Pre relatívnu chybu riešenia systému lineárnych rovníc $Ax = b$ so štvorcovou regulárnou maticou A sme našli odhad

$$R(x) \leq \frac{c(A)}{1 - c(A)R(A)} \left(R(b) + R(A)\right) \tag{1}$$

Prirodzená otázka samozrejme je, nakoľko je tento odhad dobrý a či ho nie je možné zlepšiť. Nasledujúci príklad ukazuje, že odpoveď na druhú otázku je nie,

Príklad: Pozrime sa na systém $Ax = b$ s kde

$$A = \begin{pmatrix} 1 & 0.99 \\ 0.99 & 0,98 \end{pmatrix} \qquad \text{a} \qquad b = \begin{pmatrix} 1.99 \\ 1,97 \end{pmatrix}.$$

Hneď vidíme, že presné riešenie je $\bar{x} = (1,\ 1)^T$.

Vezmime teraz systém $Ax = b + \Delta b$, pričom $\Delta b = (10^{-4},\ 10^{-4})^T$. Riešením takto pozmeneného systému je $x = (-0.97,\ 2.99)^T$. Z toho spočítame $\Delta x = x - \bar{x} = (-1.97,\ 1.99)^T$ a relatívnu chybu riešenia

$$R_\infty(x) = \frac{||\Delta x||_\infty}{||x||_\infty} = 1.99$$

Aby sme mohli použiť nájdený všeobecný odhad, potrebujeme spočítať číslo podmienenosti $c(A)$. Nájdeme preto inverznú maticu k matici A a normy matíc A a A^{-1}

$$A^{-1} = \begin{pmatrix} -9800 & 9900 \\ 9900 & -10000 \end{pmatrix}, \qquad ||A||_\infty = 1.99, \qquad ||A^-1||_\infty = 19900.$$

Z toho spočítame

$$c_\infty(A) = 1.99 \cdot 10^4$$

V našej situácii máme $R(A) = 0$ a $R_\infty(b) = 10^-4/1.99$, takže dostávame

$$1.99 = R_\infty(x) \leq c(A)R(b) = 1.99^2 \cdot 10^4 \cdot 10^4/1.99 = 1.99$$

Takže v tejto situácii sa uvedený odhad dosahuje.

Z uvedeného príkladu vidno, že získaný odhad relatívnej chyby riešenia nie je možné vylepšiť. Na druhej strane však nemôžeme povedať, že veľké číslo podmienenosti automaticky znamená, že malé chyby v pravej strany (alebo matice) sa automaticky prejavia vo veľkej relatívnej chybe riešenia, čo ukazuje nasledujúci príklad.

Príklad: Vezmime systém $Ax = b$ s kde

$$A = \begin{pmatrix} 1 & 0.99 \\ 0.99 & 0,98 \end{pmatrix} \qquad \text{a} \qquad b = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Opäť vezmime $\Delta b = (10^{-4},\ 10^{-4})^T$. Potom riešenie systému $Ax = b + \Delta b$ môžeme nájsť ako

$$x = A^{-1}b + A^{-1}\Delta b = (-1.97 \cdot 10^4, 1.99 \cdot 10^4)^T + A^{-1}\Delta b$$

Potom $A^{-1}b$ je zrejme presné riešenie $\bar{x}$ a $A^{-1}\Delta b = \Delta x$. Z toho pre relatívnu chybu riešenia máme

$$R_\infty(x) = \frac{||x||_\infty}{||\bar{x}||_\infty} = \frac{||A^{-1}\Delta b||_\infty}{||\bar{x}||_\infty} \leq \frac{||A^{-1}||_\infty ||\Delta b||_\infty}{||\bar{x}||_\infty} = \frac{1.99 \cdot 10^4}{1.99 \cdot 10^4}||\Delta b||_\infty$$
$$= ||\Delta b||_\infty = \frac{||\Delta b||_\infty}{||b||_\infty} = R_\infty(b)$$

čiže

$$R_\infty(x) \leq R_\infty(b)$$

Takže v tomto prípade chyba riešenia nepresiahne chybu pravej strany. Avšak odhad chyby z podmienenosti matice dáva

$$R_\infty(x) \leq c_\infty(A)R_\infty(b) = 1.99 \cdot 10^4 ||b||_\infty$$

Z druhého príkladu vidno, že napriek zlej podmienenosti matice chyba riešenia môže byť malá. Číslo podmienenosti teda nehovorí veľa o presnosti získaného riešenia, ale skôr o tom, nakoľko môžeme veriť, že získané riešenie je blízko presnému riešeniu. Veľké číslo podmienenosti teda znamená veľkú neistotu v presnosti riešenia.

Ako sme videli, voľba pravej strany má vplyv na relatívnu chybu riešenie a skutočná relatívna chyba riešenia sa môže od odhadu (1) výrazne líšiť. Prirodzená otázka teda je, ako veľmi sa skutočná relatívna chyba líši od odhadu (1) a ako na ňu vplýva pravá strana.

Odpoveď na túto otázku získame použitím singulárneho rozkladu. Z rozkladu $n \times n$ matice A máme $A = U\Sigma V^T$, kde $\Sigma = \mathrm{diag}(d_1, \ldots, d_n)$, $d_1 \geq d_2 \geq \cdots \geq d_n$. Potom $AV = \Sigma U$, preto $Av_i = d_i u_i$, pre $i = 1, \ldots, n$.

Uvažujme systémy

$$Ax = b, \qquad\qquad Ax = b + \Delta b. \tag{2}$$

a) Zvoľme $b = u_1$ a $\Delta b = \varepsilon u_n$. Potom riešenia sústav

$$U\Sigma V^T x = u_1, \qquad\qquad U\Sigma V^T x = u_1 + \varepsilon u_n$$

sú

$$x = V\Sigma^{-1}U^T u_1 = V\Sigma^{-1}e_1 = \frac{1}{d_1}Ve_1 = \frac{1}{d_1}v_1$$

a

$$x + \Delta x = V\Sigma^{-1}U^T(u_1 + \varepsilon u_n) = \frac{1}{d_1}v_1 + \varepsilon V\Sigma^{-1}U^T u_n = \frac{1}{d_1}v_1 + \varepsilon\frac{1}{d_n}v_n.$$

Z toho ľahko získame Δx a po aplikácii 2-normy dostaneme

$$\|\Delta x\|_2 = \varepsilon\frac{1}{d_n}\|v_n\|_2 = \frac{\varepsilon}{d_n}$$
$$\|x\|_2 = \frac{1}{d_1},$$

a preto relatívna chyba riešenia v 2-norme je

$$R_2(x) = \frac{\|\Delta x\|_2}{\|x\|_2} = \varepsilon \frac{d_1}{d_n} = \varepsilon c_2(A).$$

Ak si uvedomíme, že $R_2(b) = \frac{\|\Delta b\|_2}{\|b\|_2} = \varepsilon$, relatívna chyba riešenia bude

$$R_2(x) = \varepsilon c_2(A) = c_2(A) R_2(b).$$

b) Zvoľme teraz $b = u_n$ a Δb ľubovoľné také, aby $R(b) = \varepsilon < 1$. Potom riešenia sústav (2) sú

$$x = \frac{1}{d_n} v_n, \qquad x + \Delta x = \frac{1}{d_n} v_n + V \Sigma^{-1} U^T \Delta b.$$

Preto

$$\|x\|_2 = \frac{1}{d_n}, \qquad \text{a} \qquad \|\Delta x\|_2 = \|V \Sigma^{-1} U^T \Delta b\|_2 \le \frac{1}{d_n} \|\Delta b\|_2.$$

Potom pre relatívnu chybu riešenia platí

$$R_2(x) \le \frac{1}{d_n} \|\Delta b\|_2 \Big/ \frac{1}{d_n} = \|\Delta b\|_2 < 1.$$

Ak zvolíme $\Delta b = \varepsilon u_1$, tak

$$R_2(x) = \|\frac{\varepsilon v_1}{d_1}\|_2 \Big/ \|\frac{v_n}{d_n}\|_2 = \varepsilon \frac{d_n}{d_1} = \frac{\varepsilon}{c_2(A)}.$$

Z týchto úvah vidno, že relatívna chyba riešenia môže byť za vhodných okolností omnoho menšia než chyba pravej strany. Tieto dva prípady sú v skutočnosti hraničné, teda máme odhady

$$\frac{R_2(b)}{c_2(A)} \le R_2(x) \le c_2(A) R(b). \tag{3}$$

Číslo podmienenosti symetrických matíc

Pre 2-normu regulárnej matice A platí

$$\|A\|_2 = \left[\rho(A^T A)\right]^{\frac{1}{2}},$$

kde ρ je spektrálny polomer. Pre symetrickú maticu A potom dostaneme

$$\rho(A^T A) = \rho(A^2) = |\lambda|^2_{max}$$

t.j. druhú mocninu najväčšej vlastnej hodnoty matice A, preto $\|A\|_2 = |\lambda|_{max}$.

Podobne sa odvodí aj $\|A^{-1}\|_2 = \left|\frac{1}{\lambda}\right|_{max} = \frac{1}{|\lambda|_{min}}$. Číslo podmienenosti symetrickej matice A v 2-norma sa teda spočíta ako

$$c_2(A) = \frac{|\lambda|_{max}}{|\lambda|_{min}}.$$

Pre nesymetrické matice a ľubovoľnú normu platí vždy odhad

$$\|A\| \geq \rho(A) = |\lambda|_{max}$$

a

$$\|A^{-1}\| \geq \rho(A^{-1}) = \frac{1}{|\lambda|_{min}}.$$

Preto pre číslo podmienenosti v ľubovoľnej norme dostávame dolný odhad

$$c(A) \geq \frac{|\lambda|_{max}}{|\lambda|_{min}}.$$

Teda ak podiel absolútnych hodnôt najväčšieho a najmenšieho vlastného čísla matice A je veľký, matica je zle podmienená.
Poznámka: Ak riešime systém $Ax = b$ použitím LU-rozkladu, môže sa stať, že chyba riešenia bude väčšia než $c(A)R(b)$, pretože matice L a U môžu mať horšie čísla podmienenosti. Preto pri zle podmienených maticiach je vhodnešie použiť QR-rozklad.

Hranice rezidua ako kritérium presnosti

Klasický postup pre overenie, či nájdene riešenie systému rovníc je naozaj riešením je dosadenie riešenia do systému (tzv. skúška správnosti). Čiže pre systém $Ax = b$ nájdeme riešenie x a toto riešenie spĺňa $Ax = b$. Samozrejme pri vzniku chyby v riešení (napr. zaokrúhľovaním, alebo inou stratou presnosti) nájdené riešenie x nebude vyhovovať danému systému. Čo ale vieme zistiť je, ako veľmi sa líši hodnota $A\bar{x}$ od pravej strany b. Tento rozdiel nazveme reziduum a označíme ho r, teda máme

$$r = b - Ax$$

Prirodzene vyvstáva otázka, či reziduum podáva dobrú informáciu o presnosti riešenia. Ukazuje sa, že nie
Príklad: Uvažujme systém $Ax = b$ v maticovom tvare

$$\left(\begin{array}{cc|c} 1 + \frac{1}{1000} & 1 & 2.001 \\ 1 & 1 & 2 \end{array} \right).$$

Presné riešenie tohoto systému je $\bar{x} = (1, 1)^T$.
Vezmime $x^{(1)} = (1.5, 0.5)^T$ a $x^{(2)} = (0.99, 0.99)^T$ a spočítajme pre ne rezidua

$$r^{(1)} = b - Ax^{(1)} = \begin{pmatrix} 2.001 \\ 2 \end{pmatrix} - \begin{pmatrix} 2,0015 \\ 2 \end{pmatrix} = \begin{pmatrix} -0.0005 \\ 0 \end{pmatrix}$$
$$r^{(2)} = b - Ax^{(2)} = \begin{pmatrix} 2.001 \\ 2 \end{pmatrix} - \begin{pmatrix} 1.98099 \\ 1,98 \end{pmatrix} = \begin{pmatrix} 0.02001 \\ 0.02 \end{pmatrix}.$$

Potom pre veľkosti reziduií v 2-norme platí

$$\|r^{(1)}\|_2 < \|r^{(2)}\|_2.$$

Takže ak by sme reziduum brali ako kritérium presnosti riešenia, v tomto prípade by nám indikovalo, že $x^{(2)}$ je menej presné riešenie než $x^{(1)}$, napriek tomu, že $x^{(2)}$ je zjavne v 2-norme bližšie k presnému riešeniu než $x^{(1)}$.

Že reziduum nie je dobré kritérium presnosti môžeme nahliadnuť aj z nasledovných úvah. Ak označíme x vypočítané riešenie a $\bar{x}$ presné riešenie systému $Ax = b$, tak máme

$$r = b - Ax = A\bar{x} - Ax = A(\bar{x} - x) = A\Delta x,$$

teda $\Delta x = A^{-1}b$ a Δx môže byť veľké aj pre malé r.

Zo znalosti rezidua môžeme dostať nasledovné odhady relatívnej chyby riešenia.

$$R(x) = \frac{\|\Delta x\|}{\|\bar{x}\|} = \frac{\|A^{-1}r\|}{\|\bar{x}\|} \leq \frac{\|A^{-1}\|\|r\|}{\|\bar{x}\|} = \frac{\mathrm{c}(A)}{\|A\|}\frac{\|r\|}{\|\bar{x}\|} \leq \frac{\mathrm{c}(A)\|r\|}{\|A\bar{x}\|} = \frac{\mathrm{c}(A)\|r\|}{\|b\|}$$

$$R(x) = \frac{\|\Delta x\|}{\|\bar{x}\|} = \frac{\|A\|\|\Delta x\|}{\|A\|\|\bar{x}\|} \geq \frac{\|A\Delta x\|}{\|A\|\|\bar{x}\|} = \frac{\|r\|}{\|A\|\|A^{-1}b\|} \geq \frac{\|r\|}{\|A\|\|A^{-1}\|\|b\|} = \frac{\|r\|}{\mathrm{c}(A)\|b\|}.$$

Teda

$$\frac{1}{\mathrm{c}(A)}\frac{\|r\|}{\|b\|} \leq R(x) \leq \mathrm{c}(A)\frac{\|r\|}{\|b\|}. \qquad (4)$$

Iteračné spresnenie riešenia

Napriek tomu že reziduum nemožno vždy dobre využiť na určenie presnosti riešenia, je možné ho výhodne použiť na spresnenie už nájdeného približného riešenia. Idea je nasledovná.

Predpokladajme, že sme našli približné riešenie $x^{(1)}$ systému $Ax = b$. Potom presné riešenie $\bar{x}$ je možné zapísať ako $\bar{x} = x^{(1)} + \Delta x^{(1)}$. Potom

$$b = A\bar{x} = Ax^{(1)} + A\Delta x^{(1)},$$

a teda

$$A\Delta x^{(1)} = b - Ax = r^{(1)}, \qquad (5)$$

teda $\Delta x^{(1)}$ je riešením systému s maticou A a pravou stranou rovnou reziduu $r^{(1)}$. Numerickým riešením tohoto systému získame približne hodnotu $\Delta x^{(1)}$.Potom môžeme položiť $x^{(2)} = x^{(1)} + \Delta x^{(1)}$, pričom nové reziduum bude $r^{(2)} = b - Ax^{(2)}$ a celý postup môžeme zopakovať. Celý algoritmus môžeme zhrnúť v do niekoľkých krokov.

- Zvoľme $x^{(0)} = 0$ a $r^{(0)} = b$.
- Pre $k = 1, 2, \ldots$
  a) Vypočítame $\Delta\tilde{x}^{(k)}$ ako riešenie sústavy $Ax = r^{(k)}$.
  b) Vypočítame $x^{(k+1)} = x^{(k)} + \Delta x^{(k)}$
  c) Vypočítame $r^{(k+1)} = b - Ax^{(k+1)}$.

V tomto algoritme sa dá výhodne využiť LU rozklad, keďže v kroku a) riešime pre každé k systém s tou istou maticou A.