MATEMATISKA VETENSKAPER
Chalmers tekniska högskola
Examinator: Tony Stillfjord

TMV166 Linjär algebra för M

Datorlaboration 3: Felanalys vid lösning av linjära ekvationssystem

Allmänt

I den här laborationen skall vi se hur felkällor i indata kan påverka lösningen av ett ekvationssystem. Det är viktigt att inte lita blint på MATLAB utan kritiskt granska resultatet, samt att förstå vad eventuella varningar innebär.

Mål

- Undersöka kopplingen mellan fel i indata och fel i utdata
- Se att olika metoder för att beräkna samma sak ibland kan leda till drastiskt skilda resultat

Inledning

När man behandlar ett linjärt ekvationssystem

$$Ax = b$$

så är ofta matrisen A och högerledet b behäftade med fel eller osäkerheter. De data som matas in kan t.ex. vara resultat av mätningar, vilka har en begränsad noggrannhet. Dessutom kan oftast inte talen representeras exakt i en dator. Det senare ger visserligen mycket små fel, men dessa kan förstoras av ett stort antal räkneoperationer. Med en dålig metod kan man börja med 16 siffrors noggrannhet och få resultat som inte innehåller en enda rätt siffra.

Vi har alltså en okänd *störning* (fel) δA i matrisen A och en störning δb i högerledet. Det system som vi faktiskt löser har matrisen $A + \delta A$ och högerledet $b + \delta b$ istället för A och b. Eftersom vi löser fel system så får vi inte lösningen x utan $x + \delta x$, där δx betecknar felet i x.

Vi vill nu mäta hur vektorn δx påverkas av matrisen δA och vektorn δb. För att göra det behöver vi definiera "storleken" av vektorer och matriser. För en vektor $x \in \mathbb{R}^n$ är den *Euklidiska normen* $\|x\|_2$ ett naturligt mått. Den ges av

$$\|x\|_2 = \sqrt{\sum_{k=1}^{n} x_k^2} = \sqrt{x^T x}.$$

I $\mathbb{R}^3$ och $\mathbb{R}^2$ är detta den vanliga *längden* av en vektor och för $x \in \mathbb{R}^1$ får vi det vanliga absolutbeloppet $|x|$. Andra normer som ofta används är *summanormen*

$$\|x\|_1 = \sum_{k=1}^{n} |x_k|$$

Ursprunglig datorlaboration författad av Carl-Henrik Fant och Lennart Falk, uppdaterad av Tony Stillfjord och *maximumnormen*

$$\|x\|_\infty = \max_{1\le k\le n} |x_k|$$

men dessa ger helt andra egenskaper. I fortsättningen använder vi den Euklidiska normen och skriver bara $\|x\|$ istället för $\|x\|_2$.

För en matris $A \in \mathbb{R}^{m\times n}$ så definieras den Euklidiska normen som den maximala skalan i matristransformationen $x \mapsto Ax$, dvs.

$$\|A\|_2 = \max_{\substack{x\in\mathbb{R}^n \\ x\neq 0}} \frac{\|Ax\|_2}{\|x\|_2}. \tag{1}$$

För att faktiskt beräkna $\|A\|_2$ finns andra ekvivalenta formuleringar, men för stora matriser är det ändock en relativt kostsam operation. Man får andra matrisnormer genom att byta ut den Euklidiska vektornormen till t.ex. summanormen eller maximumnormen. Dessa är billigare att beräkna, men liksom för vektorer är $\|A\|_2$ oftast den mest naturliga. Vi skriver i fortsättningen $\|A\| = \|A\|_2$.

I Matlab kan både vektor- och matris-normer beräknas med kommandot `norm`. Med `norm(x)` får man direkt den Euklidiska normen av `x`, och med `norm(x, 1)` får man summanormen. Se dokumentationen för vilka andra alternativ som finns.

Konditionstal

Vi renodlar situationen och antar att vi endast har en störning i högerledet *eller* en störning i matrisen, inte båda samtidigt. Vi antar också att A är en inverterbar $n \times n$-matris.

Störning i högerledet

Vi *vill* lösa

$$Ax = b$$

men behandlar i själva verket

$$A(x + \delta x) = b + \delta b.$$

Genom att subtrahera ekvationerna får vi att felet i lösningen uppfyller

$$A\delta x = \delta b,$$

dvs.

$$\delta x = A^{-1}\delta b. \tag{2}$$

Nu vill vi kunna jämföra storleken av störningen i högerledet med storleken av felet i lösningen. Det är då naturligt att uttrycka sambandet mellan de relativa felen

$$\frac{\|\delta b\|}{\|b\|} \quad \text{och} \quad \frac{\|\delta x\|}{\|x\|}.$$

Av definitionen (1) av matrisnorm följer att

$$\|b\| = \|Ax\| \le \|A\|\,\|x\| \quad \Rightarrow \quad \frac{1}{\|x\|} \le \frac{\|A\|}{\|b\|},$$

och tillämpat på (2) ger detta

$$\|\delta x\| \le \left\|A^{-1}\right\| \|\delta b\|.$$

Kombinerar vi de sista olikheterna får vi slutligen

$$\frac{\|\delta x\|}{\|x\|} \leq \|A\| \left\|A^{-1}\right\| \frac{\|\delta b\|}{\|b\|}.$$

Produkten $\kappa(A) = \|A\| \left\|A^{-1}\right\|$ kallas *konditionstalet* för matrisen A, och man skriver

$$\frac{\|\delta x\|}{\|x\|} \leq \kappa(A) \frac{\|\delta b\|}{\|b\|}. \tag{3}$$

En matris med stort konditionstal sägs vara *illakonditionerad.* Det är önskvärt att om möjligt undvika sådana matriser i ekvationssystem, då små störningar i högerledet kan ge upphov till förhållandevis stora fel i lösningen. Om man likväl har erhållit ett illakonditionerat system bör man kontrollera att den underliggande modellen och implementationen är korrekt; ofta beror detta på tanke- eller räknefel snarare än ett elakartat fysikaliskt system.

Störning i matrisen

Om vi istället har en störning δA i matrisen med motsvarande fel δx i lösningen så får vi istället

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

Detta leder på samma sätt som ovan till feluppskattningen

$$\frac{\|\delta x\|}{\|x + \delta x\|} \leq \kappa(A) \frac{\|\delta A\|}{\|A\|}. \tag{4}$$

Det relativa felet i lösningen ser nu något annorlunda ut på grund av dubbeltermen $\delta A\, \delta x$. Eftersom både δx och δA antas vara små så kan vi dock i princip anta att denna term är försumbar. I så fall återfår vi den tidigare situationen: $\|\delta x\| \,/\, \|x\| \leq \kappa(A) \, \|\delta A\| \,/\, \|A\|$.

Uppgifter

Hilbertmatrisen $H_n = (h_{ij}) \in \mathbb{R}^{n \times n}$ av ordning n har elementen $h_{ij} = \frac{1}{i+j-1}$, och är ett exempel på en särskilt illakonditionerad matris (stort konditionstal). Exempelvis ges Hilbertmatrisen av ordning 3 av

$$H_3 = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \end{bmatrix}.$$

MATLAB har en funktion `hilb(n)` som genererar denna matris, och en funktion `invhilb(n)` som genererar inversen med stor noggrannhet — dess element är heltal! Därför lämpar sig Hilbertmatrisen för vår undersökning.

Uppgift 1. Titta först på H_n och H_n^{-1} för några n, samt konditionstalet $\kappa(H_n)$. Detta kan beräknas i MATLAB med kommandot `cond`. Kontrollera att `cond` ger samma resultat som när du beräknar matris-normerna explicit. □

Uppgift 2. Rita en graf som visar hur konditionstalet $\kappa(H_n)$ växer med n. Eftersom det växer mycket fort, använd `semilogy` istället för `plot`. Detta ger en linlog-skala; linjär i x-led och logaritmisk i y-led. □

Uppgift 3. För att visa att illakonditionerade matriser verkligen ger problem ska vi undersöka känsligheten i att lösa systemet $H_n x = b$. Tag en Hilbertmatris av måttligt stor ordning, t.ex. $n = 4$. Välj ett högerled b och lös systemet genom att använda `rref`. Pröva också att multiplicera med inversen: $x = H_n^{-1} b$ (använd `invhilb(n)` för att beräkna H_n^{-1}!).

Nu "stör" vi matrisen H_n genom att runda av alla dess element till 4 decimaler (kommandot `round` kan utnyttjas). Lös systemet med samma högerled och jämför resultatet. Jämför även de relativa felen $\frac{\|\delta x\|}{\|x\|}$ och $\frac{\|\delta H_n\|}{\|H_n\|}$. Använd uppskattningen (4) för att bestämma det maximala felet som kan uppstå p.g.a. dessa små fel i matrisen. □

Uppgift 4. Här ska vi se hur beräkningarna i lösningsalgoritmen påverkar resultatet. Felen som uppstår är relaterade till $\kappa(H_n)$. Vi löser ekvationssystemet $H_n x = b$ för något värde på n, välj t.ex. $n = 10$ (ej större!). Tag b som en slumptalsvektor: `b = randn(n,1)`. (Vi använder `randn` som genererar *normalfördelade* slumptal för att få både positiva och negativa tal.) Jämför de lösningar du får med `rref`, `Hn\b` och `inv(Hn)*b` samt den exakta lösningen `invhilb(n)*b`. □

Uppgift 5. Skriv en funktionsfil som löser systemet $H_n x = b$ exakt för ett stort antal (N) slumpvis valda högerled b. Tag också lika många slumpvis valda felvektorer δb och lös systemet $H_n \delta x = \delta b$ exakt. För varje par av ekvationssystem, beräkna kvoten

$$\kappa = \frac{\frac{\|\delta x\|}{\|x\|}}{\frac{\|\delta b\|}{\|b\|}}.$$

Av alla dessa kvoter κ skall bara den minsta ($\kappa_{\min}$) och den största ($\kappa_{\max}$) levereras som utdata. Detta ger oss spridningen av kvoterna. Enligt (3) bör $\kappa_{\max}$ ligga i närheten av konditionstalet $\kappa(A)$. □

Uppgift 6. Till sist, jämför med hjälp av `semilogy` spridningen $[\kappa_{\min}, \kappa_{\max}]$ gentemot konditionstalet för H_n då $n = 2, \ldots, 5$. Detta kan göras genom att bygga ut funktionsfilen i föregående uppgift. □

Uppgift 7.

- Vilka slutsatser kan du dra?
- Vilken betydelse har N (antalet slumpgenererade högerled?)
- Varför måste vi öka N med storleken på H?

□

Tips för sista frågan: det maximala relativfelet antas då b och δb har var sin speciell riktning i $\mathbb{R}^n$. I själva verket skall de vara så kallade egenvektorer till H_n, där b hör ihop med det största egenvärdet $\lambda_{\max}$ och δb hör ihop med det minsta egenvärdet $\lambda_{\min}$. Vi återkommer till detta senare i kursen, men vi kan testa påståendet redan nu. Med kommandot `[V, D] = eig(Hn)` får du två matriser `V` och `D`. Kolonnerna i `V` är egenvektorerna och diagonalmatrisen `D` innehåller egenvärdena på diagonalen. Egenvärdet på plats (i, i) i `D` motsvarar egenvektorn i kolonn i av `V`.

Kontrollera alltså vilka som är de största och minsta egenvärdena till (t.ex.) H_{10} och välj b och δb till motsvarande egenvektorer. Lös sedan systemen $H_{10}\,\delta x = \delta b$ och $H_{10} x = b$ exakt. Beräkna $\frac{\|\delta x\|}{\|x\|} / \frac{\|\delta b\|}{\|b\|}$ och konditionstalet för H_{10}. De borde vara samma eller skilja ytterst lite. Testa även att beräkna $\frac{\lambda_{\max}}{\lambda_{\min}}$. Resultatet är ingen slump, men ryms inte i den här kursen.