

Izrek: Naj bo točna rešitev Poissonar (37)
 enačbe naaf stikihnat zvezno odvedljiva
 funkcija. Za globalno napako $e = u_\delta - u$
 velja ocena

$$\max_{(x,y) \in D_\delta} |e(x,y)| \leq \underbrace{\max_{(x,y) \in \partial D_\delta} |e(x,y)|}_{\text{napaka na robu}} + \underbrace{\sqrt{\delta x^2 + \delta y^2}}_{\text{lokalna napaka}.$$

Na robu je napaka omejena z osnovo
 zaohranitveno napako enaka 0, zato
 je globalna napaka enaka kot lokalna
napaka.

Če točna rešitev obstaja, potem po pripravi
 izreku obstaja tudi rešitev diskretnih naloge
 za dovolj majhna δx in δy .

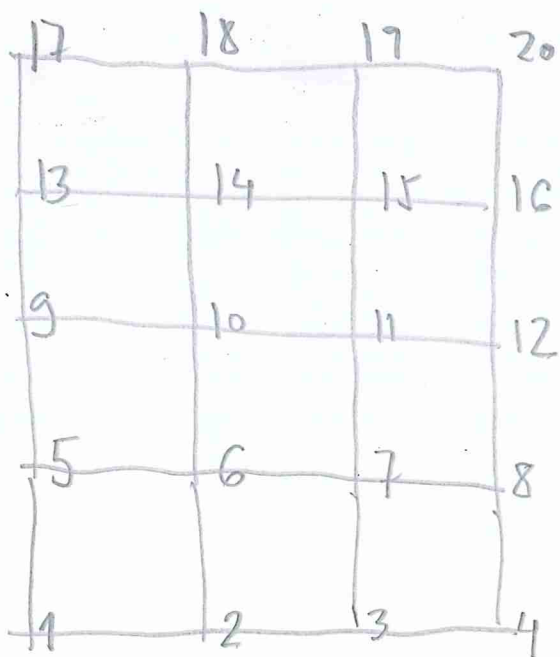
Numerično rešitev diskretnega problema (38) določimo tako, da poiščemo numerične približke u_{jk} analitičnih vrednosti $u(x_j, y_k)$.

Rešiti moramo sistem linearnih enačb za u_{jk} . Zapišemo sistem:

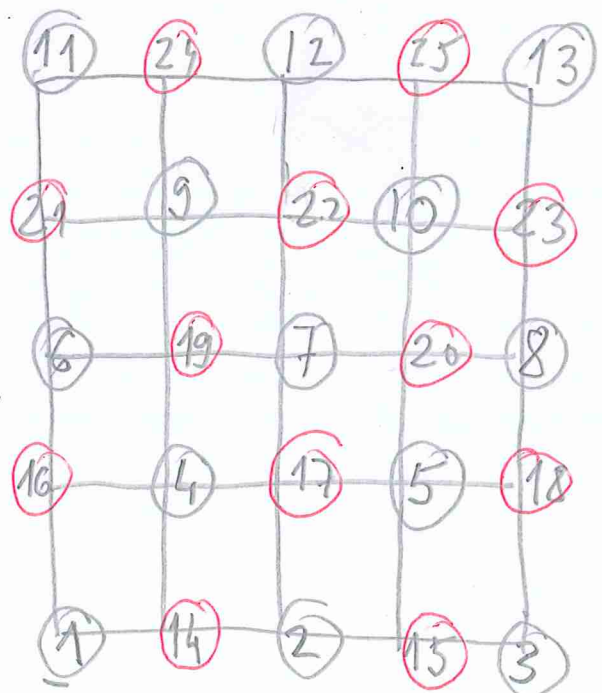
Diferenčne enačbe približujemo s

$$-\theta_x (u_{j-1,k} + u_{j+1,k}) + u_{jk} - \theta_y (u_{j,k-1} + u_{j,k+1}) = \delta^2 f_{jk}, \quad (x_j, y_k) \in D_\delta$$

Potrebno je določiti, kako enačbe po vrsti pišemo v matriko (ostanili moramo neznanke)



Številčenje po vrsticah.



Črna - rdeča številčenje

Izberimo številčenje po noticah.

(39)

Nerazneke zapišemo v vektor

$$u = \left((u_{jik})_{j=1}^J \right)_{k=1}^K \in \mathbb{R}^{J \cdot K},$$

desno stran v

$$f = \left((f_{jik})_{j=1}^J \right)_{k=1}^K \in \mathbb{R}^{J \cdot K}.$$

Matrica pa postane simetrična bločno tridiagonalna

$$\begin{matrix} JK \times JK \\ \mathbb{R} \ni A = \end{matrix} \begin{bmatrix} A_{11} & A_{12} & & & \\ A_{21} & A_{22} & A_{23} & & \\ & \ddots & \ddots & \ddots & \\ & & A_{K,K-1} & A_{K,K} & A_{K,K+1} \\ & & & \ddots & \ddots \\ & & & & A_{K-1,K-2} & A_{K-1,K-1} & A_{K-1,K} \\ & & & & & A_{K,K-1} & A_{K,K} \end{bmatrix}$$

Prvi tem so diagonalni bloki enaki

$$A_{kk} = \begin{bmatrix} 1 & -\theta_x & & & \\ -\theta_x & 1 & -\theta_x & & \\ & & \ddots & \ddots & \\ & & & -\theta_x & 1 & -\theta_x \\ & & & & -\theta_x & 1 \end{bmatrix} \in \mathbb{R}^{J \times J},$$

oddiagonalni bloki pa so diagonalne matrice

$$A_{k,k-1} = A_{k-1,k} = \begin{bmatrix} -\theta_y & & & \\ & -\theta_y & & \\ & & \ddots & \\ & & & -\theta_y & -\theta_y \end{bmatrix} \in \mathbb{R}^{J \times J}.$$

Robni pogoji so v vektorju $g \in \mathbb{R}^{JK}$:

- prvi bločni stolpec $\theta_y g_{j,0}$, $j=1,2,\dots,J$,
- zadnji bločni stolpec $\theta_y g_{j,k+1}$, $j=1,2,\dots,J$,
- ostali bločni stolpci:

$$[\theta_x g_{0,k}, 0, 0, \dots, 0, \theta_x g_{J+1,k}], k=1,2,\dots,k.$$

Sistem enačb se na kratko glasi

$$Au = \delta^2 f + g =: b$$

Enact, kjer je srednji indeks sheme (41)
(y, k) opisuje vrstica $i = (k-1)J + j$.

Matrica A lahko zapisemo kot

$$A = I - H - V, \text{ kjer } v$$

H prideta koeficienta pri $u_{j-1,k}, u_{j+1,k},$

$$H = F + F^T.$$

Pri tem smo $F \in \mathbb{R}^{JK \times JK}$ pisali matrico,
ki ima nenulne elemente (θ_x) tam,
kjer je v vrstica $i = (k-1)J + j$ matrice
 A koeficient pri memanku $u_{j-1,k}$.

Podobno V preizame koeficienta pri

$u_{j,k-1}, u_{j,k+1}, V = B + B^T$ in
 $B \in \mathbb{R}^{JK \times JK}$ določa mesto koeficientov (θ_y)
v

A za memanke $u_{j,k-1}$

Strogo spodnji del dobimo kot $L = F + B,$
strogo zgornji del pa $U = F^T + B^T.$

Neformalno vemo, da numerična rešitev obstaja (za dovolj majhne δx in δy).

(42)

Sedaj lahko dokažemo, da obstaja rešitev za pozitivna $\delta x > 0$, $\delta y > 0$.

Izrek: Sistem enačb $Au = \delta^2 f + g$ je enolično rešljiv za pozitivna pozitivna konstanti δx in δy .

Dokaz: Najprej opazimo, da je matrika A ireducibilna. To pomeni, da se ji ne da s podobno permutacijsko transformacijo pretvoriti na bločno zgornjo trikotno obliko

$$P^T A P = \begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ 0 & \tilde{A}_{22} \end{bmatrix}.$$

Res. Če bi to bilo mogoče, mora zaradi simetričnosti A in posledično $P^T A P$ veljati $\tilde{A}_{12} = 0$. To pomeni, da bi lahko enačbe in neenake predstavili tako, da bi dobili dve skupini med seboj nevezanih enačb in neenak, kar očitno ne gre.

Za dohat mitrofenosti uporabimo
Gerschgorinov izrek: spekter matrike A ,
 $\sigma(A)$, mora ležati v uniji Gerschgorinovih
krogov

$$\bigcup_{i=1}^n K_i, \quad K_i := \left\{ z \in \mathbb{C}; |z - a_{ii}| \leq \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}| \right\},$$

$$n = J \cdot K.$$

Nadgradnja tega izreka (Olga Taussky) je
naslednja: če je A irreducibilna, lahko
 $\lambda \in \sigma(A)$ leži na robu enega od krogov
le, če leži na robu vseh.

Iz lastnosti matrike sledi, da je

$$|a_{ii}| = a_{ii} \geq \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}|$$

($1 \geq 2\theta_x + 2\theta_y = 1$, toda za nekatere
vrstice je $1 > \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}|$).

Iz zgoraj omenjenega izreka za irreducibilne
matrike sledi

$$0 \in J_i, \text{ za } \forall i \Leftrightarrow a_{ii} = \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}|, \forall i, \quad (44)$$

kar očitno ni res.

Opombe:

- Matrike A ji lahko zelo velika.
V 2D, na primer, na kvadratu $[0, a] \times [0, a]$
pri $J=K=100$ dobimo matriko $A \in \mathbb{R}^{10^4 \times 10^4}$.
- Direktna metode postanejo neuporabne.
- Občutljivost matrike raste z majhnjenjem
kvadratov δ_x in δ_y .
- Pravi pristop so iterativne metode,
kjer izhajamo tudi dejstvo, da
je A razpršena (največkrat nimamo ničelnih
elementov v vrstici)
- Časoma ni pravaša zahtevnost se
bistveno bolj povečuje s
dimenzijah.

Občutljivost sistema linearnih enačb

(45)

O potencialnih napakah pri reševanju sistema $Ax = b$ odloča ponavadi občutljivost:

$$\kappa_2(A) = \|A\|_2 \|A^{-1}\|_2$$

Ker je A simetrična ($A = A^T$), je

$$\|A\|_2 = \max \sqrt{\sigma(A^T A)} = \max \sqrt{\sigma(A^2)}$$

$$= \max \sqrt{\sigma^2(A)} = \max |\sigma(A)|$$

$$= \max_{1 \leq i \leq K} |\lambda_i(A)|$$

Podobno je

$$\|A^{-1}\|_2 = \frac{1}{\min_{1 \leq i \leq K} |\lambda_i(A)|}$$

Zanimajo nas lastni pari (λ, w) ,

$$Aw = \lambda w, \quad w \neq 0.$$

Iškanje lastnih parov A je ekvivalentno (46)
 iškanju lastnih parov diskrétnega
 Laplaceovega operatorja

$$-\Delta_{\delta} w(x, y) = \frac{\lambda}{\delta^2} w(x, y); (x, y) \in D_{\delta},$$

$$w(x, y) = 0, (x, y) \in \partial D_{\delta}.$$

Funkcija w iščemo s separacijo
 spremenljivk: $w(x, y) = \phi(x) \psi(y)$, $(x, y) \in D_{\delta}$.
 Od tod dobimo enačbo

$$-\Delta_{\delta} \phi(x) \psi(y) = \frac{\lambda}{\delta^2} \phi(x) \psi(y)$$

ter robne pogoje

$$\phi(0) = 0, \phi(a) = \phi(x_{j+1}) = 0,$$

$$\psi(0) = 0, \psi(b) = \psi(y_{k+1}) = 0.$$

Zgornjo enačbo delimo z $\phi(x) \psi(y)$ in
 preuredimo v

$$-\frac{\phi(x - \delta x) + 2\phi(x) - \phi(x + \delta x)}{\delta x^2 \phi(x)} + \frac{-\psi(y - \delta y) + 2\psi(y) - \psi(y + \delta y)}{\delta y^2 \psi(y)} = \frac{\lambda}{\delta^2}$$

Ker ji prvi sumand odvisev samo od x , drugi pa le od y , morda bti
oba konstanti. Torej

(47)

$$\frac{-\phi(x-\delta x) + 2\phi(x) - \phi(x+\delta x)}{\delta x^2 \phi(x)} = \frac{\xi}{\delta^2} = ik_1$$

$$\frac{-\psi(y-\delta y) + 2\psi(y) - \psi(y+\delta y)}{\delta y^2 \psi(y)} = \frac{\eta}{\delta^2} = ik_2$$

Prvi torej $\xi + \eta = \lambda$. Poiskati
moramo torej par (ξ, ϕ) in (η, ψ) .

Za oba para ji račun podoben, zato
si ogledamo le par (ξ, ϕ) .

Vedeti mora

$$\begin{aligned} & -\phi(x_j - \delta x) + 2\phi(x_j) - \phi(x_j + \delta x) \\ & = \frac{\delta x^2}{\delta^2} \xi \phi(x_j), \quad j = 1, 2, \dots, J. \end{aligned}$$

Slednje s poenotavo v

(48)

$$-\phi(x_j - \delta x) + \left(2 - \frac{F}{\theta x}\right) \phi(x_j) - \phi(x_j + \delta x) = 0,$$

$$j = 1, 2, \dots, J.$$

Dobili smo sistem linearnih diferencialnih enačb s konstantnimi koeficienti.

Rešitev iščemo z Lagrangeovim nastavitvijo

$$\phi(x_j) = \alpha^{x_j} = (\alpha^{\delta x})^j =: \omega^j$$

Dobimo

$$\begin{aligned} & -\alpha^{x_j - \delta x} + \left(2 - \frac{F}{\theta x}\right) \alpha^{x_j} - \alpha^{x_j + \delta x} \\ &= -\frac{\omega^{j-1}}{\omega} + \left(2 - \frac{F}{\theta x}\right) \omega^j - \omega^{j+1}, \quad \omega = 0, \end{aligned}$$

Odtlej

$$\omega^{j-1} \left(-1 + \left(2 - \frac{F}{\theta x}\right) \omega - \omega^2\right) = 0.$$

Ker je $\omega \neq 0$, lahko slednje zapisemo kot

$$\frac{\Sigma}{t_x} = 2 - \omega - \frac{1}{\omega}.$$

(49)

Toda ista enačba drži tudi za $\frac{1}{\omega}$, zato se splošna rešitev sistema dof. enačb glasi

$$\phi(x_j) = C \alpha^{x_j} + D \alpha^{-x_j} = C \omega^j + D \omega^{-j}$$

S pomočjo robnih pogojev določimo konstanti C in D :

Iz $\phi(0) = 0$ dobimo $C + D = 0$ in

$$\begin{aligned} \text{zato } \phi(x_j) &= C (\alpha^{x_j} - \alpha^{-x_j}) = \\ &= C (\omega^j - \omega^{-j}), \quad C \neq 0. \end{aligned}$$

Drugi robni pogoj $\phi(x_{j+1}) = 0$ implicira

$$\phi(x_{j+1}) = C (\omega^{j+1} - \omega^{-(j+1)}) = 0,$$

$$\text{torej } \omega^{2(j+1)} - 1 = 0.$$

Rešitve so na enotni krožnici,
p-ta ji enaka

$$\omega_p = e^{i \frac{2\pi}{2(J+1)} p} = e^{i \frac{\pi}{J+1} p}, \quad p=1,2,\dots,2J+1 \quad (50)$$

Dovolj pa je omejiti le na $1 \leq p \leq J+1$,
ker za druge indekse dobimo $\phi(x_j)$, ki
so od prejšnjih ločijo le v predznakih:

$$\frac{1}{\omega_p} = e^{-i \frac{\pi}{J+1} p} = e^{+i \frac{\pi}{J+1} (2(J+1)-p)} \\ = \omega_{2(J+1)-p}$$

Če izberemo $p = J+1$, dobimo $\omega_{J+1} = -1$

$$\text{in } \omega_{J+1}^j - \omega_{J+1}^{-j} = (-1)^j - \frac{1}{(-1)^j} = 0, \text{ kar}$$

pomeni, da ta izbira ne doda nove
neodvisne rešitve. Vse neodvisne
dobimo torej za indekse $p=1,2,\dots,J$.

Konstanta C je še poljubna, zato
jo izberemo tako, da bodo komponente
lastnega vektora realne:

$$\phi_p(x_j) = C (\omega_p^j - \omega_p^{-j}) = \\ = C \left(e^{i \frac{\pi}{J+1} p j} - e^{-i \frac{\pi}{J+1} p j} \right) =$$

$$= C 2i \sin \frac{\pi}{J+1} p j$$

(51)

Torej lahko za C vzamemo $C = \frac{1}{2i} = -\frac{i}{2}$

$$\text{in} \quad \phi_j(x_j) = \sin\left(\frac{\pi}{J+1} p j\right), j=1, 2, \dots, J.$$

$$\text{Izberimo } \omega_p \text{ in} \quad iz \quad \frac{\varepsilon}{\theta_x} = 2 - \omega - \frac{1}{\omega}$$

izračunamo ε , torej

$$\varepsilon = \theta_x \left(2 - \omega_p - \frac{1}{\omega_p} \right) = \theta_x \left(2 - \left(e^{i \frac{\pi}{J+1} p} + e^{-i \frac{\pi}{J+1} p} \right) \right) = \theta_x \left(2 - 2 \cos\left(\frac{\pi}{J+1} p\right) \right)$$

$$= 2 \theta_x \left(2 \sin^2\left(\frac{\pi}{2} \frac{p}{J+1}\right) \right) = 4 \theta_x \sin^2\left(\frac{\pi}{2} \frac{p}{J+1}\right),$$

$j=1, 2, \dots, J.$

Podobno delimo

$$\gamma_q = 4 \theta_y \sin^2\left(\frac{\pi}{2} \frac{q}{K+1}\right), q=1, 2, \dots, K$$

Latence računati so torej

$$\lambda_{p2} = \varepsilon_p + \gamma_q, \quad p=1, 2, \dots, J, q=1, 2, \dots, K.$$

Za lastno vrednost

(52)

$$\lambda_{p2} = 4\theta_x^2 \operatorname{sh}^2\left(\frac{\pi}{2} \frac{p}{j+1}\right) + 4\theta_y^2 \operatorname{sh}^2\left(\frac{\pi}{2} \frac{q}{k+1}\right)$$

$$\geq \lambda_{11}^2$$

Vse lastne vrednosti so pozitivne,
najmanjša je λ_{11} .

Za majhna koraka δx in δy jo
lahko ocenimo z

$$\begin{aligned} \lambda_{11} &= 4 \left(\theta_x \operatorname{sh}^2\left(\frac{\pi}{2(j+1)}\right) + \theta_y \operatorname{sh}^2\left(\frac{\pi}{2(k+1)}\right) \right) \\ &\quad (j+1)\delta x = a, (k+1)\delta y = b \\ &\stackrel{\checkmark}{=} 4 \left(\theta_x \operatorname{sh}^2 \frac{\pi \delta x}{2a} + \theta_y \operatorname{sh}^2\left(\frac{\pi \delta y}{2b}\right) \right) \\ &= 4 \left(\theta_x \frac{\pi^2 \delta x^2}{4a^2} + \theta_y \frac{\pi^2 \delta y^2}{4b^2} \right) + O(\delta x^4 + \delta y^4) \\ &= \delta^2 \pi^2 \left(\frac{1}{a^2} + \frac{1}{b^2} \right) + O(\delta x^4 + \delta y^4) \end{aligned}$$

Najveço lastno vrednost pa ocenimo (53)

8

$$\lambda_{JK} < 4\theta_x + 4\theta_y = 2.$$

Torej smo dokazali izrek.

Izrek: Za občutljivost sistema enač $Au = \delta f + g = b$ za majhna kraka δx in δy velja

$$\kappa(A) = \|A\|_2 \|A^{-1}\|_2 = \delta \left(\frac{1}{\delta^2} \right) = \delta \left(\frac{\delta_x^2 \delta_y^2}{\delta_x^2 + \delta_y^2} \right)^{-1}$$

Če sta kraka enake, torej $\delta x = \delta y = h$,

$$\text{ji } \kappa(A) = O(h^{-2}).$$

Občutljivost torej močno narasča z manjšanjem h .