Institutionen för
Matematiska Vetenskaper
Göteborg

TENTAMEN I
LINJÄR ALGEBRA OCH NUMERISK ANALYS F1, TMA671
2011-05-27

DAG: Fredag 27 maj 2011 TID: 14.00 - 18.00 SAL: V

Ansvarig:	Ivar Gustafsson, tel: 0705-335450
Förfrågningar:	Ivar Gustafsson
Lösningar:	Anslås vid sal MVF21
Resultat:	Tentan beräknas vara rättad senast 17 juni, resultat tillsänds dig.
Betygsgränser:	30, 42, 54 av maximalt 60 poäng
	Bonus (högst 10 poäng) från inlämningsuppgifter får tillgodoräknas.
Hjälpmedel:	Inga

Iakttag följande:
- Skriv tydligt och disponera papperet på lämpligt sätt
- Börja varje ny uppgift på nytt blad
- Alla svar era skall väl motiveras

LYCKA TILL!

Uppgift 1.
a) Betrakta ett egenvärdesproblem $Ax = \lambda x,\ A \in R^{n\times n},\ \lambda \in C,\ x \in C^n$. Visa att om A är symmetrisk så är egenvärdena reella. **(4p)**
b) Visa allmänt för $A \in R^{n\times n}$, ej nödvändigtvis symmetrisk, att om egenvärdena är reella så kan egenvektorerna väljas reella. **(2p)**
c) Betrakta ett *generaliserat egenvärdesproblem* $Ax = \lambda Bx$. Anta att A och B är symmetriska och positivt definita $n \times n$-matriser. Använd en Cholesky-faktorisering $B = LL^T$ för att visa att problemet kan skrivas som ett vanligt symmetriskt egenvärdesproblem $\hat{A}y = \lambda y$ med transformerade $\hat{A}$ och y. **(3p)**
d) Anta att potensmetoden används för att lösa problemet i c)-uppgiften och att Cholesky-faktoriseringen redan är gjord. Hur mycket mer arbete per iteration behövs för det generaliserade egenvärdesproblemet jämfört med det vanliga problemet i a)-uppgiften. Vi antar att n är stort och att alla element i matriserna är icke-nollor. **(3p)**

Uppgift 2.
a) Bestäm en ortogonalbas för $P_2[0,1]$ (mängden av polynom av grad ≤ 2) med avseende på skalärprodukten $< f, g >= \int_0^1 f(t)g(t)\ dt$. **(3p)**
b) Bestäm bästa approximationen i $P_2[0,1]$ av $f = t^3$ med avseende på skalärprodukten $< f, g >= \int_0^1 f(t)g(t)\ dt$ **(3p)**

Uppgift 3.
a) Visa att lösningen till minsta-kvadrat-problemet $min_x\|Ax-b\|_2,\ A\in R^{m\times n},\ m>n$, kan skrivas $\hat{x}=V_1\Sigma_r^{-1}U_1^Tb$, där $A=U_1\Sigma_rV_1^T$ är *kompakt SVD-faktorisering* av A. **(3p)**
b) Om A inte har full rang så är lösningen i a)-uppgiften inte unik. Visa att lösningen enligt a)-uppgiften är den lösning som har minsta norm, dvs som minimerar $\|x\|_2$. **(4p)**

Uppgift 4.
a) Lös följande system av differentialekvationer med *diagonaliseringsmetoden*

$$\begin{cases} x_1'(t)=-4x_1(t)+x_2(t) & x_1(0)=-1 \\ x_2'(t)=-3x_2(t) & x_2(0)=1 \\ x_3'(t)=x_1(t)+x_2(t)-2x_3(t) & x_3(0)=1 \end{cases}.\ \textbf{(4p)}$$

b) Ange för vilka steglängder h som följande tre metoder är stabila för problemet i a)-uppgiften: *Euler framåt, Euler bakåt* och *trapetsmetoden.* **(2p)**
c) Välj steglängd $h=1$ och räkna fram en approximation till lösningen $y(1)$ till problemet i a)-uppgiften med *Eulers bakåtmetod.* **(3p)**

Uppgift 5.
a) Definiera vad som menas med ett *flyttalssystem.* **(2p)**
b) Ge exempel på ett flyttalssystem med basen 10 och bestäm UFL (”under flow limit”) och OFL (”over flow limit”) för systemet. **(2p)**
c) Vad menas med *gradual underflow*? Ge ett par exempel på tal som representerar gradual underflow i ditt system i b)-uppgiften. **(2p)**

Uppgift 6.
a) Gör en iteration med *Newtons metod* med start i origo på problemet:

$$\begin{cases} x_1^2+2x_2^2-x_3=0 \\ x_1+x_2^3=0 \\ x_1+x_2+x_3=1 \end{cases}.\ \textbf{(3p)}$$

b) Härled *Gauss-Newtons metod* för att lösa *icke-linjära minsta-kvadrat-problem.* **(3p)**

Uppgift 7.
a) Bestäm en *kvadratisk spline* s, med nod i $x=1$, som interpolerar $f(x)=x^4$ i punkterna $x=0,\ x=1,\ x=2$ och som uppfyller randvillkoret $s'(2)=f'(2)$. **(4p).**
b) Bestäm en approximation till $\int_0^2 s(x)\,dx$ med *trapetsformeln*, där s är splinen i a)-uppgiften. Välj steglängd $h=1$. **(2p)**

Uppgift 8.
a) Definiera *linjesökningsproblemet* i samband med en sökmetod för att lösa optimeringsproblem i flera variabler utan bivillkor. **(2p)**
b) Ange två metoder för linjesökning som inte kräver att derivator beräknas. **(2p)**
c) Betrakta problemet att minimera $f(x_1,x_2)=x_1^3+x_2^2-x_1+x_2$. Lös problemet approximativt med en iteration i *Steepest Descent-metoden*, med start i $x_0=[1/2,-1/2]^T$ och med linjesökning med *sekantmetoden.* Gör en iteration utgående från $\alpha_0=0$ och $\alpha_1=1$ i sekantmetoden. **(4p)**

Institutionen för
Matematik
Göteborg

F1 - Linjär algebra och numerisk analys, TMA671

Lösningar till tentamen 27 maj 2011

1a) $Ax = \lambda x,\ x \in C^n, \lambda \in C$. Multiplicera från vänster med $\bar{x}^T$ och vi får
(1) $\bar{x}^T Ax = \lambda \bar{x}^T x$
Här är vänsterledet en skalär: $\alpha = \bar{x}^T Ax = (\bar{x}^T Ax)^T = x^T A^T \bar{x} = x^T \bar{A}\bar{x} = \bar{\bar{x}}^T \bar{A}\bar{x} = \bar{\alpha} \Rightarrow$ $\alpha = \bar{x}^T Ax$ reell. Vidare är i (1) $\bar{x}^T x$ en reell skalär och därmed är λ reellt ty det är kvoten mellan två reella tal.
1b) $Av = \lambda v,\ v \in C^n, \lambda \in R$. Låt $v = x + iy$ för reella x och y. Då gäller $A(x + iy) = \lambda(x + iy)$ och identifikation av real och imaginärdelar ger $Ax = \lambda x$ och $Ay = \lambda y$ och eftersom inte både x och y kan vara nollvektorn så duger minst en av dem som egenvektor.
1c) Problemet blir $Ax = \lambda LL^T x \Leftrightarrow AL^{-T}L^T x = \lambda LL^T x \Leftrightarrow L^{-1}AL^{-T}y = \lambda y \Leftrightarrow \hat{A}y = \lambda y$, där $\hat{A} = L^{-1}AL^{-T}$ och $y = L^T x$. $\hat{A}$ är symmetrisk: $\hat{A}^T = (L^{-T})^T A^T (L^{-1})^T = L^{-1}A^T L^{-T} = \hat{A}$ eftersom $A^T = A$.
1d) Väsentliga arbetet är matris $\times$ vektor, Ax kostar $2n^2$ flops medan $L^{-1}AL^{-T}y$ dessutom innebär lösning av två triangulära system, vilket kostar n^2 flops vardera. Totalt blir det alltså ungefär dubbelt så dyrt per iteration.

2a) Vi ortogonaliserar standardbasen $\{1, t, t^2\}$ med hjälp av Gram-Schmidts process:
$b_1 = 1,\ b_2 = t - \frac{\langle t,1\rangle}{\langle 1,1\rangle}1 = t - \frac{1}{2},\ b_3 = t^2 - \frac{\langle t^2,1\rangle}{\langle 1,1\rangle} - \frac{\langle t^2, t-\frac{1}{2}\rangle}{\langle t-\frac{1}{2}, t-\frac{1}{2}\rangle}(t - \frac{1}{2}) = t^2 - \frac{1}{3} - (t - \frac{1}{2}) = t^2 - t + \frac{1}{6}$.
2b) Lösningen är $\hat{f} = c_1 b_1 + c_2 b_2 + c_3 b_3$, där $\{b_i\}_{i=1}^3$ är basen i a)-uppgiften.
$c_1 = \frac{\langle f,1\rangle}{\langle 1,1\rangle} = \frac{1}{4},\ c_2 = \frac{\langle f, t-\frac{1}{2}\rangle}{\langle t-\frac{1}{2}, t-\frac{1}{2}\rangle} = \frac{\frac{3}{40}}{\frac{1}{12}} = \frac{9}{10},\ c_3 = \frac{\langle f, t^2-t+\frac{1}{6}\rangle}{\langle t^2-t+\frac{1}{6}, t^2-t+\frac{1}{6}\rangle} = \frac{\frac{1}{120}}{\frac{1}{180}} = \frac{3}{2}$, vilket ger $\hat{f} = \frac{1}{4} + \frac{9}{10}(t - \frac{1}{2}) + \frac{3}{2}(t^2 - t + \frac{1}{6})$.

3a) Vi ska alltså hitta bästa approximation till b i $Col(A)$, som vi betecknar $\hat{b} = A\hat{x}$. Från SVD-faktoriseringen har vi att $Col(A) = Col(U_1)$, eftersom Σ_r och V_1 är "reguljära". Vidare är kolonnerna i U_1 ortonormala, så de utgör en ON-bas för $Col(A)$. Med $\hat{x} = V_1\Sigma_r^{-1}U_1^T b$ får vi, eftersom kolonnerna i V_1 är ortonormala, $A\hat{x} = U_1\Sigma_r V_1^T \hat{x} = U_1 U_1^T b$, dvs $\hat{b} = A\hat{x}$ är bästa approximation till b i $Col(A)$, enligt projektionsformeln med en ON-bas.
3b) En full SVD-faktorisering av A är $A = U\Sigma V^T$, där $V = [V_1\ V_2]$. Eftersom V är ortogonal så följer från kompakta SVD:n att $AV_2 = U_1\Sigma_r V_1^T V_2 = 0$ dvs kolonnerna i V_2 är bas för $Nul(A)$. En allmän lösning är $\hat{x} + h$ där $h \in Nul(A)$, dvs $h = V_2 y_2$ för någon vektor y_2. Nu följer för $y_1 = \Sigma_r^{-1}U_1^T b$ och $y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$ att $\|\hat{x} + h\|_2^2 = \|V_1 y_1 + V_2 y_2\|_2^2 = \|Vy\|_2^2 = \|y\|_2^2 = \|y_1\|_2^2 + \|y_2\|_2^2$, som minimeras för $y_2 = 0$ dvs för $h = 0$. Alltså har $\hat{x}$ minsta möjliga norm.

4a) Problemet är på matrisform $x' = Ax$, där $A = \begin{bmatrix} -4 & 1 & 0 \\ 0 & -3 & 0 \\ 1 & 1 & -2 \end{bmatrix}$. Egenvärden och egenvektorer beräknas. Egenvärdena är $\lambda_1 = -4,\ \lambda_2 = -3,\ \lambda_3 = -2$.
Motsvarande egenvektorer fås genom lösning av resp. homogent ekvationssystem $(A - \lambda_i I)v_i = 0$ och blir: $v_1 = (-2, 0, 1)^T,\ v_2 = (-1, -1, 2)^T$ och $v_3 = (0, 0, 1)^T$.
Lösningsformeln är sedan

$$x = c_1 e^{\lambda_1 t} v_1 + c_2 e^{\lambda_2 t} v_2 + c_3 e^{\lambda_3 t} v_3 = c_1 e^{-4t} \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix} + c_2 e^{-3t} \begin{bmatrix} -1 \\ -1 \\ 2 \end{bmatrix} + c_3 e^{-2t} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Koefficienterna $c_1,\ c_2$ och c_3 bestäms från begynnelsevillkoren genom ekvationssystemet
$\begin{bmatrix} -2 & -1 & 0 \\ 0 & -1 & 0 \\ 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$, med lösning $c_1 = 1,\ c_2 = -1,\ c_3 = 2$.

Lösningen blir alltså $x = e^{-4t} \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix} - e^{-3t} \begin{bmatrix} -1 \\ -1 \\ 2 \end{bmatrix} + 2e^{-2t} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$.

4b) Eulers framåtmetod är stabil om $\lambda h \in S, \forall \lambda$ där S är metodens stabilitetsområde $\{z \in C; |1 + z| \leq 1\}$. Detta ger villkoret $-4h \geq -2$, dvs $h \leq \frac{1}{2}$. Euler bakåt och trapetsmetoden är A-stabila dvs stabila för alla val av h.

4c) Eulers bakåtmetod ser ut så här: $y_{k+1} = y_k + hAy_{k+1}$ som kan skrivas $(I - hA)y_{k+1} = y_k$.
y_1 fås alltså från lösningen av ett ekvationssystem med matris $I - hA = \begin{bmatrix} 5 & -1 & 0 \\ 0 & 4 & 0 \\ -1 & -1 & 3 \end{bmatrix}$

och högerled $y_0 = [-1, 1,\ 1]^T$ med lösning $y_1 = \frac{1}{60} \begin{bmatrix} -9 \\ 15 \\ 22 \end{bmatrix}$.

5a) Kvadruppeln $(\beta,\ p,\ L,\ U)$ definierar tal på formen $x = m \cdot \beta^e$, där
$m = \pm d_0.d_1 d_2 ... d_{p-1},\ 0 \leq d_i < \beta,\ 0 < d_0 < \beta,\ 1 \leq |m| < \beta,\ L \leq e \leq U$.

5b) Ta exempelvis $(10,\ 5,\ -9,\ 9)$. "underlow"-gräns är minsta positiva tal i ett flyttalssystem $= \beta^L = 10^{-9}$, "overflow"-gräns är största positiva tal $= \beta^{U+1}(1 - \beta^{-p}) = 10^{10}(1 - 10^{-5})$.

5c) "Gradual underflow" representeras av alla tal mellan 0 och "underflow"-gräns dvs mellan 0 och 10^{-9}. Exempel på sådana tal är $0.9 \cdot 10^{-9}$ och 10^{-10}.

6a) Vi ska lösa $f(x) = 0$ där $f = \begin{cases} x_1^2 + 2x_2^2 - x_3 \\ x_1 + x_2^3 \\ x_1 + x_2 + x_3 - 1 \end{cases}$.

Jacobianen blir $J = \begin{bmatrix} 2x_1 & 4x_2 & -1 \\ 1 & 3x_2^2 & 0 \\ 1 & 1 & 1 \end{bmatrix}$

Vi får $\mathbf{x}_0 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$, $f_0 = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}$, $J_0 = \begin{bmatrix} 0 & 0 & -1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}$, $\mathbf{x}_1 = \mathbf{x}_0 - J_0 \backslash f_0$

Ekvationssystem $J_0 s_0 = f_0 \Leftrightarrow \begin{bmatrix} 0 & 0 & -1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} s_0 = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix} \Leftrightarrow s_0 = \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}$

$\mathbf{x}_1 = \mathbf{x}_0 - s_0 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$.

6b) Problemet är: $min_x \|f(x)\|_2$, där $f : R^n \to R^m$ med $m > n$.
Linjärisering med Taylor kring en approximation x_k: $f(x) \approx f(x_k) + J(x_k)(x - x_k)$.
En linjär modell av vårt problem är alltså: $min_x \|f(x_k) + J(x_k)(x - x_k)\|_2$.
Detta är ett linjärt minsta kvadrat-problem för obekant x, lösningen blir vår nya approximation x_{k+1}. En iteration med metoden kan alltså formuleras:
$x_{k+1} = x_k + d_k$, där d_k är lösning till överbestämda linjära ekvationssystemet $J(x_k)d_k = -f(x_k)$, som alltså ska lösas i minsta-kvadratmening.

7a) Splinen har två delar: s_1, $0 \le x \le 1$ och s_2, $1 \le x \le 2$. Vi börjar med att ansätta $s_2 = 16 + a(x-2) + b(x-2)^2$, eftersom vi har flest villkor i $x = 2$. Vi har vid ansatsen utnyttjat att $s(2) = f(2) = 16$. Derivering ger $s_2' = a + 2b(x-2)$ och villkoret $s'(2) = f'(2) = 32$ ger $a = 32$. Vidare ger interpolationsvillkoret $s_2(1) = f(1) = 1$ att $b = 17$. Vi har alltså $s_2 = 16 + 32(x-2) + 17(x-2)^2$ med $s_2' = 32 + 34(x-2)$ och $s_2'(1) = -2$.
För s_1 gör vi nu ansatsen $s_1 = 1 + c(x-1) + d(x-1)^2$, så att villkoret $s_1(1) = 1$ är uppfyllt genom ansatsen. Vi har $s_1' = c + 2d(x-1)$ och splinevillkoret $s_1'(1) = s_2'(1) = -2$ ger $c = -2$. Slutligen ger interpolationsvillkoret $s_1(0) = f(0) = 0$ att $d = -3$, dvs $s_1 = 1 - 2(x-1) - 3(x-1)^2$.
7b) Trapetsformeln med $h = 1$ ger approximationen (vi behöver inte göra a)-uppgiften):
$\int_0^2 s(x)dx \approx T = 1[\frac{1}{2}f(0) + f(1) + \frac{1}{2}f(2)] = 9$.

8a) Sökmetod: $x_{k+1} = x_k + \alpha_k d_k$, där α_k fås som lösning till linjesökningsproblemet: $min_{\alpha\in R}\ f(x_k + \alpha d_k)$.
8b) Gyllene snittet och polynomapproximation.
8c) $\nabla f = \begin{bmatrix} 3x_1^2 - 1 \\ 2x_2 + 1 \end{bmatrix}$, $H = \begin{bmatrix} 6x_1 & 0 \\ 0 & 2 \end{bmatrix}$, $x_0 = \begin{bmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{bmatrix}$, $\nabla f_0 = \begin{bmatrix} -\frac{1}{4} \\ 0 \end{bmatrix}$, $d_0 = \begin{bmatrix} \frac{1}{4} \\ 0 \end{bmatrix}$.
Linjesökning: $min_{\alpha\in R} f(x_k + \alpha d_k)$, med sekantmetoden och $\alpha_0 = 0$, $\alpha_1 = 1$:
$\alpha_2 = \alpha_1 - \frac{\nabla f(x_0+\alpha_1 d_0)^T d_0(\alpha_1-\alpha_0)}{\nabla f(x_0+\alpha_1 d_0)^T d_0 - \nabla f(x_0+\alpha_0 d_0)^T d_0} = 1 - \frac{\frac{11}{64}(1)}{\frac{11}{64}+\frac{4}{64}} = 1 - \frac{11}{15} = \frac{4}{15}$. Iterationen blir:
$x_1 = x_0 + \frac{4}{15}d_0 = \begin{bmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{bmatrix} + \frac{4}{15}\begin{bmatrix} \frac{1}{4} \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{17}{30} \\ -\frac{1}{2} \end{bmatrix}$.