio nul-vektor. Razlog za to je sasvim jednostavan, posebno u aritmetici računala. Ako je $c = 0$, onda stajemo u jednom koraku i "ne kvarimo" nule. Inače bismo nul-vektor morali dobiti kao limes vektora koji nemaju (sve) nul-komponente, tj. sigurno dolazi do kraćenja.

Međutim, niti jedan od naša 4 algoritma koje smo napisali ne provodi iteracije u tom obliku, jer se R katkad komplicirano računa, već radimo direktno s originalnim podacima A i b. Tada je zgodno zaista uzeti $x^{(0)} = 0$, za slučaj da je $b = 0$, iz istih razloga. Prethodni primjer pokazuje da sve potrebne informacije o matrici iteracija R možemo dobiti i bez da ju eksplicitno izračunamo.

Kako zaustavljamo iteracije? Najlakši način je tzv. heuristička konvergencija. Unaprijed zadamo traženu točnost ε i prekidamo iteracije čim vrijedi

$$\|x^{(m+1)} - x^{(m)}\| \leq \varepsilon,$$

u nekoj pogodno odabranoj vektorskoj normi, na primjer, ∞-normi. To znači da u svakom koraku moramo računati i ovu normu, ali to obično nije pretjerano skupo, a može se isplatiti, ako "slučajno" greška naglo padne u nekoj iteraciji.

S druge strane, ako znamo vrijednost neke operatorske norme $\|R\|$ matrice iteracija (bez pretjerano računanja), s tim da je $\|R\| < 1$, onda unaprijed možemo izračunati potreban broj iteracija. Naime, u pripadnoj vektorskoj normi vrijedi ocjena

$$\|x^{(m+1)} - x\| \leq \frac{\|R\|}{1 - \|R\|} \|x^{(m+1)} - x^{(m)}\|, \tag{1.50}$$

kao u Banachovom teoremu o fiksnoj točki. Dokaz ove relacije je jednostavan:

$$\begin{aligned} x^{(m+1)} - x &= R(x^{(m)} - x) = R(x^{(m)} - x^{(m+1)}) + R(x^{(m+1)} - x) \\ (I - R)(x^{(m+1)} - x) &= -R(x^{(m+1)} - x^{(m)}) \\ x^{(m+1)} - x &= -(I - R)^{-1} R(x^{(m+1)} - x^{(m)}), \end{aligned}$$

jer znamo da je $I - R$ regularna matrica. Primjenom norme i ocjenama izlazi (1.50). Iz te ocjene dobivamo i

$$\|x^{(m+1)} - x\| \leq \frac{\|R\|^{m+1}}{1 - \|R\|} \|x^{(1)} - x^{(0)}\|,$$

a iz ove relacije možemo, nakon prve iteracije, izračunati potreban broj iteracija $m + 1$, tako da, do na greške zaokruživanja, vrijedi

$$\|x^{(m+1)} - x\| \leq \varepsilon.$$