LINEAR ALGEBRA

and Learning

from Data

First Edition

MANUAL FOR INSTRUCTORS

Gilbert Strang

Massachusetts Institute of Technology

math.mit.edu/weborder.php (orders)

math.mit.edu/~gs

www.wellesleycambridge.com
email: linearalgebrabook@gmail.com

Wellesley - Cambridge Press

Box 812060

Wellesley, Massachusetts 02482

Problem Set I.1, page 6

1 A combination of $\boldsymbol{u}, \boldsymbol{v}$, and $\boldsymbol{u}+\boldsymbol{v}$ (vectors in $\mathbf{R}^4$) produces

$$\boldsymbol{u}+\boldsymbol{v}-(\boldsymbol{u}+\boldsymbol{v})=\mathbf{0} \qquad \begin{bmatrix} \boldsymbol{u} & \boldsymbol{v} & \boldsymbol{u}+\boldsymbol{v} \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}=\mathbf{0} \qquad A\boldsymbol{x}=\mathbf{0}$$

A is 4 by 3, $\boldsymbol{x}$ is 3 by 1, $\mathbf{0}$ is 4 by 1. Your example could use numerical vectors.

2 Suppose $A\boldsymbol{x}=A\boldsymbol{y}$. Then if $z=c\,(x-y)$ for any number c, we have $Az=0$. One candidate is always the zero vector $z=0$ (from the choice $c=0$).

3 We are given vectors $\boldsymbol{a}_1$ to $\boldsymbol{a}_n$ in $\mathbf{R}^m$ with $c_1\boldsymbol{a}_1+\cdots+c_n\boldsymbol{a}_n=\mathbf{0}$.

(1) At the matrix level $A\boldsymbol{c}=\begin{bmatrix} \boldsymbol{a}_1 & \ldots & \boldsymbol{a}_n \end{bmatrix}\boldsymbol{c}=\mathbf{0}$, with the $\boldsymbol{a}$'s in the columns of A, and c's in the vector $\boldsymbol{c}$.

(2) At the scalar level this is $\displaystyle\sum_{j=1}^{n} a_{ij}c_j=0$ for each row $i=1,2,\ldots,m$ of A.

4 Two vectors x and y out of many solutions to $A\boldsymbol{x}=\mathbf{0}$ for $A=\textbf{ones}(3,3)$ are

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}=\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} 3 \\ -3 \\ 0 \end{bmatrix}=\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

These vectors $\boldsymbol{x}=(1,1,-2)$ and $\boldsymbol{y}=(3,-3,0)$ are independent. But there is no 3rd vector $\boldsymbol{z}$ with $A\boldsymbol{z}=\mathbf{0}$ and independent $\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}$. (If there were, then combinations of $\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}$ would say that every vector $\boldsymbol{w}$ solves $A\boldsymbol{w}=\mathbf{0}$, which is not true.)

5 (a) The vector $\boldsymbol{z}=(1,-1,1)$ is perpendicular to $\boldsymbol{v}=(1,1,0)$ and $\boldsymbol{w}=(0,1,1)$. Then $\boldsymbol{z}$ is perpendicular to all combinations of $\boldsymbol{v}$ and $\boldsymbol{w}$—a whole plane in $\mathbf{R}^3$.

(b) $\boldsymbol{u}=(1,1,1)$ is NOT a combination of $\boldsymbol{v}$ and $\boldsymbol{w}$. And $\boldsymbol{u}$ is NOT perpendicular to $z=(1,-1,1)$: Their dot product is $\boldsymbol{u}^{\mathrm{T}}\boldsymbol{z}=1$.

6 If $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ are corners of a parallelogram, then $\boldsymbol{z}$ = corner 4 can be $\boldsymbol{u}+\boldsymbol{v}-\boldsymbol{w}$ **or** $\boldsymbol{u}-\boldsymbol{v}+\boldsymbol{w}$ **or** $-\boldsymbol{u}+\boldsymbol{v}+\boldsymbol{w}$. Here those 4th corners are $\boldsymbol{z}=(4,0)$ **or** $\boldsymbol{z}=(-2,2)$ or $\boldsymbol{z}=(4,4)$.

Reasoning : The corners A, B, C, D around a parallelogram have $A+C=B+D$.

7 The column space of $A=\begin{bmatrix}\boldsymbol{v} & \boldsymbol{w} & \boldsymbol{v}+2\boldsymbol{w}\end{bmatrix}$ consists of all combinations of $\boldsymbol{v}$ and $\boldsymbol{w}$.

Case 1 $\boldsymbol{v}$ and $\boldsymbol{w}$ are independent. Then $\mathbf{C}(A)$ has dimension 2 (a *plane*). A has rank 2 and its nullspace is a line (dimension 1) in $\mathbf{R}^3$: Then $2+1=3$.

Case 2 $\boldsymbol{w}$ is a multiple $c\boldsymbol{v}$ (not both zero). Then $\mathbf{C}(A)$ is a line and the nullspace is a plane : $1+2=3$.

Case 3 $\boldsymbol{v}=\boldsymbol{w}=\mathbf{0}$ and the nullspace of A (= zero matrix) is all of $\mathbf{R}^3$: $0+3=3$.

8 $A=\begin{bmatrix}1&4&9\\1&4&9\\1&4&9\end{bmatrix}=\begin{bmatrix}1\\1\\1\end{bmatrix}\begin{bmatrix}1&4&9\end{bmatrix}$ = rank-1 matrix.

9 If $\mathbf{C}(A)=\mathbf{R}^3$ then $m=3$ and $n\geq 3$ and $r=3$.

10 $A_1=\begin{bmatrix}1&3&-2\\3&9&-6\\2&6&-4\end{bmatrix}$ has $C_1=\begin{bmatrix}1\\3\\2\end{bmatrix}$ $\quad A_2=\begin{bmatrix}1&2&3\\4&5&6\\7&8&9\end{bmatrix}$ has $C_2=\begin{bmatrix}1&2\\4&5\\7&8\end{bmatrix}$

11 $A_1=C_1R_1=\begin{bmatrix}1\\3\\2\end{bmatrix}\begin{bmatrix}1&3&-2\end{bmatrix}$ $\quad A_2=C_2R_2=\begin{bmatrix}1&2\\4&5\\7&8\end{bmatrix}\begin{bmatrix}1&0&-1\\0&1&2\end{bmatrix}$

12 The vector $(1,3,2)$ is a basis for $\mathbf{C}(A_1)$. The vectors $(1,4,7)$ and $(2,5,8)$ are a basis for $\mathbf{C}(A_2)$. The dimensions are 1 and 2, so the ranks of the matrices are 1 and 2. Then A_1 and A_2 must have 1 and 2 independent rows.

13 An example is $A=\begin{bmatrix}1&1&1&1\\1&1&1&1\\1&1&0&0\\1&1&0&0\end{bmatrix}$. Then C and R are 4 by 2 and 2 by 4.

14 $A = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$ and $B = \begin{bmatrix} 0 & 2 \\ 0 & 2 \end{bmatrix}$ or $B = \begin{bmatrix} 0 & 2 & 5 \\ 0 & 2 & 5 \end{bmatrix}$ have the same column spaces but different row spaces. The basic columns chosen directly from A and B are $(1,1)$ and $(2,2)$. The rank = *number of vectors* in the column basis must be the same (1).

15 If $A = CR$, then the numbers in row 1 of C multiply the rows of R to produce row 1 of A.

16 "The rows of R are a basis for the row space of A" means: R has independent rows, and every row of A is a combination of the rows of R.

17 $A_1 = \begin{bmatrix} 0 & 1 \\ 0 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} = C_1 R_1 \quad A_2 = \begin{bmatrix} C_1 \\ C_1 \end{bmatrix} \begin{bmatrix} R_1 \end{bmatrix} \qquad A_3 = \begin{bmatrix} C_1 \\ C_1 \end{bmatrix} \begin{bmatrix} R_1 & R_1 \end{bmatrix}$

18 If $A = CR$ then $\begin{bmatrix} 0 & A \\ 0 & A \end{bmatrix} = \begin{bmatrix} C \\ C \end{bmatrix} \begin{bmatrix} 0 & R \end{bmatrix}$

19 $A = \begin{bmatrix} 1 & 3 & 8 \\ 1 & 2 & 6 \\ 0 & 1 & 2 \end{bmatrix} \to \begin{bmatrix} 1 & 3 & 8 \\ 0 & -1 & -2 \\ 0 & 1 & 2 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 2 \\ 0 & -1 & -2 \\ 0 & 1 & 2 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 2 \\ 0 & -1 & -2 \\ 0 & 0 & 0 \end{bmatrix} \to$

$\begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix} =$ **same row space as $\boldsymbol{A}$.** Remove the zero row to see R in $A = CR$.

20 $\boldsymbol{C} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ gives $\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C} = \begin{bmatrix} 13 \end{bmatrix}$. $\boldsymbol{R} = \begin{bmatrix} 2 & 4 \end{bmatrix}$ produces $\boldsymbol{R}\boldsymbol{R}^{\mathrm{T}} = \begin{bmatrix} 20 \end{bmatrix}$.

$A = \begin{bmatrix} 2 & 4 \\ 3 & 6 \end{bmatrix}$ produces $\boldsymbol{C}^{\mathrm{T}} A \boldsymbol{R}^{\mathrm{T}} = \begin{bmatrix} 130 \end{bmatrix}$. Then $\boldsymbol{M} = \frac{1}{13} \begin{bmatrix} 130 \end{bmatrix} \frac{1}{20} = \begin{bmatrix} \frac{1}{2} \end{bmatrix}$.

21 $\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C} = \begin{bmatrix} 1 & 1 & 0 \\ 3 & 2 & 1 \end{bmatrix}\begin{bmatrix} 1 & 3 \\ 1 & 2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 5 \\ 5 & 14 \end{bmatrix}$ has $(\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C})^{-1} = \dfrac{1}{3}\begin{bmatrix} 14 & -5 \\ -5 & 2 \end{bmatrix}$

$\boldsymbol{R}\boldsymbol{R}^{\mathrm{T}} = \begin{bmatrix} 1 & 3 & 8 \\ 1 & 2 & 6 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 3 & 2 \\ 8 & 6 \end{bmatrix} = \begin{bmatrix} 74 & 55 \\ 55 & 41 \end{bmatrix}$ has $(\boldsymbol{R}\boldsymbol{R}^{\mathrm{T}})^{-1} = \dfrac{1}{9}\begin{bmatrix} 41 & -55 \\ -55 & 74 \end{bmatrix}$

$$\boldsymbol{C}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{R}^{\mathrm{T}} = \begin{bmatrix} 1 & 1 & 0 \\ 3 & 2 & 1 \end{bmatrix}\begin{bmatrix} 1 & 3 & 8 \\ 1 & 2 & 6 \\ 0 & 1 & 2 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 3 & 2 \\ 8 & 6 \end{bmatrix} = \begin{bmatrix} 2 & 5 & 14 \\ 5 & 14 & 38 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 3 & 2 \\ 8 & 6 \end{bmatrix} =$$

$$\begin{bmatrix} 129 & 96 \\ 351 & 261 \end{bmatrix}$$

$$\boldsymbol{M} = \frac{1}{3}\begin{bmatrix} 14 & -5 \\ -5 & 2 \end{bmatrix}\begin{bmatrix} 129 & 96 \\ 351 & 261 \end{bmatrix}\frac{1}{9}\begin{bmatrix} 41 & -55 \\ -55 & 74 \end{bmatrix} = \textbf{?}$$

22 If $\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a & ma \\ c & mc \end{bmatrix}$ then $ad - bc = mac - mac = 0$: dependent columns !

23 $\begin{bmatrix} 2 & 4 \\ 2 & 4 \\ 2 & 4 \end{bmatrix} = CR = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}\begin{bmatrix} 1 & 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}\begin{bmatrix} \frac{1}{2} \end{bmatrix}\begin{bmatrix} 2 & 4 \end{bmatrix} = \boldsymbol{CMR}$ (row of $\boldsymbol{R}$ from A)

24 $\begin{bmatrix} 2 & 0 \\ 0 & 2 \\ 0 & 0 \end{bmatrix} = CR = \begin{bmatrix} 2 & 0 \\ 0 & 2 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = \boldsymbol{CMR}$

Problem Set I.2, page 13

1 $A\begin{bmatrix}\boldsymbol{x} & \boldsymbol{y}\end{bmatrix}=\begin{bmatrix}\mathbf{0} & \mathbf{0}\end{bmatrix}$ where $B=\begin{bmatrix}\boldsymbol{x} & \boldsymbol{y}\end{bmatrix}$ is n by 2 and $C=\begin{bmatrix}\mathbf{0} & \mathbf{0}\end{bmatrix}$ is m by 2.

2 Yes, $\boldsymbol{a}\boldsymbol{b}^{\mathrm{T}}$ is an m by n matrix. The number $a_i b_j$ is in row i, column j of $\boldsymbol{a}\boldsymbol{b}^{\mathrm{T}}$. If $\boldsymbol{b}=\boldsymbol{a}$ then $\boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}$ is a *symmetric* matrix.

3 (a) $AB=\boldsymbol{a}_1\boldsymbol{b}_1^{\mathrm{T}}+\cdots+\boldsymbol{a}_n\boldsymbol{b}_n^{\mathrm{T}}$

(b) The i,j entry of AB is $\displaystyle\sum_{k=1}^{n} a_{ik}b_{kj}$.

4 If B has one column $\begin{bmatrix} b_1 & \dots & b_n\end{bmatrix}^{\mathrm{T}}$ then $AB=\boldsymbol{a}_1 b_1+\cdots+\boldsymbol{a}_n b_n=$ combination of the columns of A (*as expected*). Each row of B is one number b_k.

5 Verify $(AB)\,C=A\,(BC)$ for $AB=\begin{bmatrix}1 & a\\ 0 & 1\end{bmatrix}\begin{bmatrix}b_1 & b_2\\ b_3 & b_4\end{bmatrix}=\begin{bmatrix}b_1+ab_3 & b_2+ab_4\\ b_3 & b_4\end{bmatrix}$

and $BC=\begin{bmatrix}b_1 & b_2\\ b_3 & b_4\end{bmatrix}\begin{bmatrix}1 & 0\\ c & 1\end{bmatrix}=\begin{bmatrix}b_1+cb_2 & b_2\\ b_3+cb_4 & b_4\end{bmatrix}$ AB was row ops, BC was col ops

Row ops then col ops $\begin{bmatrix}b_1+ab_3 & b_2+a_2b_4\\ b_3 & b_4\end{bmatrix}\begin{bmatrix}1 & 0\\ c & 1\end{bmatrix}=\begin{bmatrix}b_1+ab_3+cb_2+acb_4 & b_2+ab_4\\ b_3+cb_4 & b_4\end{bmatrix}$

Col ops then row ops $\begin{bmatrix}1 & a\\ 0 & 1\end{bmatrix}\begin{bmatrix}b_1+cb_2 & b_2\\ b_3+cb_4 & b_4\end{bmatrix}=\begin{bmatrix}b_1+ab_3+cb_2+acb_4 & b_2+ab_4\\ b_3+cb_4 & b_4\end{bmatrix}$ **SAME**

If A, C were both row operations, $(AC)\,B=(CA)\,B$ would usually be false.

6 $B=I$ has rows $\boldsymbol{b}_1,\boldsymbol{b}_2,\boldsymbol{b}_3=(1,0,0),(0,1,0),(0,0,1)$. The rank-1 matrices are

$\boldsymbol{a}_1\boldsymbol{b}_1=\begin{bmatrix}\boldsymbol{a}_1 & \mathbf{0} & \mathbf{0}\end{bmatrix}\quad \boldsymbol{a}_2\boldsymbol{b}_2=\begin{bmatrix}\mathbf{0} & \boldsymbol{a}_2 & \mathbf{0}\end{bmatrix}\quad \boldsymbol{a}_3\boldsymbol{b}_3=\begin{bmatrix}\mathbf{0} & \mathbf{0} & \boldsymbol{a}_3\end{bmatrix}$.

*The sum of those rank-*1 *matrices is* $AI=A$.

7 If $A=\begin{bmatrix}1 & 2\\ 3 & 4\end{bmatrix}$ and $B=\begin{bmatrix}0 & 0\\ 1 & 0\end{bmatrix}$ then $AB=\begin{bmatrix}2 & 0\\ 4 & 0\end{bmatrix}$ has a smaller column space than A. Note that (row space of AB) $\leq$ (row space of B).

8 **TO DO**

Problem Set I.3, page 20

1 If $B\boldsymbol{x}=\mathbf{0}$ then $AB\boldsymbol{x}=\mathbf{0}$. So every $\boldsymbol{x}$ in the nullspace of B is also in the nullspace of AB.

2 $A=\begin{bmatrix}0&1\\0&0\end{bmatrix}$ has $A^2=\begin{bmatrix}0&0\\0&0\end{bmatrix}$ and the rank has dropped.

But $A^{\mathrm{T}}A=\begin{bmatrix}0&0\\1&0\end{bmatrix}\begin{bmatrix}0&1\\0&0\end{bmatrix}=\begin{bmatrix}0&0\\0&1\end{bmatrix}$ has the same nullspace and rank as A.

3 If $C=\begin{bmatrix}A\\B\end{bmatrix}$ then $C\boldsymbol{x}=\mathbf{0}$ requires both $A\boldsymbol{x}=\mathbf{0}$ and $B\boldsymbol{x}=\mathbf{0}$. So the nullspace of C is the **intersection** $\mathbf{N}(A)\cap\mathbf{N}(B)$.

4 Actually row space = column space requires nullspace of A = nullspace of A^{T}. But it *does not* require symmetry. Choose any invertible matrix like $A=\begin{bmatrix}1&2\\3&4\end{bmatrix}$.

5 $r=m=n$ $\quad$ A_1 is any invertible square matrix

$r=m<n$ $\quad$ A_2 has extra columns like $A_2=\begin{bmatrix}1&0&0\\0&1&1\end{bmatrix}$

$r=n<m$ $\quad$ A_3 has extra rows like A_2^{T}

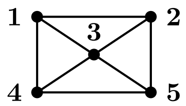
$r<m, r<n$ $\quad$ $A_4=\begin{bmatrix}1&0\\0&0\end{bmatrix}$ or $\begin{bmatrix}1&2\\1&2\end{bmatrix}$ or $\begin{bmatrix}1&2&3\\1&2&3\end{bmatrix}$.

6 First, if $A\boldsymbol{x}=\mathbf{0}$ then $A^{\mathrm{T}}A\boldsymbol{x}=\mathbf{0}$. So $\mathbf{N}(A^{\mathrm{T}}A)$ contains (or equals) $\mathbf{N}(A)$.

Second, if $A^{\mathrm{T}}A\boldsymbol{x}=0$ then $\boldsymbol{x}^{\mathrm{T}}A^{\mathrm{T}}A\boldsymbol{x}=0$ and $||A\boldsymbol{x}||^2=0$. Then $A\boldsymbol{x}=\mathbf{0}$ and $\mathbf{N}(A)$ contains (or equals) $\mathbf{N}(A^{\mathrm{T}}A)$. Altogether $\mathbf{N}(A^{\mathrm{T}}A)$ **equals** $\mathbf{N}(A)$.

7 $A=\begin{bmatrix}0&1\\0&0\end{bmatrix}$ and $A^2=\begin{bmatrix}0&0\\0&0\end{bmatrix}$ have *different* nullspaces.

8 $A=\begin{bmatrix}0&1\\0&0\end{bmatrix}$ has $\mathbf{C}(A)=\mathbf{N}(A)=$all vectors $\begin{bmatrix}x_1\\0\end{bmatrix}$. But $\mathbf{C}(A)=\mathbf{N}(A^{\mathrm{T}})$ is impossible.

9 

8 edges
5 nodes

$$A\boldsymbol{x} = \mathbf{0} \text{ for } \boldsymbol{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Columns 1 to 5 are *dependent*
Columns 1 to 4 are *independent*

$$\text{Incidence matrix } A = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 1 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix}$$

has rank 4
$\mathbf{N}(A)$ has dimension 1
$\mathbf{N}(A^{\mathrm{T}})$ has dimension 4
$8 - 4 = 4$ small loops

10 If $\mathbf{N}(A) = \{\mathbf{0}\}$, the nullspace of $B = \begin{bmatrix} A & A & A \end{bmatrix}$ contains all vectors $\begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \\ \boldsymbol{z} \end{bmatrix}$ with $\boldsymbol{x} + \boldsymbol{y} + \boldsymbol{z} = \mathbf{0}$.

11 (i) $S \cap T$ has dimension $0, 1,$ or 2

(ii) $S + T$ has dimension $7, 8,$ or 9

(iii) $S^{\perp} =$ (vectors perpendicular to S) has dimension $10 - 2 = 8$.

Problem Set I.4, page 27

1 $$\begin{bmatrix} 2 & 1 \\ 6 & 7 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 0 & 4 \end{bmatrix}$$

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

$$\begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -1/2 & 1 & 0 \\ 0 & -2/3 & 1 \end{bmatrix} \begin{bmatrix} 2 & -1 & 0 \\ 0 & 3/2 & -1 \\ 0 & 0 & 4/3 \end{bmatrix}$$

2 $a_{ij} = a_{i1}a_{1j}/a_{11}$ *Check* $A = \begin{bmatrix} 2 & 3 \\ 4 & 6 \end{bmatrix}$ and $a_{22} = (4)(3)/(2)$.

If $a_{11} = 0$ then the formula breaks down. We could still have rank 1.

3 $EA = U$ is $$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 0 \\ 0 & 4 & 2 \\ 6 & 3 & 5 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 4 & 2 \\ 0 & 0 & 5 \end{bmatrix}$$

$A = LU$ is $$\begin{bmatrix} 2 & 1 & 0 \\ 0 & 4 & 2 \\ 6 & 3 & 5 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 0 \\ 0 & 4 & 2 \\ 0 & 0 & 5 \end{bmatrix}$$ **Note $L = E^{-1}$**

4 $$E_2E_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -c & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -a & 1 & 0 \\ -b & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -a & 1 & 0 \\ ac-b & -c & 1 \end{bmatrix}$$ $ac - b$ mixes the multipliers

$$E_1^{-1}E_2^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ a & 1 & 0 \\ b & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & c & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ a & 1 & 0 \\ b & c & 1 \end{bmatrix}$$ In this order the multipliers fall into place in L

5 If zero appears in a pivot position then $A = LU$ is **not possible**. We need a *permutation* P to exchange rows and lead to nonzero pivots.

$$\begin{bmatrix} 0 & 1 \\ 2 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \ell & 1 \end{bmatrix} \begin{bmatrix} d & e \\ 0 & f \end{bmatrix} \text{ leads to } \begin{matrix} 0 = d & (1,1 \text{ entry}) \\ 2 = \ell d & (\text{impossible if } d = 0) \end{matrix}$$

$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 2 \\ 1 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ \ell & 1 & 0 \\ m & n & 1 \end{bmatrix} \begin{bmatrix} d & e & g \\ 0 & f & h \\ 0 & 0 & i \end{bmatrix} \text{ leads to } \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \ell & 1 \end{bmatrix} \begin{bmatrix} d & e \\ 0 & f \end{bmatrix}$$

second pivot is zero

Then $\begin{bmatrix} d & e \\ 0 & f \end{bmatrix}$ must be singular and $\begin{bmatrix} d & e & g \\ 0 & f & h \\ 0 & 0 & i \end{bmatrix}$ is singular. BUT $\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 2 \\ 1 & 2 & 1 \end{bmatrix}$ is invertible ! So $A = LU$ is again impossible.

6 $c = 2$ makes the second pivot zero. But A is still invertible.

$c = 1$ makes the third pivot zero. Then A is singular.

7 $A = LU$ is $\begin{bmatrix} a & a & a & a \\ a & b & b & b \\ a & b & c & c \\ a & b & c & d \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} a & a & a & a \\ 0 & b-a & b-a & b-a \\ 0 & 0 & c-b & c-b \\ 0 & 0 & 0 & d-c \end{bmatrix}$

For nonzero pivots in U, we need $a \neq 0, b \neq a, c \neq b, d \neq c$.

8 If A is tridiagonal and $A = LU$ (no row exchanges in elimination) then L and U have *two diagonals*. The only elimination steps subtract a pivot row from the row directly beneath it.

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

$$A = \begin{bmatrix} a & a & 0 \\ a & a+b & b \\ 0 & b & b+c \end{bmatrix} = \begin{bmatrix} 1 & & \\ 1 & 1 & \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} a & a & 0 \\ & b & b \\ & & c \end{bmatrix}$$

9 The second pivot in elimination depends only on the upper left 2 by 2 submatrix A_2 of A. The third pivot depends only on the upper left 3 by 3 submatrix (and so on). So if the pivots (diagonal entries in U) are $5, 9, 3$, then the pivots for A_2 are $5, 9$.

10 Continuing Problem 9, the upper left parts of L and U **come from the upper left part of $\boldsymbol{A}$**. Then $L_k U_k$ is the factorization of A_k.

$$A = \begin{bmatrix} L_k & 0 \\ * & * \end{bmatrix} \begin{bmatrix} U_k & * \\ 0 & * \end{bmatrix} \quad \text{so} \quad A_k = L_k U_k$$

11 The example could exchange rows of A to put the larger number 3 into the $(1, 1)$ position where it would become the first pivot. That would be the usual permutation in MATLAB and other systems.

This problem also exchanges columns to put the even larger number 4 into the $(1, 1)$ position. A column exchange comes from a permutation multiplying on the *right side* of A. So this problem works on both sides:

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \text{ leads to } P_1 A P_2 = \begin{bmatrix} 4 & 3 \\ 2 & 1 \end{bmatrix} \quad A = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} \text{ leads to } P_2 A P_1 = \begin{bmatrix} 4 & 2 \\ 3 & 1 \end{bmatrix}$$

12 With m rows and n columns and $m < n$, elimination normally leads from A to

$$U = \underset{m \times n \quad m \times (n-m)}{\begin{bmatrix} U_1 & U_2 \end{bmatrix}} \qquad \textit{Example}: U = \begin{bmatrix} 1 & 2 & 4 \\ 0 & 3 & 9 \end{bmatrix}$$

There must be nonzero solutions to $U\boldsymbol{x} = \mathbf{0}$. To see this, set $x_3 = 1$ and solve

$$\begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = - \begin{bmatrix} 4 \\ 9 \end{bmatrix} \text{ to find } \begin{matrix} x_1 = 2 \\ x_2 = -3 \end{matrix}. \text{ So } \boldsymbol{x} = \begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix} \text{ solves } A\boldsymbol{x} = \mathbf{0}.$$

Problem Set I.5, page 35

1 If $\boldsymbol{u}^{\mathrm{T}}\boldsymbol{v}=0$ and $\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}=1$ and $\boldsymbol{v}^{\mathrm{T}}\boldsymbol{v}=1$, then $(\boldsymbol{u}+\boldsymbol{v})^{\mathrm{T}}(\boldsymbol{u}-\boldsymbol{v})=1+0-0-1=\mathbf{0}$
Also $||\boldsymbol{u}+\boldsymbol{v}||^2=\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}+\boldsymbol{v}^{\mathrm{T}}\boldsymbol{u}+\boldsymbol{u}^{\mathrm{T}}\boldsymbol{v}+\boldsymbol{v}^{\mathrm{T}}\boldsymbol{v}=1+0+0+1=2$ and $||\boldsymbol{u}-\boldsymbol{v}||^2=2$

2 $\boldsymbol{v}$ is separated into a piece $\boldsymbol{u}(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{v})$ in the direction of $\boldsymbol{u}$ and the remaining piece $\boldsymbol{w}=\boldsymbol{v}-\boldsymbol{u}(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{v})$ perpendicular to $\boldsymbol{u}$. Check $\boldsymbol{u}^{\mathrm{T}}\boldsymbol{w}=\boldsymbol{u}^{\mathrm{T}}\boldsymbol{v}-(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u})(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{v})=0$.

3 $\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w}+\boldsymbol{z}^{\mathrm{T}}\boldsymbol{z}=(\boldsymbol{u}+\boldsymbol{v})^{\mathrm{T}}(\boldsymbol{u}+\boldsymbol{v})+(\boldsymbol{u}-\boldsymbol{v})^{\mathrm{T}}(\boldsymbol{u}-\boldsymbol{v})=(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}+\boldsymbol{v}^{\mathrm{T}}\boldsymbol{v}+\boldsymbol{u}^{\mathrm{T}}\boldsymbol{v}+\boldsymbol{v}^{\mathrm{T}}\boldsymbol{u})+(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}+\boldsymbol{v}^{\mathrm{T}}\boldsymbol{v}-\boldsymbol{u}^{\mathrm{T}}\boldsymbol{v}-\boldsymbol{v}^{\mathrm{T}}\boldsymbol{u})=2(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}+\boldsymbol{v}^{\mathrm{T}}\boldsymbol{v})$.

Sum of squares of 2 diagonals $=$ Sum of squares of 4 sides.

4 Check $(Q\boldsymbol{x})^{\mathrm{T}}(Q\boldsymbol{y})=\boldsymbol{x}^{\mathrm{T}}Q^{\mathrm{T}}Q\,\boldsymbol{y}=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$: Angles are preserved when all vectors are multiplied by Q. Remember $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}=||\boldsymbol{x}||\,||\boldsymbol{y}||\cos\theta=(Q\boldsymbol{x})^{\mathrm{T}}(Q\boldsymbol{y})$: same θ !

5 If Q is orthogonal (this word assumes a square matrix) then $Q^{\mathrm{T}}Q=I$ and $\boldsymbol{Q^{\mathrm{T}}}$ is $\boldsymbol{Q^{-1}}$. Check $(Q_1Q_2)^{\mathrm{T}}=Q_2^{\mathrm{T}}Q_1^{\mathrm{T}}=Q_2^{-1}Q_1^{-1}$ which is $(Q_1Q_2)^{-1}$.

6 Every permutation matrix has unit vectors in its columns (single 1 and $n-1$ zeros). Those columns are orthogonal because their 1's are in different positions.

7 $$PF=\begin{bmatrix}0&1&0&0\\0&0&1&0\\0&0&0&1\\1&0&0&0\end{bmatrix}\begin{bmatrix}1&1&1&1\\1&i&i^2&i^3\\1&i^2&i^4&i^6\\1&i^3&i^6&i^9\end{bmatrix}=\begin{bmatrix}1&i&i^2&i^3\\1&i^2&i^4&i^6\\1&i^3&i^6&i^9\\1&1&1&1\end{bmatrix}=\begin{bmatrix}1&1&1&1\\1&i&i^2&i^3\\1&i^2&i^4&i^6\\1&i^3&i^6&i^9\end{bmatrix}\begin{bmatrix}1&&&\\&i&&\\&&i^2&\\&&&i^3\end{bmatrix}$$

This says that P times the 4 columns of F gives those same 4 columns times $1,i,i^2,i^3=\lambda_1,\lambda_2,\lambda_3,\lambda_4=$ the 4 eigenvalues of P.

The columns of $F/2$ are orthonormal! To check, remember that for the dot product of two *complex vectors*, we take complex conjugates of the first vector: *change i to $-i$*.

8 $W^{\mathrm{T}}W = \begin{bmatrix} 4 & & & \\ & 4 & & \\ & & 2 & \\ & & & 2 \end{bmatrix}$ so that the columns of W are orthogonal but not orthonormal.

Then $W^{-1} = (W^{\mathrm{T}}W)^{-1}W^{\mathrm{T}} = \begin{bmatrix} 1/4 & & & \\ & 1/4 & & \\ & & 1/2 & \\ & & & 1/2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}$.

Problem Set I.6, page 41

1 To check : $\lambda_1+\lambda_2 =$ trace $=\cos\theta+\cos\theta$ and $\lambda_1\lambda_2 =$ determinant $=\cos^2\theta+\sin^2\theta = 1$ and $\overline{\boldsymbol{x}}_1^{\mathrm{T}}\boldsymbol{x}_2 = 0$ (orthogonal matrices have complex orthogonal eigenvectors).

$$Q^{-1} = Q^{\mathrm{T}} \text{ has eigenvalues } \frac{1}{e^{i\theta}} = e^{-i\theta} \text{ and } \frac{1}{e^{-i\theta}} = e^{i\theta}$$

2 $\det\begin{bmatrix} -\lambda & 2 \\ 1 & 1-\lambda \end{bmatrix} = \lambda^2-\lambda-2 = (\lambda-2)(\lambda+1) = 0$ gives $\lambda_1 = 2$ and $\lambda_2 = -1$. The sum $2-1$ agrees with the trace $0+1$. A^{-1} has the same eigenvectors as A, with eigenvalues $\lambda_1^{-1} = \boldsymbol{\frac{1}{2}}$ and $\lambda_2^{-1} = \boldsymbol{-1}$.

3 A has $\lambda = 3$ and 1, B has $\lambda = 1$ and 3, $A+B$ has $\lambda = 5$ and 3. Eigenvalues of $A+B$ are generally **not equal** to $\lambda(A)+\lambda(B)$. Now A and B have $\lambda = 1$ (*repeated*). AB and BA both have $\lambda^2-4\lambda+1 = 0$ (leading to $\lambda = 2\pm\sqrt{3}$ by the quadratic formula). The eigenvalues of AB and BA are the same—but not equal to $\lambda(A)$ times $\lambda(B)$.

4 A and B have $\lambda_1 = 1$ and $\lambda_2 = 1$. AB and BA have $\lambda^2-4\lambda+1$ and the quadratic formula gives $\lambda = 2\pm\sqrt{3}$. Eigenvalues of AB *are not equal* to eigenvalues of A times eigenvalues of B. Eigenvalues of AB and BA ***are equal*** (this is proved at the end of Section 6.2).

5 (a) Multiply $A\boldsymbol{x}$ to see $\lambda\boldsymbol{x}$ which reveals λ (b) Solve $(A-\lambda I)\boldsymbol{x} = \boldsymbol{0}$ to find $\boldsymbol{x}$.

6 $\det(A-\lambda I) = \lambda^2-1.4\lambda+0.4$ so A has $\lambda_1 = 1$ and $\lambda_2 = 0.4$ with $\boldsymbol{x}_1 = (1,2)$ and $\boldsymbol{x}_2 = (1,-1)$. A^∞ has $\lambda_1 = 1$ and $\lambda_2 = 0$ (same eigenvectors). A^{100} has $\lambda_1 = 1$ and $\lambda_2 = (0.4)^{100}$ which is near zero. So A^{100} is very near A^∞: same eigenvectors and close eigenvalues.

7 Set $\lambda = 0$ in $\det(A-\lambda I) = (\lambda_1-\lambda)\ldots(\lambda_n-\lambda)$ to find $\det A = (\lambda_1)(\lambda_2)\cdots(\lambda_n)$.

8 $\lambda_1 = \frac{1}{2}(a+d+\sqrt{(a-d)^2-4bc})$ and $\lambda_2 = \frac{1}{2}(a+d-\sqrt{})$ add to $a+d$. If A has $\lambda_1 = 3$ and $\lambda_2 = 4$ then $\det(A-\lambda I) = (\lambda-3)(\lambda-4) = \lambda^2-7\lambda+12$.

9 These 3 matrices have $\lambda = 4$ and 5, trace 9, det 20: $\begin{bmatrix} 4 & 0 \\ 0 & 5 \end{bmatrix}, \begin{bmatrix} 3 & 2 \\ -1 & 6 \end{bmatrix}, \begin{bmatrix} 2 & 2 \\ -3 & 7 \end{bmatrix}$.

10 $A = \begin{bmatrix} 0 & 1 \\ \mathbf{-28} & \mathbf{11} \end{bmatrix}$ has trace 11 and determinant 28, so $\lambda = 4$ and 7. Moving to a 3 by 3 companion matrix, for eigenvalues $1, 2, 3$ we want $\det(C - \lambda I) = (1-\lambda)(2-\lambda)(3-\lambda)$. Multiply out to get $-\lambda^3 + 6\lambda^2 - 11\lambda + 6$. To get those numbers $6, -11, 6$ from a companion matrix you just put them into the last row:

$C = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \mathbf{6} & \mathbf{-11} & \mathbf{6} \end{bmatrix}$ Notice the trace $6 = 1 + 2 + 3$ and determinant $6 = (1)(2)(3)$.

11 $(A - \lambda I)$ has the same determinant as $(A - \lambda I)^{\mathrm{T}}$ because every square matrix has $\det M = \det M^{\mathrm{T}}$. Pick $M = A - \lambda I$.

$\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$ and $\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$ have *different eigenvectors*.

12 $\lambda = \mathbf{0, 0, 6}$ (*notice rank* 1 *and trace* 6). Two eigenvectors of $\boldsymbol{uv}^{\mathrm{T}}$ are perpendicular to $\boldsymbol{v}$ and the third eigenvector is $\boldsymbol{u}$: $\boldsymbol{x}_1 = (0, -2, 1)$, $\boldsymbol{x}_2 = (1, -2, 0)$, $\boldsymbol{x}_3 = (1, 2, 1)$.

13 When A and B have the same n λ's and $\boldsymbol{x}$'s, look at any combination $\boldsymbol{v} = c_1\boldsymbol{x}_1 + \cdots + c_n\boldsymbol{x}_n$. Multiply by A and B: $A\boldsymbol{v} = c_1\lambda_1\boldsymbol{x}_1 + \cdots + c_n\lambda_n\boldsymbol{x}_n$ **equals** $B\boldsymbol{v} = c_1\lambda_1\boldsymbol{x}_1 + \cdots + c_n\lambda_n\boldsymbol{x}_n$ **for all vectors** $\boldsymbol{v}$. So $A = B$.

14 (a) $\boldsymbol{u}$ is a basis for the nullspace (we know $A\boldsymbol{u} = 0\boldsymbol{u}$); $\boldsymbol{v}$ and $\boldsymbol{w}$ give a basis for the column space (we know $A\boldsymbol{v}$ and $A\boldsymbol{w}$ are in the column space).

(b) $A(\boldsymbol{v}/3 + \boldsymbol{w}/5) = 3\boldsymbol{v}/3 + 5\boldsymbol{w}/5 = \boldsymbol{v} + \boldsymbol{w}$. So $\boldsymbol{x} = \boldsymbol{v}/3 + \boldsymbol{w}/5$ is a particular solution to $A\boldsymbol{x} = \boldsymbol{v} + \boldsymbol{w}$. Add any $c\boldsymbol{u}$ from the nullspace

(c) If $A\boldsymbol{x} = \boldsymbol{u}$ had a solution, $\boldsymbol{u}$ would be in the column space: wrong dimension 3.

15 Eigenvectors in $\boldsymbol{X}$ and eigenvalues in Λ. Then $A = X\Lambda X^{-1}$ is given below. The second matrix has $\lambda = 0$ (rank 1) and $\lambda = 4$ (trace $= 4$). A new $A = X\Lambda X^{-1}$:

$$\begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}. \quad \begin{bmatrix} 1 & 1 \\ 3 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} \frac{3}{4} & -\frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} \end{bmatrix}.$$

16 If $A = X\Lambda X^{-1}$ then the eigenvalue matrix for $A + 2I$ is $\Lambda + 2I$ and the eigenvector matrix is still X. So $A + 2I = X(\Lambda + 2I)X^{-1} = X\Lambda X^{-1} + X(2I)X^{-1} = A + 2I$.

17 (a) False: We are not given the λ's (b) True (c) True (d) False: For this we would need the eigenvectors of X.

18 $A = X\Lambda X^{-1} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \lambda_1 & \\ & \lambda_2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} /2 = \begin{bmatrix} \lambda_1 + \lambda_2 & \lambda_1 - \lambda_2 \\ \lambda_1 - \lambda_2 & \lambda_1 + \lambda_2 \end{bmatrix} /2.$

These are the matrices $\begin{bmatrix} a & b \\ b & a \end{bmatrix}$, their eigenvectors are $(1, 1)$ and $(1, -1)$.

19 (a) *True* (no zero eigenvalues) (b) *False* (repeated $\lambda = 2$ may have only one line of eigenvectors) (c) *False* (repeated λ may have a full set of eigenvectors)

20 (a) False: don't know if $\lambda = 0$ or not.

(b) True: an eigenvector is missing, which can only happen for a repeated eigenvalue.

(c) True: We know there is only one line of eigenvectors.

21 $A^k = X\Lambda^k X^{-1}$ approaches zero **if and only if every** $|\boldsymbol{\lambda}| < \mathbf{1}$; A_1 is a Markov matrix so $\lambda_{\text{max}} = 1$ and $A_1^k \to A_1^\infty$, A_2 has $\lambda = .6 \pm .3$ so $A_2^k \to 0$.

22 $\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} = X\Lambda X^{-1} = \frac{1}{2} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$ and

$$A^k = X\Lambda^k X^{-1} = \frac{1}{2} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 3^k \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}.$$

Multiply those last three matrices to get $A^k = \frac{1}{2} \begin{bmatrix} 1 + 3^k & 1 - 3^k \\ 1 - 3^k & 1 + 3^k \end{bmatrix}$.

23 $R = X\sqrt{\Lambda}X^{-1} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 3 & \\ & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} /2 = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ has $R^2 = A$.

$\sqrt{B}$ needs $\lambda = \sqrt{9}$ and $\sqrt{-1}$, trace (their sum) is not real so $\sqrt{B}$ cannot be real. Note that $\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$ has *two* imaginary eigenvalues $\sqrt{-1} = i$ and $-i$, real trace 0, real square root $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$.

24 $A = X\Lambda_1 X^{-1}$ and $B = X\Lambda_2 X^{-1}$. Diagonal matrices always give $\Lambda_1\Lambda_2 = \Lambda_2\Lambda_1$. Then $AB = BA$ from

$$X\Lambda_1 X^{-1} X\Lambda_2 X^{-1} = X\boldsymbol{\Lambda_1\Lambda_2}X^{-1} = X\boldsymbol{\Lambda_2\Lambda_1}X^{-1} = X\Lambda_2 X^{-1} X\Lambda_1 X^{-1} = BA.$$

25 Multiply columns of X times rows of ΛX^{-1}.

26 To have $A = B\Lambda B^{-1}$ requires A to have a full set of n independent eigenvectors. Then B is the *eigenvector matrix* and it is invertible.

Problem Set I.7, page 52

1 The key is to form $\boldsymbol{y}^{\mathrm{T}}S\boldsymbol{x}$ in two ways, using $S^{\mathrm{T}} = S$ to make them agree. One way starts with $S\boldsymbol{x} = \lambda\boldsymbol{x}$: multiply by $\boldsymbol{y}^{\mathrm{T}}$. The other way starts with $S\boldsymbol{y} = \alpha\boldsymbol{y}$ and then $\boldsymbol{y}^{\mathrm{T}}S^{\mathrm{T}} = \alpha\boldsymbol{y}^{\mathrm{T}}$.

The final step finds $0 = (\lambda - \alpha)\,\boldsymbol{y}^{\mathrm{T}}\boldsymbol{x}$ which forces $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{x} = 0$.

2 Only $S_4 = \begin{bmatrix} 1 & 10 \\ 10 & 101 \end{bmatrix}$ has two positive eigenvalues since $101 > 10^2$.

$\boldsymbol{x}^{\mathrm{T}}S_1\boldsymbol{x} = 5x_1^2 + 12x_1x_2 + 7x_2^2$ is negative for example when $x_1 = 4$ and $x_2 = -3$: A_1 is not positive definite as its determinant confirms; S_2 has trace c_0; S_3 has $\det = 0$.

3 Positive definite for $-3 < b < 3$ $\begin{bmatrix} 1 & 0 \\ b & 1 \end{bmatrix}\begin{bmatrix} 1 & b \\ 0 & 9-b^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ b & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 9-b^2 \end{bmatrix}\begin{bmatrix} 1 & b \\ 0 & 1 \end{bmatrix} = LDL^{\mathrm{T}}$

Positive definite for $c > 8$ $\begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} 2 & 4 \\ 0 & c-8 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} 2 & 0 \\ 0 & c-8 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} = LDL^{\mathrm{T}}.$

Positive definite for $c^2 > b$ $L = \begin{bmatrix} 1 & 0 \\ -b/c & 1 \end{bmatrix}$ $D = \begin{bmatrix} c & 0 \\ 0 & c - b/c \end{bmatrix}$ $S = LDL^{\mathrm{T}}.$

4 If $\boldsymbol{x}$ is not real then $\lambda = \boldsymbol{x}^{\mathrm{T}}A\boldsymbol{x}/\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}$ is *not* always real. Can't assume real eigenvectors!

5 $\begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} = 2\begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix} + 4\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$; $\begin{bmatrix} 9 & 12 \\ 12 & 16 \end{bmatrix} = 0\begin{bmatrix} .64 & -.48 \\ -.48 & .36 \end{bmatrix} + 25\begin{bmatrix} .36 & .48 \\ .48 & .64 \end{bmatrix}$

6 M is skew-symmetric and **orthogonal**; λ's must be $i, i, -i, -i$ to have trace 0, $|\lambda| = 1$.

7 $A = \begin{bmatrix} i & 1 \\ 1 & -i \end{bmatrix}$ has $\lambda = 0, 0$ and only one independent eigenvector $\boldsymbol{x} = (i, 1)$. The good property for complex matrices is not $A^{\mathrm{T}} = A$ (symmetric) but $\overline{A}^{\mathrm{T}} = A$ (Hermitian with real eigenvalues and orthogonal eigenvectors: see Problem 22 and Section 9.2).

8 Eigenvectors $(1, 0)$ and $(\mathbf{1}, \mathbf{1})$ give a $45°$ angle even with A^{T} very close to A.

9 (a) $S^{\mathrm{T}} = S$ and $S^{\mathrm{T}}S = I$ lead to $S^2 = I$.

(b) The only possible eigenvalues of S are 1 and -1.

(c) $\Lambda = \begin{bmatrix} I & 0 \\ 0 & -I \end{bmatrix}$ so $S = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \Lambda \begin{bmatrix} Q_1^{\mathrm{T}} \\ Q_2^{\mathrm{T}} \end{bmatrix} = Q_1Q_1^{\mathrm{T}} - Q_2Q_2^{\mathrm{T}}$ with $Q_1^{\mathrm{T}}Q_2 = 0$.

10 Eigenvalues of $A^{\mathrm{T}}SA$ are different from eigenvalues of S but the signs are the same: the Law of Inertia gives the same number of plus-minus-zero eigenvalues.

11 $\det(S - aI) = \begin{vmatrix} 0 & b \\ b & c-a \end{vmatrix} = -b^2$ is negative. So the point $x = a$ is between the two eigenvalues where $\det(S - \lambda_1 I) = 0$ and $\det(S - \lambda_2 I) = 0$. This $\lambda_1 \le a \le \lambda_2$ is a general rule for larger matrices too (Section II.2): Eigenvalues of the submatrix of size $n-1$ **interlace** eigenvalues of the n by n symmetric matrix.

12 $\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x} = 2x_1x_2$ comes from $\begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$. That matrix has eigenvalues 1 and -1. Conclusion: Saddle points are associated with eigenvalues *of both signs*.

13 $A^{\mathrm{T}}A = \begin{bmatrix} 1 & 2 \\ 2 & 13 \end{bmatrix}$ and $A^{\mathrm{T}}A = \begin{bmatrix} 6 & 5 \\ 5 & 6 \end{bmatrix}$ are positive definite; $A^{\mathrm{T}}A = \begin{bmatrix} 2 & 3 & 3 \\ 3 & 5 & 4 \\ 3 & 4 & 5 \end{bmatrix}$ is singular (and positive semidefinite). The first two A's have independent columns. The 2 by 3 A cannot have full column rank 3, with only 2 rows; $A^{\mathrm{T}}A$ is singular.

14 $S = \begin{bmatrix} 4 & -4 & 8 \\ -4 & 4 & -8 \\ 8 & -8 & 16 \end{bmatrix}$ has only one pivot $= 4$, rank $S = 1$, eigenvalues are $24, 0, 0$, $\det S = 0$.

15 Corner determinants $|S_1| = 2$, $|S_2| = 6$, $|S_3| = 30$. The pivots are $2/1, 6/2, 30/6$.

16 S is positive definite for $c > 1$; determinants $c, c^2 - 1$, and $(c-1)^2(c+2) > 0$. T is *never* positive definite (determinants $d - 4$ and $-4d + 12$ are never both positive).

17 $S = \begin{bmatrix} 1 & 5 \\ 5 & 10 \end{bmatrix}$ is an example with $a + c > 2b$ but $ac < b^2$, so not positive definite.

18 $\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}$ is zero when $(x_1, x_2, x_3) = (0, 1, 0)$ because of the zero on the diagonal. Actually $\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}$ goes *negative* for $\boldsymbol{x} = (1, -10, 0)$ because the second pivot is *negative*.

19 If a_{jj} were smaller than all λ's, $S - a_{jj}I$ would have all eigenvalues > 0 (positive definite). But $S - a_{jj}I$ has a *zero* in the (j, j) position; impossible by Problem 18.

20 $A = \dfrac{\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}}{\sqrt{2}} \begin{bmatrix} \sqrt{9} & \\ & \sqrt{1} \end{bmatrix} \dfrac{\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}}{\sqrt{2}} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}; A = Q \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix} Q^{\mathrm{T}} = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}.$

21 The ellipse $x^2 + xy + y^2 = 1$ has axes with half-lengths $1/\sqrt{\lambda} = \sqrt{2}$ and $\sqrt{2/3}$.

22 The Cholesky factors $A = \left(L\sqrt{D}\right)^{\mathrm{T}} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 2 \end{bmatrix}$ and $A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & \sqrt{5} \end{bmatrix}$ have *square roots* of the pivots from D. Note again $A^{\mathrm{T}}A = LDL^{\mathrm{T}} = S$.

23 The energy test gives $\boldsymbol{x}^{\mathrm{T}}(A^{\mathrm{T}}CA)\,\boldsymbol{x} = (A\boldsymbol{x})^{\mathrm{T}}C(A\boldsymbol{x}) = \boldsymbol{y}^{\mathrm{T}}C\boldsymbol{y} > 0$ since C is positive definite and $\boldsymbol{y} = A\boldsymbol{x}$ is only zero if $\boldsymbol{x}$ is zero. (A was assumed to have independent columns.)

This is just like the $A^{\mathrm{T}}A$ discussion, but now with a positive definite C in $A^{\mathrm{T}}CA$.

24 $S_1 = \begin{bmatrix} 6x^2 & 2x \\ 2x & 2 \end{bmatrix}$ is semidefinite; $f_1 = (\frac{1}{2}x^2 + y)^2 = 0$ on the curve $\frac{1}{2}x^2 + y = 0$; $S_2 = \begin{bmatrix} 6x & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ is indefinite at $(0, 1)$ where first derivatives $= 0$. Then $x = 0, y = 1$ is a saddle point of the function $f_2(x, y)$.

25 If $c > 9$ the graph of z is a bowl, if $c < 9$ the graph has a saddle point. When $c = 9$ the graph of $z = (2x + 3y)^2$ is a "trough" staying at zero along the line $2x + 3y = 0$.

26 $\det S = (1)(10)(1) = 10$; $\lambda = 2$ and 5; $\boldsymbol{x}_1 = (\cos\theta, \sin\theta)$, $\boldsymbol{x}_2 = (-\sin\theta, \cos\theta)$; the λ's are positive. So S is positive definite.

27 Energy $\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x} = a\,(x_1+x_2+x_3)^2 + c\,(x_2-x_3)^2 \geq 0$ if $a \geq 0$ and $c \geq 0$: semidefinite.

S has rank ≤ 2 and determinant $= 0$; cannot be positive definite for any a and c.

28 (a) The eigenvalues of $\lambda_1 I - S$ are $\lambda_1 - \lambda_1, \lambda_1 - \lambda_2, \ldots, \lambda_1 - \lambda_n$. Those are ≥ 0; $\lambda_1 I - S$ is semidefinite.

(b) Semidefinite matrices have energy $\boldsymbol{x}^{\mathrm{T}}(\lambda_1 I - S)\,\boldsymbol{x} \geq 0$. Then $\lambda_1 \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} \geq \boldsymbol{x}^{\mathrm{T}} S \boldsymbol{x}$.

(c) Part (b) says $\boldsymbol{x}^{\mathrm{T}} S \boldsymbol{x} / \boldsymbol{x}^{\mathrm{T}} \boldsymbol{x} \leq \lambda_1$ for all $\boldsymbol{x}$. Equality holds at the leading eigenvector with $S\boldsymbol{x} = \lambda_1 \boldsymbol{x}$.

(Note that the maximum is λ_1—the first printing missed the subscript "one").

Problem Set I.8, page 68

1 $(c_1\boldsymbol{v}_1^{\mathrm{T}}+\cdots+c_n\boldsymbol{v}_n^{\mathrm{T}})\,(c_1\boldsymbol{v}_1+\cdots+c_n\boldsymbol{v}_n)=c_1^2+\cdots+c_n^2$ because the $\boldsymbol{v}$'s are orthonormal.

$(c_1\boldsymbol{v}_1^{\mathrm{T}}+\cdots+c_n\boldsymbol{v}_n^{\mathrm{T}})\,S\,(c_1\boldsymbol{v}_1+\cdots+c_n\boldsymbol{v}_n)=(\qquad\qquad\qquad)\,(c_1\lambda_1\boldsymbol{v}_1+\cdots+c_n\lambda_n\boldsymbol{v}_n)$
$=c_1^2\lambda_1+\cdots+c_n^2\lambda_n$.

2 Remember that $\lambda_1\geq\lambda_2\geq\ldots\geq\lambda_n$. Then $\lambda_1c_1^2+\cdots+\lambda_nc_n^2\leq(c_1^2+\cdots+c_n^2)$.

Therefore the ratio $R(\boldsymbol{x})$ is $\leq\lambda_1$. It equals λ_1 when $\boldsymbol{x}=\boldsymbol{v}_1$.

3 Notice that $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{v}_1=(c_1\boldsymbol{v}_1^{\mathrm{T}}+\cdots+c_n\boldsymbol{v}_n^{\mathrm{T}})\,\boldsymbol{v}_1=c_1$. Then $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{v}_1=0$ means $c_1=0$.

Now $R(\boldsymbol{x})=\dfrac{\lambda_2c_2^2+\cdots+\lambda_nc_n^2}{c_2^2+\cdots+c_n^2}$ is a maximum when $\boldsymbol{x}=\boldsymbol{v}_2$ and $c_2=1$ and other c's $=0$.

4 The maximum of $R(\boldsymbol{x})=\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}/\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}$ is λ_3 when $\boldsymbol{x}$ is restricted by $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{v}_1=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{v}_2=0$.

5 If $A=U\Sigma V^{\mathrm{T}}$ then $A^{\mathrm{T}}=V\Sigma^{\mathrm{T}}U^{\mathrm{T}}$ (singular vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ are reversed but the numbers $\sigma_1,\ldots,\sigma_r$ do not change. Then $A\boldsymbol{v}=\sigma\boldsymbol{u}$ and $A^{\mathrm{T}}\boldsymbol{u}=\sigma\boldsymbol{v}$ for each pair of singular vectors.

For example $A=\begin{bmatrix}0&5\\0&0\end{bmatrix}$ has $\sigma_1=5$ and so does $A^{\mathrm{T}}=\begin{bmatrix}0&0\\5&0\end{bmatrix}$. But $||A\boldsymbol{x}||\neq||A^{\mathrm{T}}\boldsymbol{x}||$ for most $\boldsymbol{x}$.

6 Exchange $\boldsymbol{u}$'s and $\boldsymbol{v}$'s (and keep $\sigma=\sqrt{45}$ and $\sigma=\sqrt{5}$) in equation (12) $=$ the SVD of $\begin{bmatrix}3&0\\4&5\end{bmatrix}$.

7 This question should have told us which matrix norm to use! If we use $||A||=\sigma_1$ then removing $\sigma_1\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}}$ will leave the norm as σ_2. If we use the Frobenius norm $(\sigma_1^2+\cdots+\sigma_r^2)^{1/2}$, then removing $\sigma_1\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}}$ will leave $(\sigma_2^2+\cdots+\sigma_r^2)^{1/2}$.

8 $$\begin{bmatrix}0&2&0\\0&0&3\\0&0&0\end{bmatrix}\begin{bmatrix}0&1&0\\1&0&0\\0&0&1\end{bmatrix}\begin{bmatrix}3&0&0\\0&2&0\\0&0&0\end{bmatrix}\begin{bmatrix}0&0&1\\0&1&0\\1&0&0\end{bmatrix}=U\Sigma V^{\mathrm{T}}$$

9 (*Correction to first printing*) Remove both factors $\frac{1}{2}$ that multiply $\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}$. Then maximizing $\boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x}$ with $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} = 1$ is the same as maximizing their ratio $R(\boldsymbol{x})$.

Now the gradient of $\boldsymbol{L} = \boldsymbol{x}^{\mathrm{T}}S\boldsymbol{x} + \lambda(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} - 1)$ is $2S\boldsymbol{x} - 2\lambda\boldsymbol{x}$. This gives *gradient* $= 0$ at all eigenvectors $\boldsymbol{v}_1$ to $\boldsymbol{v}_n$. Testing $R(\boldsymbol{x})$ at each eigenvector gives $R(\boldsymbol{v}_k) = \lambda_k$ so $\boldsymbol{x} = \boldsymbol{v}_1$ maximizes $R(\boldsymbol{x})$.

10 If you remove columns of a matrix, this cannot increase the norm. Reason: We still have norm $=$ max $||A\boldsymbol{v}||/||\boldsymbol{v}||$ but we are only keeping the $\boldsymbol{v}$'s with zeros in the positions corresponding to removed columns. So the maximum can only move down and never up.

Then removing columns of the transpose (rows of the original matrix) can only reduce the norm further. So a submatrix of A cannot have larger norm than A.

11 The trace of $S = \begin{bmatrix} 0 & A & ; & A^{\mathrm{T}} & 0 \end{bmatrix}$ is zero. The eigenvalues of S come in plus-minus pairs so they add to zero. If $A = \text{diag}\,(1, 2, \ldots, n)$ is diagonal, these $2n$ eigenvalues of S are 1 to n and -1 to $-n$. The $2n$ eigenvectors of S have 1 in positions 1 to n with all $+1$ or all -1 in positions $n+1$ to $2n$.

12 $A = \begin{bmatrix} 2 & 4 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \end{bmatrix}$ means that $A = \dfrac{\begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix}}{\sqrt{5}} \begin{bmatrix} 5 & 0 \\ 0 & 0 \end{bmatrix} \dfrac{\begin{bmatrix} 1 & 2 \\ 2 & -1 \end{bmatrix}}{\sqrt{5}} = U\Sigma V^{\mathrm{T}}$.

13 The homemade proof depends on this step: If Λ is diagonal and $\Sigma\Lambda = \Lambda\Sigma$ then Σ is also diagonal. That step fails when $\Lambda = I$ because $\Sigma I = I\Sigma$ for all Σ. The step fails unless the numbers $\lambda_1, \ldots, \lambda_n$ are all different (which is usually true—but not always, and we want a proof that always works).

Note: If $\lambda_1 \neq \lambda_2$ then comparing the $(1, 2)$ entries of $\Sigma\Lambda$ and $\Lambda\Sigma$ gives $\lambda_2\sigma_{12} = \lambda_1\sigma_{12}$ which forces $\sigma_{12} = 0$. Similarly, all the other off-diagonal σ's will be zero. Repeated eigenvalues $\lambda_1 = \lambda_2$ or singular values always bring extra steps.

14 For a 2 by 3 matrix $A = U\Sigma V^{\mathrm{T}}$, U has 1 parameter (angle θ) and Σ has 2 parameters (σ_1 and σ_2) and V has 3 parameters (3 angles like roll, pitch, and yaw for an aircraft in 3D flight). Total 6 parameters in $U\Sigma V^{\mathrm{T}}$ agrees with 6 in the 2 by 3 matrix A.

15 For 3 by 3 matrices, U and Σ and V have 3 parameters each. For 4 by 4, Σ has 4 singular values and U and V involve 6 angles each: $6+4+6=16$ parameters in A. (*See also the last Appendix*.)

16 4 numbers give a direction in $\mathbf{R}^5$. A unit vector orthogonal to that direction has 3 parameters. The remaining columns of Q have $2,1,0$ parameters (not counting $+/-$ decisions). Total $4+3+2+1+0=10$ parameters in Q.

17 If $A^{\mathrm{T}}A\boldsymbol{v}=\lambda\boldsymbol{v}$ with $\lambda\neq 0$, *multiply by* A: $(AA^{\mathrm{T}})A\boldsymbol{v}=\lambda A\boldsymbol{v}$ with eigenvector $A\boldsymbol{v}$.

18 $A=U\Sigma V^{\mathrm{T}}$ gives $\boldsymbol{A^{-1}=V\Sigma^{-1}U^{\mathrm{T}}}$ when A is invertible. The singular values of $A^{\mathrm{T}}A$ are $\sigma_1^2,\ldots,\sigma_r^2$ (squares of singular values of A).

19 (Correction to 1st printing: Change S to A: *not symmetric* !) If A has orthogonal columns of lengths $2,3,4$ then $A^{\mathrm{T}}A=\text{diag}\,(4,9,16)$ and $\Sigma=\text{diag}\,(2,3,4)$. We can choose $V=$ identity matrix and $U=A\Sigma^{-1}$ has orthogonal *unit vectors*: the original columns divided by $2,3,4$.

20 This matrix has **TO DO**

21 We know that $AA^{\mathrm{T}}A=(U\Sigma V^{\mathrm{T}})\,(V\Sigma^{\mathrm{T}}U^{\mathrm{T}})\,(U\Sigma V^{\mathrm{T}})=U\,(\Sigma\Sigma^{\mathrm{T}}\Sigma)\,V^{\mathrm{T}}$. So the singular values from $\Sigma\Sigma^{\mathrm{T}}\Sigma$ are σ_1^3 to σ_r^3.

22 To go from the reduced form $AV_r=U_r\Sigma_r$ to $A=U_r\Sigma_rV_r^{\mathrm{T}}$, we cannot just multiply both sides by V_r^{T} (Since V_r only has r columns and rank r, possibly a small number, and then $V_rV_r^{\mathrm{T}}$ is not the identity matrix). But the result $A=U_r\Sigma_rV_r^{\mathrm{T}}$ is still correct, since both sides give the zero vector when they multiply the basis vectors $\boldsymbol{v}_{r+1},\ldots,\boldsymbol{v}_n$ in the nullspace of A.

23 This problem is solved in the final appendix of the book. Note for $r=1$ those rank-one matrices have $m+n-1$ free parameters: vectors $\boldsymbol{u}_1$ and $\boldsymbol{v}_1$ have $m+n$ parameters but there is freedom to make one of them a unit vector: $A=(\boldsymbol{u}_1/||\boldsymbol{u}_1||)\,(||\boldsymbol{u}_1||\boldsymbol{v}^{\mathrm{T}})$.

Problem Set I.9, page 80

1 The singular values of $A - A_k$ are $\sigma_{k+1} \geq \sigma_{k+2} \geq \ldots \geq \sigma_r$ (the smallest $r-k$ singular values of A).

2 The closest rank 1 approximations are $A = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ $\quad A = \begin{bmatrix} 0 & 3 \\ 0 & 0 \end{bmatrix}$ $\quad A = \dfrac{3}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$

3 Since this A is orthogonal, its singular values are $\sigma_1 = \sigma_2 = 1$. So we cannot reduce its spectral norm σ_{max} by subtracting a rank-one matrix. On the other hand, we can reduce its Frobenius norm from $||A||_F = \sqrt{2}$ to $||A - \boldsymbol{u}_1\sigma_1\boldsymbol{v}_1^{\mathrm{T}}||_F = \sqrt{1}$.

4 $A - A_1 = \dfrac{1}{2}\begin{bmatrix} 3 & -3 \\ -1 & 1 \end{bmatrix}$ has $||A - A_1||_\infty =$ max row sum $= \mathbf{3}$. But in this "∞ norm" (which is not determined by the singular values) we can find a rank-one matrix B that is closer to A than A_1 is.

$B = \begin{bmatrix} 1 & .75 \\ 4 & 3 \end{bmatrix}$ has $A - B = \begin{bmatrix} 2 & -.75 \\ 0 & 2 \end{bmatrix}$ and $||A - B||_\infty = 2.75$.

5 If $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ then $QA = \begin{bmatrix} \cos\theta & 0 \\ \sin\theta & 0 \end{bmatrix}$. Those matrices have $||A||_\infty = 1$ different from $||QA||_\infty = |\cos\theta|$.

6 $S = Q\Lambda Q^{\mathrm{T}} = \lambda_1\boldsymbol{q}_1\boldsymbol{q}_1^{\mathrm{T}} + \cdots$ is the eigenvalue decomposition and *also the singular value decomposition* of S. So the Eckart-Young theorem applied to $\lambda_1\boldsymbol{q}_1\boldsymbol{q}_1^{\mathrm{T}}$ is the nearest rank-one matrix.

7 Express $E = ||A - CR||_F^2$ as $E = \displaystyle\sum_{i,j}(A_{ij} - \sum_k C_{ik}R_{kj})^2$. Take the derivative with respect to each particular C_{IK}.

$$\frac{\partial E}{\partial C_{IK}} = 2\sum_j \left(A_{Ij} - C_{IK}R_{Kj}\right) R_{Kj}$$

The $(1,1)$ entry of $A - CR$ is $a_{11} - c_{11}r_{11} - c_{12}r_{21}$. The $(1,1)$ entry of $A - (C + \Delta C)\,R$ is $a_{11} - c_{11}r_{11} - c_{12}r_{21} - \Delta c_{11}r_{11} - \Delta c_{12}r_{21}$. **TO COMPLETE**

Squaring and subtracting, the leading terms (first-order) are $2(a_{11}-c_{11}r_{11}-c_{12}r_{21})\,(\Delta c_{11}r_{11}+\Delta c_{12}r_{12})$.

8 $||A-A_1||_2=\sigma_2(A)$ and $||A-A_2||_2=\sigma_3(A)$. (The 2-norm for a matrix is its largest singular value.) So those norms are equal when $\sigma_2=\sigma_3$.

9 Our matrix has 1's below the parabola $y=1-x^2$ and 0's above that parabola. The parabola has slope $dy/dx=-2x=-1$ where $x=\frac{1}{2}$ and $y=\frac{3}{4}$. Remove the rectangle (filled with 1's and therefore rank $=1$) below $y=\frac{3}{4}$ and to the left of $x=\frac{1}{2}$. Above that rectangle, between $y=\frac{3}{4}$ and $y=1$, the rows of A are independent. Beyond that rectangle, between $x=\frac{1}{2}$ and $x=1$, the columns of A are independent. Since $\frac{1}{4}+\frac{1}{2}=\frac{3}{4}$, the rank of A is approximately $\frac{3}{4}N$.

10 A is invertible so $A^{-1}=V\Sigma^{-1}U^{\mathrm{T}}$ has singular values $1/\sigma_1$ and $1/\sigma_2$. Then $||A^{-1}||_2=$ max singular value $=1/\sigma_2$. And $||A^{-1}||_F^2=(1/\sigma_1)^2+(1/\sigma_2)^2$.

Problem Set I.10, page 87

1 $H = M^{-1/2}SM^{-1/2} = \begin{bmatrix} 1 & 0 \\ 0 & 1/2 \end{bmatrix} \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1/2 \end{bmatrix} = \begin{bmatrix} 5 & 4/2 \\ 4/2 & 5/4 \end{bmatrix}$

$\det(S - \lambda M) = \det \begin{bmatrix} 5-\lambda & 4 \\ 4 & 5-4\lambda \end{bmatrix} = 4\lambda^2 - 25\lambda + 9 = 0$

$\det(H - \lambda I) = \det \begin{bmatrix} 5-\lambda & 2 \\ 2 & \frac{5}{4}-\lambda \end{bmatrix} = \lambda^2 - \frac{25}{4}\lambda + \frac{9}{4} = 0$

By the quadratic formula, $d = \dfrac{25 \pm \sqrt{25^2 - 144}}{8} = \dfrac{25 \pm \sqrt{481}}{8}$

The first equation agrees with the second equation (times 4). The eigenvectors will be too complicated for convenient computation by hand.

2 **TO DO**

Problem Set I.11, page 96

1 $||\boldsymbol{v}||_2^2 = v_1^2 + \cdots + v_n^2 \leq (\max|v_i|)\,(|v_1| + \cdots + |v_n|) = ||v||_\infty ||v||_1$

2 (Length)^2 is never negative. We have to simplify that (length)^2:

$$\left(\boldsymbol{v} - \frac{\boldsymbol{v}^{\mathrm{T}}\boldsymbol{w}}{\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w}}\,\boldsymbol{w}\right)^{\mathrm{T}}\left(\boldsymbol{v} - \frac{\boldsymbol{v}^{\mathrm{T}}\boldsymbol{w}}{\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w}}\,\boldsymbol{w}\right) = \boldsymbol{v}^{\mathrm{T}}\boldsymbol{v} - 2\frac{(\boldsymbol{v}^{\mathrm{T}}\boldsymbol{w})^2}{\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w}} + \frac{(\boldsymbol{v}^{\mathrm{T}}\boldsymbol{w})^2}{\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w}} = \boldsymbol{v}^{\mathrm{T}}\boldsymbol{v} - \frac{(\boldsymbol{v}^{\mathrm{T}}\boldsymbol{w})^2}{\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w}} \geq 0.$$

Multiply by $\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w}$.

3 $||\boldsymbol{v}||_2^2 = v_1^2 + \cdots + v_n^2 \leq n\ \max|v_i|^2$ so $||\boldsymbol{v}||_2 \leq \sqrt{n}\ \max|v_i|$.

For the second part, choose $\boldsymbol{w} = (1, 1, \ldots, 1)$ and use Cauchy-Schwarz:

$||\boldsymbol{v}||_1 = |v_1|w_1 + \cdots + |v_n|w_n \leq ||\boldsymbol{v}||_2\, ||w||_2 = \sqrt{n}\, ||\boldsymbol{v}||_2$

4 For $p = 1$ and $q = \infty$, Hölder's inequality says that

$|\boldsymbol{v}^{\mathrm{T}}\boldsymbol{w}| \leq ||\boldsymbol{v}||_1\, ||\boldsymbol{w}||_\infty = (|v_1| + \cdots + |v_n|)\ \max|w_i|$

Problem Set I.12, page 109

1 The v_1 derivative of $(a-v_1u_1)^2+(b-v_1u_2)^2$ is $-2u_1(a-v_1u_1)-2u_2(b-v_1u_2)=0$. Dividing by 2 gives $(u_1^2+u_2)^2v_1 = u_1a+u_2b$. In II.2, this will be the normal equation for the best solution v_1 to the 1D least squares problem $\boldsymbol{u}v_1 = \boldsymbol{a}_1$.

2 Same problem as $\mathbf{1}$, stated in vector notation.

3 This is the same question but now for the second component v_2. Together with $\mathbf{1}$, the combined problem is to find the minimizing numbers (v_1, v_2) for $||\boldsymbol{a} - \boldsymbol{v}\boldsymbol{u}||^2$ when $\boldsymbol{u}$ is fixed.

4 The combined problem when U is fixed is to choose V to minimize $||A - UV||_F^2$. The best V solves $(U^{\mathrm{T}}U)V = U^{\mathrm{T}}A$.

5 This alternating algorithm is important! Here the matrices are small and convergence can be tested and seen computationally.

6 Rank 1 requires $A = \boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}$. All columns of A must be multiples of one nonzero column. (Then all rows will automatically be multiples of one nonzero row.)

7 For the fibers of T in each of the three directions, all slices must be in multiples of one nonzero fiber. (**Question**: If this holds in two directions, does it automatically hold in the third direction?)

8 **TO DO**

9 (a) The sum of all row sums must equal the sum of all column sums.

(b) In each separate direction, add the totals for all slices in that direction. For each direction, the sum of those totals must be the total sum of all entries in the tensor.

10 **TO DO**

11 **TO DO**