Numerisk Analys, MMG410. Lecture 6.

Störningsteori för $Ax = b$

Vi använder normer för att studera konditionstalet för problemet $Ax = b$. Vi vill veta vad som händer med x när A och/eller b ändras. Vi kommer endast att ändra b.

Sats

Låt A vara ickesingulär och $Ax = b \neq 0$. Om $Ay = b + f$ så gäller

$$\frac{||x - y||}{||x||} \leq ||A||\,||A^{-1}||\frac{||f||}{||b||}$$

Bevis.

$$Ay = b + f \text{ och } Ax = b \Rightarrow A(y - x) = f \Rightarrow$$

$$y - x = A^{-1}f \Rightarrow \; ||y - x|| = ||A^{-1}f|| \leq ||A^{-1}||\,||f||.$$

Men $Ax = b$, varför $\|x\| \leq \|A^{-1}\|\|b\|$ eller $1/||x|| \leq ||A||/||b||$. $\square$

Man kan visa liknande satser för fallen när A eller A *och* b störs. Konditionstalet betecknas med kappa, dvs. $\kappa(A) = ||A||\,||A^{-1}||$.

Konditionstal för $Ax = b$

Antag att $||\cdot||$ är en operatornorm. Då gäller:

- $\kappa(A) \geq 1$, ty $1 = ||AA^{-1}|| \leq ||A||\,||A^{-1}||$
- I är perfekt konditionerad, ty $\kappa(I) = 1$
- konditionstalet är skalningsoberoende: $\kappa(\alpha A) = \kappa(A)$
- $\kappa(A) = \infty$ om A är singulär

Om A är singulär kan det finnas ingen eller oändligt många lösningar. Vi förväntar oss problem om A *nästan* är singulär. Om $\kappa(A)$ är stort så finns en matris E med liten norm, så att $A + E$ blir singulär. Vi säger att A "ligger nära" mängden av singulära matriser. Om däremot $\kappa(A) \approx 1$, måste $||E||$ vara stor för att $A + E$ ska bli singulär.
Man kan visa att de E som gör $A + E$ singulär och har minst norm uppfyller $||E|| = ||A||/\kappa(A)$.

Konditionstal för $Ax = b$

Determinanten för A är *inte* ett bra mått på att A nästan är singulär vilket följande exempel illustrerar

Exempel

$$\det(\alpha I) = \alpha^n \text{ medan } \kappa(\alpha I) = 1.$$

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0.1 \end{bmatrix}, \det(A) = 0.1, \kappa(A) = 10$$

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0.1 & 0 & 0 \\ 0 & 0 & 0.1 & 0 \\ 0 & 0 & 0 & 0.1 \end{bmatrix}, \det(A) = 0.001, \kappa(A) = 10$$

Konditionstal för $Ax = b$

Exempel (Hur väl stämmer satsen?)

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 10^{-8} \end{bmatrix}, \; b = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \; x = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \; \kappa_\infty(A) = 10^8$$

$$\text{Låt } f = \begin{bmatrix} 10^{-9} \\ 10^{-9} \end{bmatrix} \Rightarrow y - x = \begin{bmatrix} 10^{-9} \\ 10^{-1} \end{bmatrix}, \frac{||x-y||_\infty}{||x||_\infty} = \frac{0.1}{1} = 0.1$$

$$\text{och dessutom } \kappa_\infty(A)\frac{||f||_\infty}{||b||_\infty} = 10^8 \frac{10^{-9}}{1} = 0.1$$

dvs. likhet i gränsen. Tar vi istället

$$f = \begin{bmatrix} 10^{-9} \\ 0 \end{bmatrix} \Rightarrow y - x = \begin{bmatrix} 10^{-9} \\ 0 \end{bmatrix}, \frac{||x-y||_\infty}{||x||_\infty} = \frac{10^{-9}}{1} = 10^{-9}$$

$$\text{men fortfarande } \kappa_\infty(A)\frac{||f||_\infty}{||b||_\infty} = 10^8 \frac{10^{-9}}{1} = 0.1$$

Tolkning av satsen

Antag att elementen i x respektive y är ungefär lika stora. Då gäller $x \approx x_k e$ och $y \approx y_k e$ där $e = (1, 1, ..., 1)$. Vi får

$$\frac{||x - y||}{||x||} \approx \frac{||(x_k - y_k)e||}{||x_k e||} = \frac{|x_k - y_k|}{|x_k|}$$

dvs. normen uppskattar det elementvisa felet. För detta specialfall gäller

$$\frac{|x_k - y_k|}{|x_k|} \lessapprox \kappa(A)\frac{||f||}{||b||}$$

Ovanstående säger att det relativa felet i varje komponent begränsas av det relativa felet i indata multiplicerat med konditionstalet för A.
Om t.ex. $||f||/||b|| = 0.5 \cdot 10^{-k}$ (k decimaler) och $\kappa(A) \approx 10^p$ så

$$\frac{|x_k - y_k|}{|x_k|} \lessapprox 10^p \cdot 0.5 \cdot 10^{-k} = 0.5 \cdot 10^{p-k}$$

Som tumregel får vi således följande:

Tolkning av satsen

$$\boxed{\text{Om } \kappa(A) = 10^p \text{ så riskerar vi att tappa } p \text{ siffror.}}$$

Antag nu att x innehåller element av olika storleksordning, t.ex. $x = [1, 10^{-3}]^T$ och att vi använder $||\cdot||_\infty$. Om $p - k = -3$ så

$$\max\{|1 - y_1|, |10^{-3} - y_2|\} \leq 0.5 \cdot 10^{-3}$$

så att

$$1 - 0.5 \cdot 10^{-3} \leq y_1 \leq 1 + 0.5 \cdot 10^{-3}$$

och

$$10^{-3} - 0.5 \cdot 10^{-3} \leq y_2 \leq 10^{-3} + 0.5 \cdot 10^{-3}$$

Normer kan vara trubbiga instrument.

Tolkning av satsen

Hur stora fel har vi i indata? Låt oss studera två fall.
Exakt indata: Vi får eventuellt avrundningsfel när $a_{j,k}$ och b_k lagras i minnet. Relativa felet (per komponent) är cirka ϵ_{mach}. Vi får även avrundningsfel när vi löser $Ax = b$.
Vi kan troligen tillåta ganska stora $\kappa(A)$, men det beror på hur många siffror vi behöver. Om vi har stora krav, eller för väldigt stort $\kappa(A)$ kan vi minska ϵ_{mach} genom att använda t.ex. Maple eller Mathematica. Att räkna med många fler siffror går dock mycket långsammare (mjukvara, inte hårdvara).
Indata med osäkerhet (mätdata): Ger normalt större begränsningar på hur stort $\kappa(A)$ vi kan tillåta, eftersom vi oftast inte mäter väldigt noga.
Att räkna med mindre ϵ_{mach} ger oftast inte en mer exakt lösning, ty om $\kappa(A)$ är måttligt stort så dominerar mätfelen över avrundningsfelen.

Uppskattning av $\kappa(A)$

Att beräkna A^{-1} tar mycket tid och minne om A är stor. **cond** i Matlab använder **svd** för $||\cdot||_2$ och explicit **inv** för andra normer. För stora matriser kan man använda **condest** som uppskattar $||A^{-1}||$ genom att lösa det linjära ekvationssystem (samt andra listigheter). Även LAPACK kan ge en sådan uppskattning när man löser $Ax = b$. Uppskattningen kostar nästan ingenting, eftersom man uppnyttjar den LU-faktorisering som redan beräknats.

Tolkning av residualen

Vad säger residualen $r = b - A\hat{x}$? ($\hat{x}$ är den beräknade lösningen).

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 10^{-8} \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \hat{x} = \begin{bmatrix} 1 \\ 10^4 \end{bmatrix}, \quad r = \begin{bmatrix} 0 \\ -10^{-4} \end{bmatrix}$$

Kan visa $(A+E)\hat{x} = b$, $||E||_2 = ||r||_2/||\hat{x}||_2$. ($||E||_2 \approx 10^{-8}$ i exemplet.)

En liten residual betyder att vi har löst nästan rätt problem. I framåtriktningen kommer $\kappa(A)$ in:

$$r = b - A\hat{x} = Ax - A\hat{x} = A(x - \hat{x}) \Leftrightarrow x - \hat{x} = A^{-1}r.$$

Vi får $||x - \hat{x}|| \leq ||A^{-1}||\,||r||$. Vidare om $b \neq 0$ får vi

$$||b|| \leq ||A||\,||x|| \Leftrightarrow \frac{1}{||x||} \leq \frac{||A||}{||b||}$$

Kombinerar vi dessa olikheter får vi

$$\boxed{\frac{||x - \hat{x}||}{||x||} \leq \kappa(A)\frac{||r||}{||b||}}$$

Felet i lösningen kan vara godtyckligt stort även om $||r||$ är liten.

En mer generell störningssats

Sats

Låt $Ax = b \neq 0$ *och* $(A+F)y = b+f$ *där* $||F|| \leq \mu\, ||A||$ *och* $||f|| \leq \mu\, ||b||$. *Om* $r := \mu\, \kappa(A) < 1$ *så är* $(A+F)$ *ickesingulär och*

$$\frac{||y-x||}{||x||} \leq \frac{2\,r}{1-r}$$

Olikheten kan även formuleras som

$$\frac{||y-x||}{||x||} \leq \frac{2\,\kappa(A)}{1-r} \max\left\{\frac{||F||}{||A||}, \frac{||f||}{||b||}\right\}.$$

Om man enbart stör A (och inte b) kan man ta bort tvåan från olikhetens högerled.

En mer generell störningssats

Bevis:
$\delta A = F, \hat{x} = x + \delta x = y, \delta b = f.$

$$\begin{aligned} (1) \quad & Ax = b, \\ (2) \quad & (A + \delta A)\hat{x} = b + \delta b. \end{aligned}$$

(2)-(1):

$$\begin{aligned} Ax + A\delta x + \delta A x + \delta A \delta x - Ax &= \delta b, \\ A\delta x + \delta A(\delta x + x) &= \delta b, \\ A\delta x &= \delta b - \delta A \hat{x}, \\ \delta x &= A^{-1}(\delta b - \delta A \hat{x}). \end{aligned}$$

En mer generell störningssats

- Vi har:

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

Skriver om:

$$\begin{aligned} A\delta x &= -\delta A\hat{x} + \delta b, \\ A\delta x + \delta A\hat{x} &= \delta b, \\ A\delta x + \delta A(x + \delta x) &= \delta b, \\ \delta A x + (A + \delta A)\delta x &= \delta b. \end{aligned}$$

Löser $\delta A x + (A + \delta A)\delta x = \delta b$ för δx och får:

En mer generell störningssats

- Vi har:

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

  Skriver om:

$$\begin{aligned} A\delta x &= -\delta A\hat{x} + \delta b, \\ A\delta x + \delta A\hat{x} &= \delta b, \\ A\delta x + \delta A(x + \delta x) &= \delta b, \\ \delta A x + (A + \delta A)\delta x &= \delta b. \end{aligned}$$

  Löser $\delta A x + (A + \delta A)\delta x = \delta b$ för δx och får:

-

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

En mer generell störningssats

Lemma

Let $\|\cdot\|$ satisfy $\|AB\| \leq \|A\| \cdot \|B\|$. Then $\|X\| < 1$ implies that $I - X$ is invertible. $(I - X)^{-1} = \sum_{i=0}^{\infty} X^i$, and

$$\|(I - X)^{-1}\| \leq \frac{1}{1 - \|X\|}. \tag{1}$$

En mer generell störningssats

- Ta normer och dividera med $\|x\|$:

$$
\begin{aligned}
\frac{\|\delta x\|}{\|x\|} &\leq \|(I + A^{-1}\delta A)^{-1}\| \cdot \|A^{-1}\| \left(\|\delta A\| + \frac{\|\delta b\|}{\|x\|} \right) \\
&\leq \frac{\|A^{-1}\|}{1 - \|A^{-1}\| \cdot \|\delta A\|} \left(\|\delta A\| + \frac{\|\delta b\|}{\|x\|} \right) \quad (\textit{Lemma}\ ) \\
&= \frac{\|A^{-1}\| \cdot \|A\|}{1 - \|A^{-1}\| \cdot \|A\| \frac{\|\delta A\|}{\|A\|}} \left(\frac{\|\delta A\|}{\|A\|} + \frac{\|\delta b\|}{\|A\| \cdot \|x\|} \right) \\
&\leq \frac{k(A)}{1 - k(A) \frac{\|\delta A\|}{\|A\|}} \left(\frac{\|\delta A\|}{\|A\|} + \frac{\|\delta b\|}{\|b\|} \right)
\end{aligned} \tag{2}
$$

En mer generell störningssats

- Ta normer och dividera med $\|x\|$:

$$
\begin{aligned}
\frac{\|\delta x\|}{\|x\|} &\leq \|(I + A^{-1}\delta A)^{-1}\| \cdot \|A^{-1}\| \left(\|\delta A\| + \frac{\|\delta b\|}{\|x\|} \right) \\
&\leq \frac{\|A^{-1}\|}{1 - \|A^{-1}\| \cdot \|\delta A\|} \left(\|\delta A\| + \frac{\|\delta b\|}{\|x\|} \right) \ (\textit{Lemma}\ ) \\
&= \frac{\|A^{-1}\| \cdot \|A\|}{1 - \|A^{-1}\| \cdot \|A\| \frac{\|\delta A\|}{\|A\|}} \left(\frac{\|\delta A\|}{\|A\|} + \frac{\|\delta b\|}{\|A\| \cdot \|x\|} \right) \\
&\leq \frac{k(A)}{1 - k(A) \frac{\|\delta A\|}{\|A\|}} \left(\frac{\|\delta A\|}{\|A\|} + \frac{\|\delta b\|}{\|b\|} \right)
\end{aligned}
\tag{2}
$$

- relativa felet $\frac{\|\delta x\|}{\|x\|}$ är beroende på relativa felet $\frac{\|\delta A\|}{\|A\|}$ och $\frac{\|\delta b\|}{\|b\|}$ i input data.