1.2 Lösningar till kapitel 2

Låt $\mathbf{e}$ beteckna vektorn av ettor (`ones(n, 1)`) och låt $\mathbf{e}_k$ vara kolonn k i enhetsmatrisen.

1: Det gäller att $\mathbf{Ae} = \mathbf{0}$ varför $\mathbf{A}$ har ett icketrivialt nollrum.

2: Eftersom radsummorna alla är lika, σ, gäller $\mathbf{M}_n\mathbf{e} = \sigma\mathbf{e}$. Summan av de n radsummorna måste vara $n\sigma = 1 + 2 + \cdots + n^2 = n^2(n^2+1)/2$ varför $\sigma = n(n^2+1)/2$.

3: a) $\mathbf{A}[1, -1, 1]^T = \mathbf{0}$. b)$\mathbf{Ae} = [2, 4, 6]^T$, så det existerar ∞ många lösningar.

4: När man någon enstaka gång vill beräkna en invers utförs det normalt med hjälp av LU-faktorisering. Låt oss beteckna inversen med $\mathbf{X}$. Det gäller då att $\mathbf{AX} = \mathbf{I}$. Om vi skriver denna ekvation kolonnvis får vi $\mathbf{Ax}_k = \mathbf{e}_k, k = 1, \ldots, n$. Här är $\mathbf{x}_k$ kolonn k i $\mathbf{X}$ och $\mathbf{e}_k$ är kolonn k i $\mathbf{I}$. Vi har alltså n linjära ekvationssystem att lösa. I detta exempel underlättas lösandet av att $\mathbf{A}$ redan är triangulär. Vi kan alltså, i detta speciella fall, beräkna inversen med tre framåtsubstitutioner. Ännu enklare blir det om vi observerar att $x_{j,k} = 0$ då $j < k$ (pga högerledets, $\mathbf{e}_k$, speciella form).

Ett annat sätt att lösa problemet är via ansats. En triangulär matris har en triangulär invers (om den existerar) med $(\mathbf{A}^{-1})_{k,k} = 1/a_{k,k}$. Vi kan därför göra följande ansats:

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 \\ \alpha & -1 & 0 \\ \beta & \gamma & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 1 & -1 & 0 \\ 1 & -2 & 1 \end{bmatrix}$$

vilket ger systemet

$$\begin{cases} \alpha - 1 = 0 \\ \beta + \gamma + 1 = 0 \\ -\gamma - 2 = 0 \end{cases}$$

så att $\alpha = \beta = 1$ och $\gamma = -2$.

5: Uppgiften kan lösas på flera olika sätt, t.ex. $0 = \det(\mathbf{A}^2) = (\det(\mathbf{A}))^2 \Rightarrow \det(\mathbf{A}) = 0$, så att $\mathbf{A}$ är singulär. **Om** $\mathbf{A}$ är ickesingulär så gäller att $\mathbf{A}^2 = \mathbf{0} \Rightarrow \mathbf{A}^{-1}\mathbf{A}^2 = \mathbf{A}^{-1}\mathbf{0}$ så $\mathbf{A} = \mathbf{0}$, motsägelse.

6: a) För att få lite enklare beteckningar visar jag istället att $(\mathbf{A}^T\mathbf{B})^T = \mathbf{B}^T\mathbf{A}$. Det är ekvivalent med det som efterfrågas i uppgiften om man låter $\mathbf{C} = \mathbf{A}^T$. Låt oss partitionera matriserna kolonnvis, dvs. $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_n]$ och $\mathbf{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_n]$. Vi får

$$\mathbf{A}^T\mathbf{B} = \begin{bmatrix} \mathbf{a}_1^T \\ \mathbf{a}_2^T \\ \vdots \\ \mathbf{a}_n^T \end{bmatrix} [\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n] = \begin{bmatrix} \mathbf{a}_1^T\mathbf{b}_1 & \mathbf{a}_1^T\mathbf{b}_2 & \cdots & \mathbf{a}_1^T\mathbf{b}_n \\ \mathbf{a}_2^T\mathbf{b}_1 & \mathbf{a}_2^T\mathbf{b}_2 & \cdots & \mathbf{a}_2^T\mathbf{b}_n \\ \vdots & \vdots & & \vdots \\ \mathbf{a}_n^T\mathbf{b}_1 & \mathbf{a}_n^T\mathbf{b}_2 & \cdots & \mathbf{a}_n^T\mathbf{b}_n \end{bmatrix}$$

Transponatet av ovanstående matris är:

$$\begin{bmatrix} \mathbf{a}_1^T\mathbf{b}_1 & \mathbf{a}_2^T\mathbf{b}_1 & \cdots & \mathbf{a}_n^T\mathbf{b}_1 \\ \mathbf{a}_1^T\mathbf{b}_2 & \mathbf{a}_2^T\mathbf{b}_2 & \cdots & \mathbf{a}_n^T\mathbf{b}_2 \\ \vdots & \vdots & & \vdots \\ \mathbf{a}_1^T\mathbf{b}_n & \mathbf{a}_2^T\mathbf{b}_n & \cdots & \mathbf{a}_n^T\mathbf{b}_n \end{bmatrix} = \begin{bmatrix} \mathbf{b}_1^T\mathbf{a}_1 & \mathbf{b}_1^T\mathbf{a}_2 & \cdots & \mathbf{b}_1^T\mathbf{a}_n \\ \mathbf{b}_2^T\mathbf{a}_1 & \mathbf{b}_2^T\mathbf{a}_2 & \cdots & \mathbf{b}_2^T\mathbf{a}_n \\ \vdots & \vdots & & \vdots \\ \mathbf{b}_n^T\mathbf{a}_1 & \mathbf{b}_n^T\mathbf{a}_2 & \cdots & \mathbf{b}_2^T\mathbf{a}_n \end{bmatrix}$$

vilket är lika med $\mathbf{B}^T\mathbf{A}$. Likheten ovan följer av att $\mathbf{x}^T\mathbf{y} = \sum_k x_k y_k = \sum_k y_k x_k = \mathbf{y}^T\mathbf{x}$.
b) Det enklaste sättet att kontrollera om en matris är en invers till en given matris, är att multiplicera ihop

matriserna och se om man får enhetsmatrisen. Så, $(\mathbf{AB})(\mathbf{B}^{-1}\mathbf{A}^{-1}) = \mathbf{A}(\mathbf{BB}^{-1})\mathbf{A}^{-1} = \mathbf{AIA}^{-1} = \mathbf{AA}^{-1} = \mathbf{I}$.

7: Vi börjar med att visa att inversen (om den existerar) är entydigt bestämd. Antag att $\mathbf{C}$ är en ickesingulär matris och att $\mathbf{CX} = \mathbf{I}$ och att $\mathbf{CY} = \mathbf{I}$. Detta medför att $\mathbf{C}(\mathbf{X}-\mathbf{Y}) = \mathbf{0}$. Men eftersom $\mathbf{C}$ är ickesingulär så måste (varför?) $\mathbf{X}-\mathbf{Y} = \mathbf{0}$. Nu till $\mathbf{A}^T$. Inversen till denna matris är $(\mathbf{A}^T)^{-1}$ och det gäller att $\mathbf{A}^T(\mathbf{A}^T)^{-1} = \mathbf{I}$. Men det gäller även att $\mathbf{I} = \mathbf{A}^{-1}\mathbf{A} = (\mathbf{A}^{-1}\mathbf{A})^T = \mathbf{A}^T(\mathbf{A}^{-1})^T$, där den andra likheten följer av att $\mathbf{I}$ är symmetrisk och den tredje likheten följer av föregående övning. Inversens entydighet implicerar nu att $(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T$.

8: Systemet är ekvivalent med

$$\begin{aligned} \mathbf{L}_1\mathbf{x} &= \mathbf{b} \\ \mathbf{Bx} + \mathbf{L}_2\mathbf{y} &= \mathbf{c} \end{aligned}$$

Algoritm: lös $\mathbf{L}_1\mathbf{x} = \mathbf{b}$, bilda sedan $\mathbf{t} = \mathbf{c} - \mathbf{Bx}$ och lös slutligen $\mathbf{L}_2\mathbf{y} = \mathbf{t}$.

9: Determinanten av en triangulär matris är produkten av diagonalelementen. Om denna produkt är skild från noll är matrisen ickesingulär. 1) Matrisen $\mathbf{L}_k$ är triangulär och har ettor på diagonalen.
2) Via inspektion, alternativt: $\mathbf{Ae}_j$ or den j-te kolonnen i $\mathbf{A}$, så

$$\mathbf{L}_k\mathbf{e}_j = (\mathbf{I} - \mathbf{me}_k^T)\mathbf{e}_j = \mathbf{e}_j - \mathbf{me}_k^T\mathbf{e}_j = \begin{cases} \mathbf{e}_j, j \neq k \\ \mathbf{e}_k - \mathbf{m}, j = k \end{cases}$$

$$3) \quad (\mathbf{I} + \mathbf{me}_k^T)(\mathbf{I} - \mathbf{me}_k^T) = \mathbf{I} - \mathbf{me}_k^T + \mathbf{me}_k^T - \mathbf{m}\underbrace{\mathbf{e}_k^T\mathbf{m}}_{0}\mathbf{e}_k^T = \mathbf{I}$$

$$4) \quad \mathbf{L}_k\mathbf{L}_j = (\mathbf{I} - \mathbf{m}_k\mathbf{e}_k^T)(\mathbf{I} - \mathbf{m}_j\mathbf{e}_j^T) = \mathbf{I} - \mathbf{m}_j\mathbf{e}_j^T - \mathbf{m}_k\mathbf{e}_k^T + \mathbf{m}_k \underbrace{\mathbf{e}_k^T\mathbf{m}_j}_{0,\text{ ty } k<j} \mathbf{e}_j^T$$

10:

$$\mathbf{A} = \begin{bmatrix} 1 & a \\ c & b \end{bmatrix}, \quad \mathbf{L}_1 = \begin{bmatrix} 1 & 0 \\ -c & 1 \end{bmatrix}, \quad \mathbf{L}_1\mathbf{A} = \begin{bmatrix} 1 & a \\ 0 & b-ca \end{bmatrix}$$

så

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ c & 1 \end{bmatrix}\begin{bmatrix} 1 & a \\ 0 & b-ca \end{bmatrix}$$

singulär om $b = ca$ (ty $\det\mathbf{A} = \det(\mathbf{LU}) = \det\mathbf{L}\det\mathbf{U} = 1\cdot(b-ca)$).

11: Definition: $\mathbf{x}^T\mathbf{Ax} > 0$ för alla $\mathbf{x} \neq \mathbf{0}$.
a) Tag $\mathbf{x} = \mathbf{e}_k$, $\mathbf{e}_k^T\mathbf{Ae}_k = a_{k,k}$.
b) Med $\sigma = -\text{sign}(a_{j,k})$ och $\mathbf{x} = \mathbf{e}_j + \sigma\mathbf{e}_k$

$$\mathbf{x}^T\mathbf{Ax} = a_{j,j} + 2\sigma a_{j,k} + a_{k,k} > 0 \Rightarrow \frac{a_{j,j} + a_{k,k}}{2} > |a_{j,k}|$$

$$\text{c)} \quad |a_{j,k}| < \frac{a_{j,j} + a_{k,k}}{2} \leq \max(a_{j,j}, a_{k,k})$$

För att visa d) använder vi definitionen av positivt definit, $\mathbf{x}^T\mathbf{Ax} > 0$ för alla $\mathbf{x} \neq \mathbf{0}$, och väljer en speciell $\mathbf{x}$-vektor. Om $\mathbf{A}$ är positivt definit så kan man visa att $\mathbf{LDL}^T$-faktoriseringen alltid existerar (inget pivot-element kan bli noll). $\mathbf{L}$ är därför ickesingulär och vi kan ta $\mathbf{x} = \mathbf{L}^{-T}\mathbf{e}_k$ (där $\mathbf{e}_k$ är kolonn k i $\mathbf{I}$). $\mathbf{x}$ kan inte vara noll (varför?) och vi får:

$$0 < \mathbf{x}^T\mathbf{Ax} = [\mathbf{L}^{-T}\mathbf{e}_k]^T\mathbf{LDL}^T[\mathbf{L}^{-T}\mathbf{e}_k] = \mathbf{e}_k^T\mathbf{De}_k = d_{k,k}$$

12: Jag har skrivit ut hela sekvensen av $\mathbf{L}_k$-matriser. Detta gör man normalt inte då man handräknar, utan man skriver upp sekvensen av reducerade $\mathbf{A}$-matriser.

$$\mathbf{A} = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 5 \end{bmatrix}, \quad \mathbf{L}_1 = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{L}_1\mathbf{A} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & -1 & 5 \end{bmatrix}, \quad \mathbf{L}_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

$$\mathbf{U} = \mathbf{L}_2\mathbf{L}_1\mathbf{A} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 4 \end{bmatrix}, \text{ och } \mathbf{A} = \mathbf{LU} \text{ med } \mathbf{L} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}$$

Choleskyfaktoriseringen får från $\mathbf{U}$ genom att bryta ut "roten ur diagonalen":

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 2 \end{bmatrix}\begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 2 \end{bmatrix}$$

13: Gör ansatsen

$$\begin{bmatrix} l_1 & 0 \\ l_2 & l_3 \end{bmatrix}\begin{bmatrix} u_1 & u_2 \\ 0 & u_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \Rightarrow \begin{cases} l_1u_1 = 0 \\ l_1u_2 = 1 \\ l_2u_1 = 1 \\ l_2u_2 + l_3u_3 = 0 \end{cases}$$

$l_1u_1 = 0 \Rightarrow l_1 = 0$ eller $u_1 = 0$, men då kan inte $l_1u_2 = 1$ och $l_2u_1 = 1$.

14: Vi måste antaga att $\alpha \neq 0$ för att ta första steget i Gauesseliminationen:

$$\begin{bmatrix} \alpha & 1 \\ 1 & 2 \end{bmatrix} \rightarrow \begin{bmatrix} \alpha & 1 \\ 0 & 2-1/\alpha \end{bmatrix} \Rightarrow \mathbf{A} = \mathbf{CC}^T = \begin{bmatrix} \sqrt{\alpha} & 0 \\ 1/\sqrt{\alpha} & \sqrt{2-1/\alpha} \end{bmatrix}\begin{bmatrix} \sqrt{\alpha} & 1/\sqrt{\alpha} \\ 0 & \sqrt{2-1/\alpha} \end{bmatrix}$$

För att kunna dra roten ur diagonalen får vi tydligen kräva att $\alpha > 0$ och att $2 - 1/\alpha > 0$. Sammanfattningsvis måste alltså $\alpha > 1/2$. Nu med definitionen, $\mathbf{x}^T\mathbf{Ax} > 0, \mathbf{x} \neq \mathbf{0}$. Vi använder kvadratkomplettering:

$$\mathbf{x}^T\mathbf{Ax} = \alpha x_1^2 + 2x_1x_2 + 2x_2^2 = (\sqrt{\alpha}x_1 + x_2/\sqrt{\alpha})^2 + x_2^2(2 - 1/\alpha)$$

Vi får alltså kräva att $\alpha > 0$ och $2 - 1/\alpha > 0$ (vi kan ju göra den första kvadraten noll genom att ta $x_1 = 1/\sqrt{\alpha}$ och $x_2 = -\sqrt{\alpha}$ t.ex.).
En symmetrisk matris har reella egenvärden och är matrisen positivt definit så är egenvärdena positiva. Karakteristiska ekvationen blir: $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$, dvs. $\lambda^2 - (2+\alpha)\lambda + 2\alpha - 1 = 0$. Rötterna är

$$1 + \frac{\alpha}{2} \pm \sqrt{\left[1 + \frac{\alpha}{2}\right]^2 + 1 - 2\alpha} = 1 + \frac{\alpha}{2} \pm \frac{1}{2}\sqrt{(\alpha-2)^2 + 4}$$

Rooten är alltså alltid reell och vi kräver att

$$1 + \frac{\alpha}{2} > \frac{1}{2}\sqrt{(\alpha-2)^2+4}$$

vilket medför att $\alpha > 1/2$.

15: $a_{k,k} = \mathbf{C}(k,:) * \mathbf{C}(k,:)^T$ (med användning av Matlabnotation), så $a_{k,k} = c_{k,1}^2 + \ldots + c_{k,k}^2$. Det gäller alltså att $|c_{k,j}| \leq \sqrt{a_{k,k}}$.

16: I dessa typer av uppgifter skall man vara uppmärksam på operationer som skapar mycket mer minne än vad som man utgår ifrån, t.ex.

- inverser av glesa matriser tenderar att bli rätt fulla
- i en ytterprodukt går man från två vektorer till en matris

Man skall också tänka på beräkningstiden (antalet flyttalsoperationer):

- att använda inversen för $\mathbf{Ax} = \mathbf{b}$ tar mer tid än LU-faktorisering
- man vill hålla nere dimensionerna, så att man utför hellre en matris-vektor-multiplikation än en matris-matris-multiplikation

Här några typiska operationer och det ungefärliga antalet flyttalsoperationer.
Antag att $\mathbf{x}$, $\mathbf{y}$ är kolonnvektorer med n element och att $\mathbf{A}$ och $\mathbf{B}$ är $n \times n$-matriser.

- $\mathbf{x}^T\mathbf{y}$, en innerprodukt kräver ungefär n multiplikationer och additioner
- $\mathbf{Ax}$, en matris-vektor-multiplikation kräver ungefär n^2 multiplikationer och additioner
- $\mathbf{AB}$, en matris-matris-multiplikation kräver ungefär n^3 multiplikationer och additioner

Så, om vi skall bilda

$$\mathbf{y} = \mathbf{ABx}$$

skall vi arbetar från höger till vänster, vilket kräver två matris-vektor-multiplikationer. Utför vi operationerna från vänster till höger får vi en matris-matris-multiplikation och en matris-vektor-multiplikation. Dessutom måste vi allokera utrymme för en temporär matris.

Nu till problemet: Man skall **inte** bilda inverser explicit! Lös system istället; $\mathbf{C}^{-1}\mathbf{b}$ är ju t.ex. lösningen till systemet: $\mathbf{Cy} = \mathbf{b}$. Här följer algoritmen:
Lös $\mathbf{Cy} = \mathbf{b}$ och bilda $\mathbf{z} = \mathbf{Ab}$.
Bilda $\mathbf{y} = \mathbf{y} + \mathbf{z}$ (vi slösar givetvis inte med minne heller). = står för tilldelning här.
Bilda $\mathbf{z} = 2(\mathbf{Ay})$ (varför är parenteserna viktiga?)
Bilda $\mathbf{y} = \mathbf{y} + \mathbf{z}$ och lös slutligen $\mathbf{Bx} = \mathbf{y}$. Eftersom vi skriver över $\mathbf{x}$ på slutet skulle vi kunnat använda $\mathbf{x}$ som en arbetssvektor (istället för $\mathbf{z}$).

17: a) $\mathbf{uv}^T = [\mathbf{u}v_1, \ldots, \mathbf{u}v_n]$ så alla kolonnerna är parallella med $\mathbf{u}$, alltså är rangen högst ett. Eftersom $\mathbf{u}, \mathbf{v} \neq \mathbf{0}$ måste $\mathbf{uv}^T \neq \mathbf{0}$ varför rangen är ett.
b) rang $\mathbf{A} = 1 \Rightarrow \mathcal{R}(\mathbf{A}) = \text{span}(\mathbf{u})$ för $\mathbf{u} \neq \mathbf{0}$. Alltså gäller att $\mathbf{A} = [\mathbf{u}v_1, \ldots, \mathbf{u}v_n] = \mathbf{uv}^T$ med något $v_k \neq 0$.

18: a) Om matrisen är singulär så existerar $\mathbf{x} \neq \mathbf{0}$ så att $(\mathbf{I} - \mathbf{uv}^T)\mathbf{x} = \mathbf{0}$, dvs. så att $\mathbf{x} = \mathbf{u}(\mathbf{v}^T\mathbf{x})$, dvs. $\mathbf{x}$ måste vara parallell med $\mathbf{u}$. Tag $\mathbf{x} = \mathbf{u}$ (längden spelar ingen roll eftersom vi har nollvektorn i högerledet). Detta ger $\mathbf{u} = \mathbf{u}(\mathbf{v}^T\mathbf{u})$ dvs. $\mathbf{v}^T\mathbf{u} = 1$. Om $\mathbf{v}^T\mathbf{u} \neq 1$ så är matrisen ickesingulär.

$$b) \quad (\mathbf{I} - \mathbf{uv}^T)(\mathbf{I} - \sigma\mathbf{uv}^T) = \mathbf{I} - \sigma\mathbf{uv}^T - \mathbf{uv}^T + \mathbf{uv}^T\sigma\mathbf{uv}^T = \mathbf{I} - \mathbf{uv}^T(1 + \sigma - \sigma\mathbf{v}^T\mathbf{u})$$

Detta är enhetsmatrisen om $\sigma = 1/(\mathbf{v}^T\mathbf{u} - 1)$ och $\mathbf{v}^T\mathbf{u} \neq 1$.

19: Sätt $\mathbf{C} = (\mathbf{I} - \mathbf{V}^T\mathbf{A}^{-1}\mathbf{U})^{-1}$ Vi får:

$$(\mathbf{A} - \mathbf{UV}^T)(\mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{UCV}^T\mathbf{A}^{-1}) = \mathbf{I} + \mathbf{UCV}^T\mathbf{A}^{-1} - \mathbf{UV}^T\mathbf{A}^{-1} - \mathbf{UV}^T\mathbf{A}^{-1}\mathbf{UCV}^T\mathbf{A}^{-1} =$$

$$\mathbf{I} + \mathbf{U}\left[\mathbf{C} - \mathbf{I} - \mathbf{V}^T\mathbf{A}^{-1}\mathbf{UC}\right]\mathbf{V}^T\mathbf{A}^{-1} = \mathbf{I} + \mathbf{U}[(\mathbf{I} - \mathbf{V}^T\mathbf{A}^{-1}\mathbf{U})\mathbf{C} - \mathbf{I}]\mathbf{V}^T\mathbf{A}^{-1} = \mathbf{I}$$

20: $p = 1$: 1) $||\mathbf{x}||_1 = \sum_{k=1}^n |x_k| > 0$ om $\mathbf{x} \neq \mathbf{0}$.
2) $||\gamma\mathbf{x}||_1 = \sum_{k=1}^n |\gamma x_k| = |\gamma| \sum_{k=1}^n |x_k| = |\gamma| ||\mathbf{x}||_1$.
3) $||\mathbf{x} + \mathbf{y}||_1 = \sum_{k=1}^n |x_k + y_k| \leq \sum_{k=1}^n (|x_k| + |y_k|) = \sum_{k=1}^n |x_k| + \sum_{k=1}^n |y_k| = ||\mathbf{x}||_1 + ||\mathbf{y}||_1$

$p = 2$: 1) och 2) enkla, visar tredje punkten.
$||\mathbf{x}+\mathbf{y}||_2^2 = (\mathbf{x}+\mathbf{y})^T(\mathbf{x}+\mathbf{y}) = \mathbf{x}^T\mathbf{x} + 2\mathbf{x}^T\mathbf{y} + \mathbf{y}^T\mathbf{y} \le ||\mathbf{x}||_2^2 + 2||\mathbf{x}||_2||\mathbf{y}||_2 + ||\mathbf{y}||_2^2 = (||\mathbf{x}||_2 + ||\mathbf{y}||_2)^2$, så $||\mathbf{x}+\mathbf{y}||_2 \le ||\mathbf{x}||_2 + ||\mathbf{y}||_2$.

$p = \infty$: 1) och 2) enkla, visar tredje punkten.
$||\mathbf{x}+\mathbf{y}||_\infty = \max_k |x_k + y_k| \le \max_k(|x_k| + |y_k|) \le \max_k |x_k| + \max_k |y_k| = ||\mathbf{x}||_\infty + ||\mathbf{y}||_\infty$.

21: $p = 1$: 1) $||\mathbf{A}||_1 = \max_j \sum_i |a_{i,j}|$, så om $\mathbf{A} \ne \mathbf{0}$ så finns något $a_{i,j} \ne 0$ varför $||\mathbf{A}||_1 > 0$.
2) $||\gamma\mathbf{A}||_1 = \max_j \sum_i |\gamma a_{i,j}| = \max_j \sum_i (|\gamma||a_{i,j}|) = |\gamma| \max_j \sum_i |a_{i,j}| = |\gamma|||\mathbf{A}||_1$.
3) $||\mathbf{A}+\mathbf{B}||_1 = \max_j \sum_i |a_{i,j} + b_{i,j}| \le \max_j \sum_i (|a_{i,j}| + |b_{i,j}|) \le \max_j \sum_i |a_{i,j}| + \max_j \sum_i |b_{i,j}| = ||\mathbf{A}||_1 + ||\mathbf{B}||_1$.
Nu till submultiplikativiteten. Vi visar $||\mathbf{A}\mathbf{x}||_1 \le ||\mathbf{A}||_1||\mathbf{x}||_1$ först. Det enklaste är att använda definitionen av normen:

$$||\mathbf{A}||_1 = \max_{\mathbf{x}\ne\mathbf{0}} \frac{||\mathbf{A}\mathbf{x}||_1}{||\mathbf{x}||_1}$$

Max antas för en speciell vektor. För andra vektorer gäller därför att $||\mathbf{A}||_1 \ge ||\mathbf{A}\mathbf{x}||_1/||\mathbf{x}||_1$.
Och nu till $||\mathbf{AB}||_1$. Antag att max antages för kolonn k i $\mathbf{B}$;

$$||\mathbf{AB}||_1 = ||\mathbf{A}\mathbf{b}_k||_1 \le ||\mathbf{A}||_1||\mathbf{b}_k||_1 \le ||\mathbf{A}||_1||\mathbf{B}||_1$$

där den första olikheten följer av 5).
$p = \infty$ kan visas analogt. Ett trick man kan använda är att $||\mathbf{A}^T||_1 = ||\mathbf{A}||_\infty$, så att vissa egenskaper hos maxnormen kan härledas från motsvarande egenskaper hos ettnormen.

22: Låt $\mathbf{A} = \mathbf{C}\mathbf{C}^T$ vara Choleskyfaktoriseringen av $\mathbf{A}$. Då är

$$||\mathbf{x}||_\mathbf{A} = (\mathbf{x}^T\mathbf{A}\mathbf{x})^{1/2} = (\mathbf{x}^T\mathbf{C}\mathbf{C}^T\mathbf{x})^{1/2} = ((\mathbf{C}^T\mathbf{x})^T(\mathbf{C}^T\mathbf{x}))^{1/2} = ||\mathbf{C}^T\mathbf{x}||_2$$

Vi kan alltså återföra $||\mathbf{x}||_\mathbf{A}$ på tvånormen. Eftersom $\mathbf{A}$ är positivt definit och därmed ickesingulär så är även $\mathbf{C}$ ickesingulär varför $\mathbf{C}^T\mathbf{x} = \mathbf{0} \Leftrightarrow \mathbf{x} = \mathbf{0}$. Vi testar nu de tre normvillkoren:
1) $||\mathbf{x}||_\mathbf{A} > 0$ om $\mathbf{x} \ne \mathbf{0}$ ty $||\mathbf{C}^T\mathbf{x}||_2 > 0$ om $\mathbf{x} \ne \mathbf{0}$ ty $||\ ||_2$ är en norm och $\mathbf{C}^T\mathbf{x} \ne \mathbf{0}$.
2) $||\alpha\mathbf{x}||_\mathbf{A} = ||\alpha\mathbf{C}^T\mathbf{x}||_2 = |\alpha|\ ||\mathbf{C}^T\mathbf{x}||_2 = |\alpha|\ ||\mathbf{x}||_\mathbf{A}$.
3) $||\mathbf{x}+\mathbf{y}||_\mathbf{A} = ||\mathbf{C}^T(\mathbf{x}+\mathbf{y})||_2 \le ||\mathbf{C}^T\mathbf{x}||_2 + ||\mathbf{C}^T\mathbf{y}||_2 = ||\mathbf{x}||_\mathbf{A} + ||\mathbf{y}||_\mathbf{A}$.

23: a) 1) $||\mathbf{A}||_{\max} = \max_{j,k} |a_{j,k}| > 0$ om något $a_{j,k} \ne 0$.
2) $||\mathbf{A}||_{\max} = \max_{j,k} |\gamma a_{j,k}| = \max_{j,k} |\gamma||a_{j,k}| = |\gamma| \max_{j,k} |a_{j,k}| = |\gamma|||\mathbf{A}||_{\max}$.
3) $||\mathbf{A}+\mathbf{B}||_{\max} = \max_{j,k} |a_{j,k} + b_{j,k}| \le \max_{j,k}(|a_{j,k}| + |b_{j,k}|) \le \max_{j,k} |a_{j,k}| + \max_{j,k} |b_{j,k}| = ||\mathbf{A}||_{\max} + ||\mathbf{B}||_{\max}$.
Notera (står inte i uppgiften) att denna norm **inte** är submultiplikativ. Tag $\mathbf{A}$ =`ones(2)`. Då är $||\mathbf{AA}||_{\max} = 2$, men $||\mathbf{A}||_{\max}||\mathbf{A}||_{\max} = 1 \cdot 1 = 1$.

b) Låt vec($\mathbf{A}$) vara den vektor man får om man staplar alla $\mathbf{A}$:s kolonner på varandra (om $\mathbf{A}$ är en $m \times n$-matris är vec($\mathbf{A}$) en vektor om mn element). Vi ser att $||\mathbf{A}||_F = ||\mathrm{vec}(\mathbf{A})||_2$.
1) $||\mathbf{A}||_F = ||\mathrm{vec}(\mathbf{A})||_2 > 0$ om något $a_{j,k} \ne 0$.
2) $||\gamma\mathbf{A}||_F = ||\gamma\mathrm{vec}(\mathbf{A})||_2 = |\gamma|||\mathrm{vec}(\mathbf{A})||_2 = |\gamma|||\mathbf{A}||_F$.
3) $||\mathbf{A}+\mathbf{B}||_F = ||\mathrm{vec}(\mathbf{A}+\mathbf{B})||_2 = ||\mathrm{vec}(\mathbf{A}) + \mathrm{vec}(\mathbf{B})||_2 \le ||\mathrm{vec}(\mathbf{A})||_2 + ||\mathrm{vec}(\mathbf{B})||_2 = ||\mathbf{A}||_F + ||\mathbf{B}||_F$.

24: $\mathbf{D}^{-1} = \mathrm{diag}(1/d_1, \ldots, 1/d_n)$. För en diagonalmatris gäller att $||\mathbf{D}|| = \max |d_k|$ (för de tre normer vi använder). Så $\kappa(\mathbf{D}) = \max |d_k|\ \max |1/d_k| = \max |d_k|/\min |d_k|$.

25: Om $\alpha = 1$ så är matrisen singulär och vi säger att $\kappa(\mathbf{A}) = \infty$. I annat fall gäller att

$$\kappa(\mathbf{A}) = \left|\left|\begin{bmatrix} 1 & \alpha \\ 1 & 1 \end{bmatrix}\right|\right|_1 \left|\left|\frac{1}{1-\alpha}\begin{bmatrix} 1 & -\alpha \\ -1 & 1 \end{bmatrix}\right|\right|_1 = \max(1+|\alpha|, 2)\ \max(1+|\alpha|, 2)/|1-\alpha|$$

Med andra ord, $\kappa(\mathbf{A}) = 4/|1-\alpha|$ om $|\alpha| < 1$ och $(1+|\alpha|)^2/|1-\alpha|$ annars.

26: a) **Om** är $\mathbf{A}$ singulär så existerar $\mathbf{x} \neq \mathbf{0}$ så att $\mathbf{Ax} = \mathbf{0}$. Det medför att $\mathbf{x}^T\mathbf{Ax} = 0$ (och inte > 0), motsägelse!
b) Vi kräver inte att $\mathbf{A}$ är symmetrisk utan vet bara att $\mathbf{x}^T\mathbf{Ax} > 0$ om $\mathbf{x} \neq \mathbf{0}$. Tag $\mathbf{x} = \mathbf{A}^{-1}\mathbf{y}$ (notera att $\mathbf{x} = \mathbf{0} \Leftrightarrow \mathbf{y} = \mathbf{0}$). Vi får $0 < \mathbf{x}^T\mathbf{Ax} = (\mathbf{A}^{-1}\mathbf{y})^T\mathbf{A}(\mathbf{A}^{-1}\mathbf{y}) = \mathbf{y}^T\mathbf{A}^{-T}\mathbf{y}$ som är en skalär så att $(\mathbf{y}^T\mathbf{A}^{-T}\mathbf{y})^T = \mathbf{y}^T\mathbf{A}^{-T}\mathbf{y}$. Men $(\mathbf{y}^T\mathbf{A}^{-T}\mathbf{y})^T = \mathbf{y}^T\mathbf{A}^{-1}\mathbf{y}$. Alltså är $0 < \mathbf{y}^T\mathbf{A}^{-1}\mathbf{y}$.

27: Symmetrisk ty $(\mathbf{BB}^T)^T = (\mathbf{B}^T)^T\mathbf{B}^T = \mathbf{BB}^T$.
Positivt definit ty $\mathbf{x}^T\mathbf{Ax} = \mathbf{x}^T\mathbf{BB}^T\mathbf{x} = (\mathbf{B}^T\mathbf{x})^T\mathbf{B}^T\mathbf{x} = ||\mathbf{B}^T\mathbf{x}||_2^2 > 0$ om $\mathbf{x} \neq \mathbf{0}$ ($\mathbf{x} \neq \mathbf{0}$ kan ej medföra att $\mathbf{B}^T\mathbf{x} = \mathbf{0}$ eftersom $\mathbf{B}$ är ickesingulär).

28: $\mathbf{x}^T\mathbf{Bx} > 0$ om $\mathbf{x} \neq \mathbf{0}$. Tag $\mathbf{x} = \mathbf{e}_1 \neq \mathbf{0}$, det ger $0 < \mathbf{e}_1^T\mathbf{Be}_1 = \mathbf{e}_1^T[\alpha,\ \mathbf{a}^T]^T = \alpha$.
Tag $\mathbf{x}^T = [0,\ \mathbf{y}^T]$ med godtycklig $\mathbf{y} \neq \mathbf{0}$. detta medför att

$$0 < \mathbf{x}^T\mathbf{Bx} = \begin{bmatrix} 0 & \mathbf{y}^T \end{bmatrix} \underbrace{\begin{bmatrix} \alpha & \mathbf{a}^T \\ \mathbf{a} & \mathbf{A} \end{bmatrix} \begin{bmatrix} 0 \\ \mathbf{y} \end{bmatrix}}_{\begin{bmatrix} \mathbf{a}^T\mathbf{y} \\ \mathbf{Ay} \end{bmatrix}} = \mathbf{y}^T\mathbf{Ay}$$

b) Gör ansatsen ($\mathbf{L}$ är en matris, $\mathbf{l}$ en vektor och λ är en skalär):

$$\begin{bmatrix} \alpha & \mathbf{a}^T \\ \mathbf{a} & \mathbf{A} \end{bmatrix} = \begin{bmatrix} \lambda & \mathbf{0} \\ \mathbf{l} & \mathbf{L} \end{bmatrix} \begin{bmatrix} \lambda & \mathbf{l}^T \\ \mathbf{0} & \mathbf{L}^T \end{bmatrix} = \begin{bmatrix} \lambda^2 & \lambda\mathbf{l}^T \\ \mathbf{l}\lambda & \mathbf{l}\,\mathbf{l}^T + \mathbf{LL}^T \end{bmatrix}$$

så $\lambda = \sqrt{\alpha}$, $\mathbf{l} = \mathbf{a}/\lambda = \mathbf{a}/\sqrt{\alpha}$ och $\mathbf{LL}^T = \mathbf{A} - \mathbf{l}\,\mathbf{l}^T = \mathbf{A} - \mathbf{aa}^T/\alpha$.

29: Studera föregående övning. Om $\mathbf{a}$ har k element gäller (i det steget) att $\mathbf{l} = \mathbf{a}/\sqrt{\alpha}$ kräver k multiplikationer och en division $\mathbf{l} = \mathbf{a}(1/\sqrt{\alpha})$. $\mathbf{A} - \mathbf{l}\,\mathbf{l}^T$ kräver $k(k+1)/2$ additioner och multiplikationer (utnyttja symmetrin), så totalt

$$\sum_{k=1}^{n-1} \frac{k(k+1)}{2} = \frac{1}{2}\left[\sum_{k=1}^{n-1} k + \sum_{k=1}^{n-1} k^2\right] = \frac{1}{2}\left[\frac{(n-1)n}{2} + \frac{(n-1)n(2(n-1)+1)}{6}\right] = \frac{n^3}{6} + \cdots$$

De k multiplikationerna bidrar endast med en n^2-term.

30: Här är en del av mitt svar i typsatt form.
No, it need not be the case. Consider:

$$\mathbf{S} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix},\ \mathbf{D} = \begin{bmatrix} 1 & 0 \\ 0 & 0.01 \end{bmatrix}.\ \text{Then}\ \mathbf{DS} = \begin{bmatrix} 2 & 1 \\ 0.01 & 0.02 \end{bmatrix}$$

Take $\mathbf{x}^T = [1 - 10]$ then $\mathbf{x}^T\mathbf{DSx} = -6.1$.
Looking at the slightly more general 2×2-case we get.

$$\mathbf{S} = \begin{bmatrix} a & b \\ b & c \end{bmatrix},\ \mathbf{D} = \begin{bmatrix} 1 & 0 \\ 0 & d \end{bmatrix}$$

with b not zero.
If $\mathbf{B}$ is positive definite then so is $\mathbf{B} + \mathbf{B}^T$, since

$$\mathbf{x}^T\mathbf{Bx} = \frac{\mathbf{x}^T\mathbf{Bx} + (\mathbf{x}^T\mathbf{Bx})^T}{2} = \mathbf{x}^T\frac{\mathbf{B} + \mathbf{B}^T}{2}\mathbf{x}$$

Numerisk analys

In our case:

$$\mathbf{DS} + (\mathbf{DS})^T = \begin{bmatrix} 2a & b(1+d) \\ b(1+d) & 2dc \end{bmatrix}$$

which is positive definite provided the determinant is positive, so if $4adc > b^2(1+d)^2$ or $4ac/b^2 > (1+d)^2/d$ which cannot hold when $d \to 0$. So, in my numerical example d must be smaller than the smallest root to the equation $d^2 - 14d + 1 = 0$, i.e. $d < 0.0717967697...$. When d is equal to this root, $\mathbf{DS} + (\mathbf{DS})^T$ is singular.
Looking at the 2×2-case, one can say that $\mathbf{DS}$ is positive definite if $\mathbf{D}$ is not too "badly scaled".

Med mera matematik kan man visa mer och då även i det allmänna $n \times n$-fallet. I den efterföljande brevväxlingen stod det klart att frågeställaren inte riktigt visste vad positivt definit var i det osymmetriska fallet. Han hade blandat ihop positivt definit med positiva egenvärden (han ville egentligen ha ett bevis för att $\mathbf{DS}$ har positiva egenvärden). Om en matris är symmetrisk så är den positivt definit om dess egenvärden är positiva. Det är dock inte alltid sant om matrisen är osymmetrisk. Jag skickade honom följande exempel:

$$\begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & -10 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = -8$$

31:

$$(\mathbf{A} + i\mathbf{B})(\mathbf{x} + i\mathbf{y}) = \mathbf{b} + i\mathbf{c} \Leftrightarrow \mathbf{Ax} - \mathbf{By} + i(\mathbf{Ay} + \mathbf{Bx}) = \mathbf{b} + i\mathbf{c} \Leftrightarrow \begin{bmatrix} \mathbf{A} & -\mathbf{B} \\ \mathbf{B} & \mathbf{A} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{b} \\ \mathbf{c} \end{bmatrix}$$

vilket kräver dubbelt så mycket plats. Angående operationerna: en komplex multiplikation, $(\alpha + i\beta)(\gamma + i\delta) = \alpha\gamma - \beta\delta + i(\alpha\delta + \beta\gamma)$, tar sex operationer och en komplex addition, $(\alpha + i\beta) + (\gamma + i\delta) = (\alpha + \gamma) + i(\beta + \delta)$, tar två operationer. Ett komplext system av ordningen n tar därmed ungefär $(6+2)n^3/3 = 8n^3/3$ operationer. Ett reellt system av ordningen $2n$ tar ungefär $2/3(2n)^3 = 16n^3/3$ operationer, dvs. det dubbla antalet. I det reella systemet kommer operationerna i (*, +)-par. Detta gäller även det komplexa, eftersom en komplex * kräver fyra reella * och två reella +. En komplex + kräver två reella +, så att den komplexa lösningen kräver dubbelt så många (*, +)-par.

32: a) Att lösa systemet med GE kräver, för eliminationssteget, $(n-1)$ +, * och / om vi inte behöver pivotera och det krävs inget extra utrymme (om vi får förstöra matrisen) dvs. vi klarar oss på de $3n-2$ elementen. Vi pivotering kan den den övre bandbredden öka med ett, som i bilden nedan, där vi pivoterar de två första stegen:

```
x x 0 0 0 0    x x x 0 0 0    x x x 0 0 0
x x x 0 0 0    x x 0 0 0 0    0 x x x 0 0
0 x x x 0 0    0 x x x 0 0    0 x x 0 0 0
  etc.
```

Om vi pivoterar i **varje** steg så krävs det $n-1$ / och + samt $2(n-1)$ *. Att lösa det triangulära systemet kostar (om vi inte pivoterar) ungefär $2n$ * och + samt n /. Sammanfattningsvis: att lösa ett tridiagonalt problem kräver $c \cdot n$ operationer, där c är ett litet tal.
Om vi beräknar inversen måste vi kunna lagra en $n \times n$-matris (dvs. n^2 element, se nästa övning). Att beräkna inversen görs via LU-faktorisering (vi måste lösa n system) vilket kräver $\mathcal{O}(n^2)$ operationer. Att slutligen bilda $\mathbf{A}^{-1}\mathbf{b}$, givet inversen, kostar bara det ungefär n^2 + och *. Att använda inversen kräver alltså $\mathcal{O}(n^2)$ operationer och element. Gausselimination (LU-faktorisering) kräver $\mathcal{O}(n)$ operationer och element.

b) Man kan göra en väldigt enkel algoritm när matrisen, $\mathbf{T}$, är både positivt definit och tridiagonal. Låt oss beräkna Choleskyfaktoriseringen $\mathbf{T} = \mathbf{C}\mathbf{C}^T$. Här följer ett exempel där $n = 5$:

$$\underbrace{\begin{bmatrix} \alpha_1 & \beta_1 & 0 & 0 & 0 \\ \beta_1 & \alpha_2 & \beta_2 & 0 & 0 \\ 0 & \beta_2 & \alpha_3 & \beta_3 & 0 \\ 0 & 0 & \beta_3 & \alpha_4 & \beta_4 \\ 0 & 0 & 0 & \beta_4 & \alpha_5 \end{bmatrix}}_{\mathbf{T}} = \underbrace{\begin{bmatrix} d_1 & 0 & 0 & 0 & 0 \\ s_1 & d_2 & 0 & 0 & 0 \\ 0 & s_2 & d_3 & 0 & 0 \\ 0 & 0 & s_3 & d_4 & 0 \\ 0 & 0 & 0 & s_4 & d_5 \end{bmatrix}}_{\mathbf{C}} \underbrace{\begin{bmatrix} d_1 & s_1 & 0 & 0 & 0 \\ 0 & d_2 & s_2 & 0 & 0 \\ 0 & 0 & d_3 & s_3 & 0 \\ 0 & 0 & 0 & d_4 & s_4 \\ 0 & 0 & 0 & 0 & d_5 \end{bmatrix}}_{\mathbf{C}^T}$$

Om vi identifierar element ser vi att $\alpha_1 = d_1^2$, $\beta_1 = d_1 s_1$, $\alpha_2 = s_1^2 + d_2^2$, $\beta_2 = d_2 s_2$, $\alpha_3 = s_2^2 + d_3^2$, $\beta_3 = d_3 s_3$ osv. Om vi löser dessa ekvationer med avseende på d_k och s_k (och i rätt ordning får vi) kan vi skriva ner följande Matlab-algoritm:

```
d(1) = sqrt(alpha(1));
for k = 1:n-1
  s(k)     = beta(k) / d(k);
  d(k + 1) = sqrt(alpha(k + 1) - s(k) * s(k));
end
```

När väl har använt α- och β-värdena kan de skrivas över av motsvarande d och s värden. Så, om vi får förstöra $\mathbf{T}$ kan vi skriva programmet på följande vis:

```
alpha(1) = sqrt(alpha(1));
for k = 1:n-1
  beta(k)      = beta(k) / alpha(k);
  alpha(k + 1) = sqrt(alpha(k + 1) - beta(k) * beta(k));
end
```

Om någon kvadratrot inte existerar eller om en nämnare är noll, så är matrisen inte positivt definit. Om högerledet lagras i $\mathbf{b}$ löser följande rutiner de triangulära systemen och skriver över $\mathbf{b}$ med $\mathbf{x}$.

```
b(1) = b(1) / alpha(1);
for k = 2:n
  b(k) = (b(k) - beta(k - 1) * b(k - 1)) / alpha(k);
end

b(n) = b(n) / alpha(n);
for k = n-1:-1:1
  b(k) = (b(k) - beta(k) * b(k + 1)) / alpha(k);
end
```

Antalet operationer framgår klart av ovanstående rutiner. Notera att inget extra minne behövs.

33: a) Låt oss först bestämma oss för vad vi skall bevisa. En tridiagonal matris, $\mathbf{T}$, satisfierar $t_{j,k} = 0$ för alla $|j - k| > 1$. Det finns många tridiagonala matriser som inte har full invers; ett exempel är enhetsmatrisen (som dessutom är diagonal). Matrisen är irreducibel om de sub- och superdiagonala elementen $(t_{k+1,k}, t_{k,k+1})$ alla är skilda från noll. Även inversen av en irreducibel matris kan innehålla nollor, ett exempel ges av:

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 & -1 & 1 \\ 0 & 0 & 1 & -1 \\ -1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 1 \end{bmatrix}$$

Antalet ickenollor i en stor (ickesingulär) matris, av ovanstående konstruktion, är mindre än hälften. Sådana exempel är dock **otypiska**. Beräknar man inversen av en tridiagonal slumpmatris blir den (nästan) alltid full.

Elementen i inversen är komplicerade funktioner (kvoter av determinanter) av elementen i $\mathbf{T}$. $(1,2)$-elementet i inversen av en tridiagonal matris av ordning fyra är t.ex.:

$$\frac{t_{1,2}(t_{4,3}t_{3,4} - t_{3,3}t_{4,4})}{\det \mathbf{T}}$$

Vi ser att elementet i inversen är noll om $t_{1,2} = 0$. Det är också noll om $t_{4,3}t_{3,4} - t_{3,3}t_{4,4} = 0$, men det är otypiskt i den meningen att vi måste ha bestämda värden (som ger exakt kancellation). Här följer en stringent formulering av det jag har valt att bevisa:

Sats: Låt $\mathbf{T}$ vara en godtycklig tridiagonal matris. För varje $\epsilon > 0$ existerar det en ickesingulär tridiagonal matris $\tilde{\mathbf{T}}$ vars invers inte innehåller några nollor och där $|\mathbf{T} - \tilde{\mathbf{T}}| \le \epsilon$ (elementvis jämförelse).

Satsen säger att det finns "gott om" tridiagonala matriser med fulla inverser. Lägg märke till att det inte finns lika gott om tridiagonala matriser med glesa inverser. I till exempel 2×2-fallet innehåller inversen nollor bara om någon av α_1, γ_2, β_2 eller α_2 är noll, ty

$$\begin{bmatrix} \alpha_1 & \gamma_2 \\ \beta_2 & \alpha_2 \end{bmatrix}^{-1} = \frac{1}{\alpha_1\alpha_2 - \beta_2\gamma_2} \begin{bmatrix} \alpha_2 & -\gamma_2 \\ -\beta_2 & \alpha_1 \end{bmatrix}$$

Så om $\tilde{\mathbf{T}}^{-1}$ är full så finns det inte en $\mathbf{T}$, godtyckligt nära $\tilde{\mathbf{T}}$, där $\mathbf{T}^{-1}$ är gles.

Nu till beviset av satsen. Vi använder induktion med avseende på matrisens ordning, n. Betrakta först fallet då $n = 1$. $\mathbf{T} = (\alpha)$. Om $\alpha \neq 0$ så existerar inversen och den är full. Om $\alpha = 0$ så existerar inte inversen, men en godtyckligt liten störning ($\neq 0$) av $\mathbf{T}$ ger oss en inverterbar matris som är full. Låt nu $\mathbf{T}$ vara en ickesingulär tridiagonal matris av ordning n och antag att $\mathbf{T}^{-1}$ är full ($\mathbf{T}$ kan alltså vara en störd matris). Vi bildar nu en matris av ordning $n+1$ och ställer upp en ansats för inversen. Vi antar att α, β och γ nedan alla är skilda från noll (antingen från början eller via en störning).

$$\underbrace{\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix}}_{\text{enhetsmatris}} = \underbrace{\begin{bmatrix} \mathbf{T} & \gamma\mathbf{e}_n \\ \beta\mathbf{e}_n^T & \alpha \end{bmatrix}}_{\text{ordning } n+1} \underbrace{\begin{bmatrix} \mathbf{G} & \mathbf{c} \\ \mathbf{b}^T & \delta \end{bmatrix}}_{\text{invers}} = \begin{bmatrix} \mathbf{TG} + \gamma\mathbf{e}_n\mathbf{b}^T & \mathbf{Tc} + \delta\gamma\mathbf{e}_n \\ \beta\mathbf{e}_n^T\mathbf{G} + \alpha\mathbf{b}^T & \beta c_n + \alpha\delta \end{bmatrix}$$

Detta ger oss fyra ekvationer. $\mathbf{Tc} + \delta\gamma\mathbf{e}_n = \mathbf{0}$ medför att $\mathbf{c} = -\delta\gamma\mathbf{T}^{-1}\mathbf{e}_n$, vilken är full om $\delta \neq 0$. $\delta = 0$ är inte möjligt, ty det medför att $\mathbf{c} = \mathbf{0}$ vilket medför att $\beta c_n + \alpha\delta = 1$ inte kan vara sant. Alltså är $\delta \neq 0$ och $\mathbf{c}$ är full. Det kvarstår att visa att $\mathbf{G}$ och $\mathbf{b}$ är fulla. $\beta\mathbf{e}_n^T\mathbf{G} + \alpha\mathbf{b}^T = \mathbf{0}$ ger oss ett uttryck för $\mathbf{b}$ som vi kan använda i $\mathbf{TG} + \gamma\mathbf{e}_n\mathbf{b}^T = \mathbf{I}$. Detta ger

$$\left[\mathbf{T} - \frac{\beta\gamma}{\alpha}\mathbf{e}_n\mathbf{e}_n^T\right]\mathbf{G} = \mathbf{I}$$

Använder vi övning 17, i detta kapitel, ser vi att

$$\mathbf{G} = \mathbf{T}^{-1} + \mathbf{T}^{-1}\mathbf{e}_n\mathbf{e}_n^T\mathbf{T}^{-1}\frac{\beta\gamma}{\alpha(1 - \frac{\beta\gamma}{\alpha}\mathbf{e}_n^T\mathbf{T}^{-1}\mathbf{e}_n)}$$

$\mathbf{G}$ existerar om $1 - \frac{\beta\gamma}{\alpha}\mathbf{e}_n^T\mathbf{T}^{-1}\mathbf{e}_n \neq 0$. Genom att störa α, β och γ kan vi alltid se till att så är fallet. Vi kan också se till att $\mathbf{G}$ blir full (eftersom $\mathbf{T}^{-1}$ är det). Eftersom slutligen $\mathbf{b}^T = -(\beta/\alpha)\mathbf{e}_n^T\mathbf{G}$ är även $\mathbf{b}$ full.