Numerisk Analys, MMG410. Lecture 7.

Några inledande exempel

Ett vanligt problem är att vi har en matematisk modell och uppmätta värden, och vill bestämma värden på parametrar i modellen. En vanlig modell är $b = ce^{\lambda t}$ (halveringstid, befolkningstillväxt, etc.) b skulle kunna vara befolkningen vid tiden t och c är befolkningsmängden vid tiden $t = 0$.
Vi vill bestämma parametern λ genom att utföra m mätningar av b vid olika tidpunkter t_k. Vi får således m stycken par (t_k, b_k), $k = 1, 2, ..., m$. Vi antar att c är känd. Hur ska vi beräkna λ? Vi har m *olika* ekvationer

$$b_1 = ce^{\lambda t_1}, b_2 = ce^{\lambda t_2}, ..., b_m = ce^{\lambda t_m}.$$

Det är inte sannolikt att samma λ satisfierar alla ekvationerna. Vi är heller inte intresserade av att få m olika värden på λ.

Några inledande exempel

En rimlig kompromiss är att hitta ett λ som approximativt satisfierar alla ekvationerna:

$$b_1 \approx ce^{\lambda t_1}, b_2 \approx ce^{\lambda t_2}, ..., b_m \approx ce^{\lambda t_m}.$$

Detta kan formuleras på följande sätt:
Försök att göra residualerna

$$r_1 = b_1 - ce^{\lambda t_1}, r_2 = b_2 - ce^{\lambda t_2}, ..., r_m = b_m - ce^{\lambda t_m}.$$

så små som möjligt. Vi kan definiera "små" på många olika sätt (normer), t.ex.

$$\min_{\lambda} \sum_{k=1}^{m} |b_k - ce^{\lambda t_k}|$$

$$\min_{\lambda} \left(\sum_{k=1}^{m} (b_k - ce^{\lambda t_k})^2 \right)^{1/2}, \quad \min_{\lambda} \max_{1 \leq k \leq m} |b_k - ce^{\lambda t_k}|$$

Några inledande exempel

Dessa förslag är inte slumpvis utvalda. Vi inför residualvektorn $r = [r_1, ..., r_m]^T$. Då kan de tre måtten på föregående sida skrivas som

$$\min_{\lambda} ||r||_1, \quad \min_{\lambda} ||r||_2, \quad \min_{\lambda} ||r||_\infty.$$

Olika normer kommer i regel att ge olika värden på λ. Varje λ är dock det bästa valet för den givna normen.
Det finns oändligt många frågor, med olika svar. Varje svar är dock korrekt svar på den givna frågan. Det finns normalt inte ett bästa värde på λ.

Några inledande exempel

Man kan ha modeller med fler än en parameter. Ett vanligt exempel är att man vill anpassa mätpunkter till en rät linje. Modellen kan skrivas $b = x_1 + x_2 t$, där x_1 och x_2 är parametrar och (t_k, b_k) uppmätta värden. Residualvektorn blir

$$r = \begin{bmatrix} x_1 + x_2 t_1 - b_1 \\ x_1 + x_2 t_2 - b_2 \\ \vdots \\ x_1 + x_2 t_m - b_m \end{bmatrix} = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

Dvs. $r = Ax - b$. Vi vill lösa minimeringsproblemet

$$\min_x ||Ax - b||$$

i någon lämplig norm.

Linjära problem

Observera att detta normalt inte är ett linjärt ekvationssystem. Det går normalt inte att i lösa $Ax = b$ pga. avvikelser i ekvationerna. Slarvigt kan vi skriva $Ax \approx b$.
Om vi hade kunnat lösa $Ax = b$ så hade residualvektorn varit $r = Ax - b = 0$ och mätpunkterna hade följt modellen exakt.
Notera även att matrisen A har fler rader än kolonner.
När residualvektorn kan skrivas $r = Ax - b$ säger vi att problemet är linjärt. Modellen har då utseendet:

$$b = \text{ uttryck}_1 \text{ parameter}_1 + \ldots + \text{uttryck}_n \text{ parameter}_n,$$

där uttryck_k beror av mätvärdena och inte på någon parameter.
Vår första modell är ickelinjär eftersom parametern λ inte ingår linjärt i modellen. I vissa fall kan vi via substitutioner eller andra transformationer skapa en linjär modell utifrån en ickelinjär sådan.

Linjära problem

Vår första modell är enkel att transformera, förutsatt att b och c har samma tecken. Låt oss anta att både b och c är positiva. Vi får

$$b = ce^{\lambda t} \Leftrightarrow \log b = \log c + \lambda t$$

λ ingår nu linjärt i modellen.
Om vi antar att c inte är känd (vi mätte aldrig b för $t = 0$) så är c en parameter som ingår ickelinjärt i modellen. Sätt $x_1 = \log c$ och vi får en linjär modell som är identisk med modellen för vår räta linje: $\log b = x_1 + \lambda t$. För att göra analogin tydligare sätter vi också $x_2 = \lambda$. Då får vi

$$\min_x \left\| \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} \log b_1 \\ \log b_2 \\ \vdots \\ \log b_m \end{bmatrix} \right\|$$

När x är beräknad sätter vi $c = e^{x_1}$ och $\lambda = x_2$.

När vi gör transformationer på detta sätt ändrar vi (ibland) på normen. Logaritmering, t.ex. har en utjämnande effekt och minskar de stora residualernas inflytande.

Detta kan jämföras med att minimera i en annan norm. Vi ställer en annan fråga, men den kan vara lika relevant.

Ibland fäster vi olika stor vikt vid de olika residualerna. Mätapparaturen kanske mäter olika noga i olika mätområden. Det är då rimligt att ett osäkert värde får mindre inflytande än ett säkert. Vi kan åstadkomma detta med en viktad norm, t.ex.

$$\min_x \|V(Ax - b)\|, \; V = \text{diag}(v_1, v_2, ..., v_m)$$

Residual r_k multipliceras alltså med vikten v_k.

Optimalitet

Vi studerar nu det linjära minstakvadratproblemet:

$$\min_x \|Ax - b\|_2$$

Det är enkelt att beskriva den optimala lösningen till detta problem. Vi ser på specialfallet när A har två kolonner, a_1 och a_2; kan enkelt generaliseras till ett godtyckligt fall. För en godtycklig $x \in \mathbb{R}^2$ gäller att $Ax = a_1x_1 + a_2x_2$ är en linjärkombination av As kolonner. När x varierar över alla vektorer med två element så kommer mängden $a_1x_1 + a_2x_2$ att bilda ett plan, As bildrum, $\mathcal{R}(A)$.

Om b tillhör detta plan så existerar (minst) ett x så att $Ax = b$ med likhet. Residualvektorn blir då noll. T.ex.

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}, \; b = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} \Rightarrow x = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Optimalitet

Normalt bildar dock b en vinkel mot planet, tag t.ex. $b = [2, 1, 2]^T$. Vektorn b kan då inte skrivas som en linjärkombination av As kolonner, men vi vill minimera avvikelsen, längden av residualvektorn $Ax - b$.

Deta upp b i två komponenter, b_A som ligger i planet, och $b_\perp$ som är ortogonalt mot planet. Oavsätt hur vi väljer x så kan vi inte nollställa någon del av $b_\perp$, eftersom $b_\perp$ är ortogonal mot alla linjärkombinationer, Ax. Däremot kan vi nollställa b_A, eftersom b_A ligger i planet och därmed är en linjärkombination av As kolonner, dvs. det finns (minst) ett x så att $b_A = Ax$. Det är detta x vi söker.
Residualvektorn blir $r = Ax - b = Ax - b_A - b_\perp = -b_\perp$.

Optimalitet

Här följer samma resonemang med normer:

Sats (Pythagoras)

Om y och z är ortogonala vektorer gäller:

$$\|y + z\|_2^2 = \|y\|_2^2 + \|z\|_2^2$$

ty

$$\|y+z\|_2^2 = (y+z)^T(y+z) = y^Ty + \underbrace{y^Tz}_{=0} + \underbrace{z^Ty}_{=0} + z^Tz = \|y\|_2^2 + \|z\|_2^2$$

Optimalitet

Det x som löser $Ax = b_A$ är optimalt. Ty om så inte vore fallet existerar $z \neq 0$ så att $x + z$ ger ett mindre värde på normen. Testa:

$$\|A(x+z) - b\|_2^2 = \|A(x+z) - b_A - b_\perp\|_2^2 =$$
$$\|\underbrace{Ax - b_A}_{=0} + Az - b_\perp\|_2^2 = \|Az\|_2^2 + \|b_\perp\|_2^2 \geq \|b_\perp\|_2^2$$

Med minimum då $z = 0$ (om A har linjärt oberoende kolonner).

Residualvektorn $r = -b_\perp$ är ju ortogonal mot bildrummet. Bildrummet utgörs av alla linjärkombinationer av a_1 och a_2 (i vårt specialfall) vilket medför att $a_1^T r = a_2^T r = 0$. Vi kan skriva dessa likheter på följande form:

$$0 = \begin{bmatrix} a_1^T r \\ a_2^T r \end{bmatrix}, = \begin{bmatrix} a_1^T \\ a_2^T \end{bmatrix} r = \begin{bmatrix} a_1^T & a_2^T \end{bmatrix}^T r = A^T r = A^T(Ax - b)$$

vilket ger oss <u>normalekvationerna:</u>

$$A^T A x = A^T b$$

Normal Equations

Our goal is to minimize the residual $\|r(x)\|_2^2 = ||Ax - b||_2^2$. To find minimum of this functional and derive the *normal equations*, we look for the x where the gradient of $||Ax - b||_2^2 = (Ax - b)^T(Ax - b)$ vanishes, or where $(r^T(x)r(x))' = 0$. So we want

$$\begin{aligned} 0 &= \lim_{e \to 0} \frac{r^T(x+e)r(x+e) - r^T(x)r(x)}{||e||_2} \\ &= \lim_{e \to 0} \frac{(A(x+e) - b)^T(A(x+e) - b) - (Ax - b)^T(Ax - b)}{||e||_2} \\ &= \lim_{e \to 0} \frac{2e^T(A^TAx - A^Tb) + e^TA^TAe}{||e||_2} \end{aligned}$$

The second term $\frac{|e^TA^TAe|}{||e||_2} \leq \frac{||A||_2^2||e||_2^2}{||e||_2} = ||A||_2^2||e||_2^2$ approaches 0 as e goes to 0, so the factor $A^TAx - A^Tb$ in the first term must also be zero, or $A^TAx = A^Tb$. This is a system of n linear equations in n unknowns, the normal equations.

Entydighet

rang$(A) = n \Rightarrow A^T A$ symmetrisk och positivt definit. Kan lösa normalekvationerna med Choleskyfaktorisering.

Entydighet?

- Om A har linjärt obereoende kolonner så har minstakvadratproblemet en entydlig lösning. Matrisen har full rang.
- Om A har linjärt beroende kolonner (är rangdefekt) så finns det oändligt många lösningar som ger samma residualvektor, ty tag $z \in \mathcal{N}(A)$. Då gäller $A(x + z) = Ax$.

Om A har nästan linjärt beroende kolonner, så är problemet illa konditionerat. Normalekvationerna förvärrar konditionen på problemet, ett elakt problem kan bli omöjligt att lösa. Det gäller att $\kappa(A^T A) = \kappa(A)^2$.

Entydighet

Det finns bättre metoder baserade på så kallad QR-faktorisering.
”x = A \ b” i Matlab använder QR-faktorisering.
Observera att operatorn \ är överlagrad. Om A är kvadratisk används LU-faktorisering, annars QR-faktorisering. Matlabkoderna för de olika fallen har ingen gemensam del.