Numerical Linear Algebra
Solution of Exercise Problems

Yan Zeng

Version 0.1.1, last revised on 2009-09-01.

Abstract

This is a solution manual of the textbook *Numerical Linear Algebra*, by Lloyd N. Trefethen and David Bau Ⅲ (SIAM, 1997). This version omits Exercise 9.3, 10.4.

Contents

1 Matrix-Vector Multiplication

1.1. (a)

Proof. The basic principle applied here is the following: if A is a rectangular matrix, then any elementary row operation on A may be carried out by multiplying A from the left by the corresponding elementary matrix; any elementary column operation on A may be carried out by multiplying A from the right by the corresponding elementary matrix. See, for example, Munkres [2] §2, Theorem 2.1 for details.

$$\begin{bmatrix}1 & -1 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & -1 & 1 & 0\\ 0 & -1 & 0 & 1\end{bmatrix}\begin{bmatrix}1 & 0 & 1 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{bmatrix}\begin{bmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & \frac{1}{2} & 0\\ 0 & 0 & 0 & 1\end{bmatrix} B \begin{bmatrix}1 & 0 & 0 & 0\\ 0 & 2 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{bmatrix}\begin{bmatrix}0 & 0 & 0 & 1\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 1 & 0 & 0 & 0\end{bmatrix}\begin{bmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 1\\ 0 & 0 & 0 & 0\end{bmatrix}\begin{bmatrix}0 & 0 & 0\\ 1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}$$

□

(b)

Proof.

$$A = \begin{bmatrix}1 & -1 & \frac{1}{2} & 0\\ 0 & 1 & 0 & 0\\ 0 & -1 & \frac{1}{2} & 0\\ 0 & -1 & 0 & 1\end{bmatrix}, \quad C = \begin{bmatrix}0 & 0 & 0\\ 2 & 0 & 0\\ 0 & 1 & 1\\ 0 & 0 & 0\end{bmatrix}.$$

□

1.2. Not sure if I understand the problem correctly. Anyway, here's the solution.

(a)

Proof. We have

$$\begin{bmatrix}f_1\\ f_2\\ f_3\\ f_4\end{bmatrix} = \begin{bmatrix}-k_{12} & k_{12} & 0 & 0\\ 0 & -k_{23} & k_{23} & 0\\ 0 & 0 & -k_{34} & k_{34}\\ 0 & 0 & 0 & 0\end{bmatrix}\begin{bmatrix}x_1\\ x_2\\ x_3\\ x_4\end{bmatrix} - \begin{bmatrix}k_{12}l_{12}\\ k_{23}l_{23}\\ k_{34}l_{34}\\ 0\end{bmatrix}.$$

□

(2)

Proof. The entries of K are spring constants. By Hooke's law (force = spring constant × displacement), the dimension of the entries of K is $\frac{\text{force}}{\text{distance}} = \frac{\text{mass}}{\text{time}^2}$. □

(3)

Proof. $\left[\frac{\text{mass}}{\text{time}^2}\right]^4$. □

(4)

Proof. $K = 1000K'$ and $\det(K) = 10^{12} \cdot \det(K)$. □

1.3.

Proof. We write $I_{m\times m}$ and R^{-1} in the column form: $I_{m\times m} = [e_1, e_2, \cdots, e_m]$, $R^{-1} = [a_1, a_2, \cdots, a_m]$. Suppose $R = (r_{ij})$ has the form

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ 0 & r_{22} & \cdots & r_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & r_{mm} \end{bmatrix}.$$

Then $I_{m\times m} = R^{-1}R$ can be written as

$$[e_1, e_2, \cdots, e_m] = [a_1, a_2, \cdots, a_m] \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ 0 & r_{22} & \cdots & r_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & r_{mm} \end{bmatrix}.$$

Since R is nonsingular and det(R)=$\prod_{i=1}^m r_{ii}$, we conclude the diagonal entries r_{ii} are non-zero.

To show R^{-1} is necessarily upper-triangular, we work by induction. To begin with, we have $e_1 = r_{11}a_1$. So $a_1 = r_{11}^{-1}e_1$ has zero entries except the first one. For convenience, we denote by $\mathbb{C}^m(k)$ the column space $\{v = \begin{bmatrix} v_1 \\ v_2 \\ \cdots \\ v_m \end{bmatrix} \in \mathbb{C}^m : v_i = 0 \text{ for } i > k\}$. Then $\mathbb{C}^m(1) \subset \mathbb{C}^m(2) \subset \cdots \subset \mathbb{C}^m(m) = \mathbb{C}^m$ and each $\mathbb{C}^m(k)$ is a linear subspace of $\mathbb{C}^m$. We have shown $a_1 \in \mathbb{C}^m(1)$. Assume for any $k \le i$, $a_k \in \mathbb{C}^m(k)$. Then by $I_{m\times m} = RR^{-1}$, we have

$$e_{i+1} = \sum_{k=1}^m a_k r_{k(i+1)} = \sum_{k=1}^i a_k r_{k(i+1)} + a_{i+1} r_{(i+1)(i+1)}.$$

Therefore

$$a_{i+1} = r_{(i+1)(i+1)}^{-1} \left(e_{i+1} - \sum_{k=1}^m a_k r_{k(i+1)} \right) \in \mathbb{C}^m(i+1).$$

By induction, we have proved $a_k \in \mathbb{C}^m(k)$ for $1 \le k \le m$, which is equivalent to R^{-1} being upper-triangular. □

1.4. (a)

Proof. Denote by d the column vector $\begin{bmatrix} d_1 \\ d_2 \\ \cdots \\ d_8 \end{bmatrix}$ and by c the column vector $\begin{bmatrix} c_1 \\ c_2 \\ \cdots \\ c_8 \end{bmatrix}$. Let F be the matrix whose (i,j)-entry is $f_j(i)$. Then the given condition can be rephrased as: for any $d \in \mathbb{C}^8$, we can find $c \in \mathbb{C}^8$ such that $Fc = d$. This means range(F)=$\mathbb{C}^8$. By Theorem 1.3, null(F)=$\{0\}$, which implies the mapping $d \mapsto c$ is one-to-one. □

(b)

Proof. Note $Ad = c$ for any $c \in \mathbb{C}^8$ implies $AFc = c$ for any $c \in \mathbb{C}^8$. So $A^{-1} = F$ and the (i,j)-entry of A^{-1} is $F_{ij} = f_j(i)$. □

2 Orthogonal Vectors and Matrices

2.1.

Proof. If A is upper-triangular, then $A^* = A^{-1}$ is also upper-triangular (Exercise 1.3). So A must be diagonal. The case of lower-triangular can be proved similarly. □

2.2. (a)

Proof.

$$\|x_1 + x_2\|^2 = \sum_{i=1}^{m}(x_{1i} + x_{2i})^2 = \sum_{i=1}^{m}(x_{1i}^2 + x_{2i}^2) + 2\sum_{i=1}^{m} x_{1i}x_{2i} = \|x_1\|^2 + \|x_2\|^2 + 2x_1^* x_2 = \|x_1\|^2 + \|x_2\|^2,$$

where the last equality is due to orthogonality. □

(b)

Proof. Assume $\| \sum_{i=1}^{k} x_i\|^2 = \sum_{i=1}^{k} \|x_i\|^2$ for any $1 \le k \le n$, then

$$\| \sum_{i=1}^{n+1} x_i\|^2 = \| \sum_{i=1}^{n} x_i\|^2 + \|x_{n+1}\|^2 = \sum_{i=1}^{n} \|x_i\|^2 + \|x_{n+1}\|^2 = \sum_{i=1}^{n+1} \|x_i\|^2.$$

By mathematical induction, we conclude the Pythagorean Theorem is true for any $n \in \mathbb{N}$. □

2.3. (a)

Proof. If $Ax = \lambda x$ for some $x \in \mathbb{C} \setminus \{0\}$, then

$$\lambda\|x\|^2 = \lambda x^* x = x^*(\lambda x) = x^* Ax = x^* A^* x = (Ax)^* x = (\lambda x)^* x = \lambda^* x^* x = \lambda^* \|x\|^2.$$

Since $x \neq 0$, it must be the case that $\lambda = \lambda^*$. □

(b)

Proof. Suppose $Ax = \lambda x$ and $Ay = \mu y$ with $\lambda \neq \mu$. Then

$$\lambda y^* x = y^*(Ax) = (y^* A)x = (y^* A^*)x = (Ay)^* x = (\mu y)^* x = \mu y^* x,$$

where we have used the result of part (a) to get $\mu^* = \mu$. Since $\lambda \neq \mu$, we must have $y^* x = 0$, i.e. x and y are orthogonal to each other. □

2.4.

Proof. Suppose A is a unitary matrix, λ is an eigenvalue of A with x a corresponding eigenvector. Then

$$x^* x = x^*(A^* A)x = (Ax)^*(Ax) = (\lambda x)^*(\lambda x) = \|\lambda\|^2 x^* x.$$

Since $x \neq 0$, we must have $\|\lambda\| = 1$. □

2.5.

Remark 1. *The point of skew-Hermitian matrix is that any matrix A can be decomposed into the sum of a Hermitian matrix $\frac{1}{2}(A + A^*)$ and a skew-Hermitian matrix $\frac{1}{2}(A - A^*)$. For more details on skew-Hermitian matrix, see Lax [1], Chapter 8.*

(a)

Proof. We note iS is Hermitian: $(iS)^* = i^* S^* = -i(-S) = iS$. Since λ is an eigenvalue of S if and only if λi is an eigenvalue of iS, by Exercise 2.3 we conclude λ must be pure imaginary. □

(b)

Proof. Suppose x is such that $(I - S)x = 0$. Then $x = Sx$ and

$$x^* x = (Sx)^* x = x^* S^* x = -x^* Sx = -x^* x.$$

So $x^* x = 0$ and x must be zero. This shows $\text{null}(I - S) = \{0\}$. By Theorem 1.3, $I - S$ is nonsingualr. □

(c)

Proof. Define $A = [(I-S)^{-1}(I+S)]^*[(I-S)^{-1}(I+S)]$. Then (note for a nonsingular matrix U, $(U^{-1})^* = (U^*)^{-1}$)

$$\begin{aligned} A &= (I+S)^*(I-S)^{-*}(I-S)^{-1}(I+S) \\ &= (I+S^*)[(I-S)(I-S)^*]^{-1}(I+S) \\ &= (I-S)[(I-S)(I+S)]^{-1}(I+S) \\ &= (I-S)(I-S^2)^{-1}(I+S). \end{aligned}$$

Therefore, $A = A(I-S)(I-S)^{-1} = (I-S)[(I-S^2)^{-1}(I-S^2)](I-S)^{-1} = (I-S)(I-S)^{-1} = I$. This shows the matrix Q is unitary. □

2.6.

Proof. If $u = 0$, the problem is trivial. So without loss of generality, we assume $u \neq 0$.

We first find the necessary and sufficient condition under which A is singular. For necessity, suppose there exists $x \in \mathbb{C} \setminus \{0\}$ such that $Ax = x + uv^*x = 0$. Then $x = -u(v^*x)$ is a scalar multiple of u. So we can assume $x = \alpha u$ form some $\alpha \in \mathbb{C}$. Then the equation $Ax = 0$ beomces

$$\alpha u + u(v^*(\alpha u)) = \alpha u(1 + v^*u) = 0.$$

So we must have $v^*u = -1$. For sufficiency, assume $v^*u = -1$. Then for any $\alpha \in \setminus\{0\}$, $A(\alpha u) = \alpha u + uv^*(\alpha u) = \alpha u(1 + v^*u) = 0$, which implies A is singular. Combined, we conclude A is singular if and only if $v^*u = -1$. In this case, $\text{null}(A) = \{\alpha u : \alpha \in \mathbb{C}\}$, the linear subspace spanned by u.

Now assume A is nonsingular. We shall find out A^{-1}. Write A^{-1} in column form $[x_1, \cdots, x_m]$. Then

$$AA^{-1} = (I + uv^*)[x_1, \cdots, x_m] = [x_1 + uv^*x_1, \cdots, x_m + uv^*x_m] = I = [e_1, \cdots, e_m].$$

So $x_i + uv^*x_i = e_i$ $(1 \leq i \leq m)$ and A^{-1} has the general $[e_1 - \theta_1 u, \cdots, e_m - \theta_m u] = I - u\theta^*$, where $\theta = \begin{bmatrix} \theta_1 \\ \cdots \\ \theta_m \end{bmatrix}$. Note

$$I = AA^{-1} = (I + uv^*)(I - u\theta^*) = I - u\theta^* + uv^* - uv^*u\theta^*,$$

which implies $0 = -u\theta_i + uv_i - u(v^*u)\theta_i$ $(1 \leq i \leq m)$. Solving for θ_i gives $\theta_i = \frac{v_i}{1+v^*u}$. So $A^{-1} = I - \frac{uv^*}{1+v^*u}$. □

2.7.

Proof. Inspired by Cramer's rule, we can easily verify

$$H_{k+1}^{-1} = -\frac{1}{2}(H_k^{-1})^2 \begin{bmatrix} -H_k & -H_k \\ -H_k & H_k \end{bmatrix} = \frac{1}{2}\begin{bmatrix} H_k^{-1} & H_k^{-1} \\ H_k^{-1} & -H_k^{-1} \end{bmatrix}.$$

Assume H_k is a Hadamard matrix and denote by α_k the corresponding constant factor. Then $\alpha_0 = 1$ and we have

$$H_{k+1}^T = \begin{bmatrix} H_k^T & H_k^T \\ H_k^T & -H_k^T \end{bmatrix} = \alpha_k \begin{bmatrix} H_k & H_k^{-1} \\ H_k^{-1} & -H_k^{-1} \end{bmatrix} = 2\alpha_k H_{k+1}^{-1}.$$

By induction, we can conclude H_k $(k \in \mathbb{N})$ is a Hadamard matrix and the corresponding constant factor is $\alpha_k = 2^k$. □

3 Norms

3.1.

Proof. To verify property (1) of (3.1), we note $\|x\|_W = \|Wx\| \geq 0$ is obvious. Also $\|x\|_W = \|Wx\| = 0$ if and only if $Wx = 0$, which is further equivalent to $x = 0$ by the nonsingularity of W. To verify property (2) of (3.1), we note

$$\|x+y\|_W = \|W(x+y)\| = \|Wx + Wy\| \leq \|Wx\| + \|Wy\| = \|x\|_W + \|y\|_W.$$

Finally, $\|\alpha x\|_W = \|W(\alpha x)\| = \|\alpha(Wx)\| = |\alpha|\|Wx\| = |\alpha|\|x\|_W$. □

3.2.

Proof. If λ is an eigenvalue of A, we can find $x \in \mathbb{C}^m \setminus \{0\}$ such that $Ax = \lambda x$. So

$$|\lambda| = \frac{\|Ax\|}{\|x\|} \leq \sup_{y \in \mathbb{C}^m \setminus \{0\}} \frac{\|Ay\|}{\|y\|} = \|A\|.$$

□

3.3. (a)

Proof. Assume $|x_{i_0}| = \max_{1 \leq i \leq m} |x_i|$. Then $\|x\|_\infty = |x_{i_0}| \leq \sqrt{\sum_{i=1}^m |x_i|^2} = \|x\|_2$. For $x = \alpha e_i$ ($\alpha \in \mathbb{C}, 1 \leq i \leq m$), the equality holds. □

(b)

Proof. Assume $|x_{i_0}| = \max_{1 \leq i \leq m} |x_i|$. Then $\|x\|_2 = \sqrt{\sum_{i=1}^m |x_i|^2} \leq \sqrt{\sum_{i=1}^m |x_{i_0}|^2} = \sqrt{m}|x_{i_0}| = \sqrt{m}\|x\|_\infty$. For $x = \alpha \begin{bmatrix} 1 \\ 1 \\ \cdots \\ 1 \end{bmatrix}$ ($\alpha \in \mathbb{C}$), the equality holds. □

(c)

Proof. For any $x \in \mathbb{C}^n \setminus \{0\}$, by part (a) and (b),

$$\frac{\|Ax\|_\infty}{\|x\|_\infty} \leq \frac{\|Ax\|_2}{\frac{1}{\sqrt{n}}\|x\|_2} = \sqrt{n}\frac{\|Ax\|_2}{\|x\|_2}.$$

Take supremum on both sides, we have $\|A\|_\infty \leq \sqrt{n}\|A\|_2$. □

(d)

Proof. For any $x \in \mathbb{C}^n \setminus \{0\}$, by part (a) and (b),

$$\frac{\|Ax\|_2}{\|x\|_2} \leq \frac{\sqrt{m}\|Ax\|_\infty}{\|x\|_\infty}.$$

Take supremum on both sides, we have $\|A\|_2 \leq \sqrt{m}\|A\|_\infty$. □

3.4. (a)

Proof. If A is multiplied from the right by identity matrix with the i-th column removed, the result is the matrix obtained by removing the i-th column of A. Similarly, if A is multiplied from the left by the identity matrix with the i-th row removed, the result is the matrix obtained by removing the i-th row of A. To obtain B, we can multiply A from the left and the right by the appropriate identity matrices with certain rows/columns removed. □

(b)

Proof. By inequality (3.14), it suffices to prove any matrix obtained by removing a column or a row from an identity matrix has p-norm less than or equal to 1. This is easy to see by straightforward computation and we omit the details. □

3.5.

Proof. Assume $u \in \mathbb{C}^m$ and $v \in \mathbb{C}^n$. If $E = uv^*$, for any $x \in \mathbb{C}^n \setminus \{0\}$,

$$\frac{\|Ex\|_2}{\|x\|_2} = \frac{\|uv^*x\|_2}{\|x\|_2} = \frac{|v^*x|\|u\|_2}{\|x\|_2} \leq \|u\|_2\|v\|_2,$$

where the last inequality is by Cauchy-Schwarz inequality. This shows $\|E\|_2 \leq \|u\|_2\|v\|_2$. By letting $x = v$ in the above reasoning, we can show the equality holds, i.e. $\|E\|_2 = \|u\|_2\|v\|_2$.

Moreover, we note $E_{ij} = u_i v_j$. Therefore

$$\|E\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |E_{ij}|^2} = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |u_i|^2|v_j|^2} = \sqrt{(\sum_{i=1}^m |u_i|^2)(\sum_{j=1}^n |v_j|^2)} = \|u\|_F\|v\|_F.$$

□

3.6. (a)

Proof. $\|x\|' \geq 0$ is straightforward. Note $\|x\|' \geq |(\frac{x}{\|x\|})^*x| = \|x\|$. So $\|x\|' = 0$ if and only if $x = 0$. Also,

$$\|x_1 + x_2\|' = \sup_{\|y\|=1} |y^*(x_1 + x_2)| \leq \sup_{\|y\|=1} (|y^*x_1| + |y^*x_2|) \leq \sup_{\|y\|=1} |y^*x_1| + \sup_{\|y\|=1} |y^*x_2| = \|x_1\|' + \|x_2\|'.$$

Finally, for any $\alpha \in \mathbb{C}$, $\|\alpha x\|' = \sup_{\|y\|=1} |y^*(\alpha x)| = \sup_{\|y\|=1} |\alpha||y^*x| = |\alpha|\|x\|'$. □

(b)

Proof. We follow the hint. For the given x, we can find a nonzero $z_0 \in \mathbb{C}^m$ such that $|z_0^*x| = \|z_0\|'\|x\|$. Define $z = e^{i\theta}z_0/\|z_0\|'$, where $\theta = \arg(z_0x)$. Then $\|z\|' = 1$ and

$$z^*x = e^{-i\theta}\frac{z_0^*}{\|z_0\|'}x = \frac{|z_0^*x|}{\|z_0\|'} = \|x\| = 1.$$

Therefore, $Bx = (yz^*)x = y(z^*x) = y$. Furthermore, we note

$$\|B\| = \sup_{\|w\|=1} \|Bw\| = \sup_{\|w\|=1} \|yz^*w\| = \sup_{\|w\|=1} |z^*w|\|y\| = \sup_{\|w\|=1} |w^*z| = \|z\|' = 1.$$

Combined, we conclude for the above defined z^* and $B = yz^*$, we have $Bx = y$ and $\|B\| = 1$.

Remark 2. *Note the desired properties of B, $Bx = y$ and $\|B\| = 1$, are equivalent to (a) $z^*x = 1 = \|x\|$; (b) $\|z\|' = 1$. This immediately reminds us of the following well-known corollary of Han-Banach Theorem: Given an element x in a normed linear space X, there exists a continuous linear functional l on X, such that $l(x) = \|x\|$ and $\|l\| = 1$ (Yosida [3], page 108). Note $l(x) = l(\sum_{i=1}^m x_ie_i) = [l(e_1), \cdots, l(e_m)]\begin{bmatrix} x_1 \\ \cdots \\ x_m \end{bmatrix}$. We can therefore define z such that $z^* = [l(e_1), \cdots, l(e_m)]$. Then $l(w) = z^*w$ for any $w \in \mathbb{C}^m$ (in particular, $z^*x = l(x) = \|x\|$). Consequently, $\|z\|' = \|l\| = 1$. This proves the hint.*

□

4 The Singular Value Decomposition

Remark 3. *In the proof of Theorem 4.1, the part on uniqueness is a bit confusing, esp. the following sentence: "Since $\|A\|_2 = \sigma_1$, $\|Av_2\|_2 \le \sigma_1$; but this must be an equality, for otherwise, since $w = v_1 c + v_2 s$ for some constants c and s with $|c|^2 + |s|^2 = 1$, we would have $\|Aw\|_2 < \sigma_1$."*

I don't see why "we would have $\|Aw\|_2 < \sigma_1$." An alternative proof for the uniqueness goes as follows: The uniqueness of the singular values $\{\sigma_j\}$ are given by the observation that $\{\sigma_j^2\}$ are eigenvalues of AA^* or A^*A. Note $\{u_j\}$ are eigenvectors of AA^*, with σ_j^2's the corresponding eigenvalues. So when σ_j's are distinct, the eigenvalues of AA^* are also distinct, and the eigenspace of each σ_j^2 (for AA^*) is of dimension one. Due to normalization, each u_j is unique up to a complex sign. The result for $\{v_j\}$ can be proven similarly.

4.1. To find the SVD of a matrix A, we note that if $A = U\Sigma V^*$ is the SVD of A, then $AA^* = U(\Sigma\Sigma^*)U^*$, or equivalently, $(AA^*)U = U(\Sigma\Sigma^*)$. Writing U in its column form $[u_1, u_2, \cdots, u_m]$, the equation becomes

$$(AA^*)[u_1, u_2, \cdots, u_m] = \begin{cases} [\sigma_1^2 u_1, \sigma_2^2 u_2, \cdots, \sigma_m^2 u_m] & \text{if } m \le n \\ [\sigma_1^2 u_1, \cdots, \sigma_n^2 u_n, 0, \cdots, 0] & \text{if } m > n. \end{cases}$$

This shows for $p = \min\{m, n\}$, σ_1^2, σ_2^2, $\cdots$, σ_p^2 are eigenvalues of AA^* and u_1, u_2, $\cdots$, u_p are the corresponding eigenvectors. This gives us a way to find the singular values and U: if $m \le n$, U consists of the eigenvectors of σ_1^2, σ_2^2, $\cdots$, σ_m^2; if $m > n$, the first n columns of U are eigenvectors of σ_1^2, σ_2^2, $\cdots$, σ_n^2 and the last $m - n$ columns of U can be chosen from null(AA^*) so that U becomes unitary. If A is not of full rank, then p needs to be replaced by rank(A) in the above algorithm.

A similar algorithm for V can be found by noting $(A^*A)V = V(\Sigma\Sigma^*)$. Note $\Sigma\Sigma^*$ and $\Sigma^*\Sigma$ have the same nonzero eigenvalues. we don't have to recalculate the eigenvalues. In practice, various tricks often directly yield the SVD without resorting to the above method.

The solution below has been verified by **Mathematica**.[1]

(a)

Proof. $U = I_{2\times 2}$, $\Sigma = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$, $V = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$. □

(b)

Proof. $U = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, $\Sigma = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$, $V = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. □

(c)

Proof. $U = I_{3\times 3}$, $\Sigma = \begin{bmatrix} 2 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$, $V = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. □

(d)

Proof. $U = I_{2\times 2}$, $\Sigma = \begin{bmatrix} \sqrt{2} & 0 \\ 0 & 0 \end{bmatrix}$, $V = \frac{\sqrt{2}}{2}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$. □

(e)

Proof. $U = \frac{\sqrt{2}}{2}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$, $\Sigma = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$, $V = \frac{\sqrt{2}}{2}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$. □

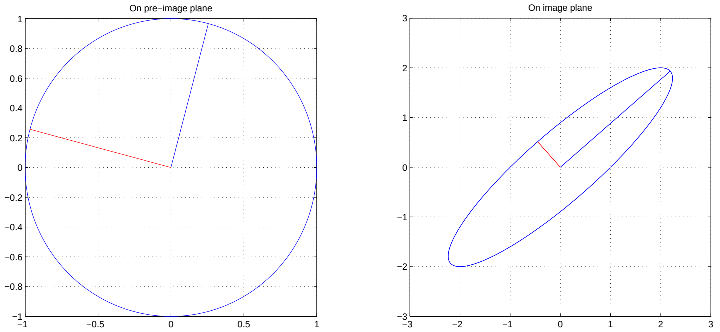
4.2.

[1] Example: $\{u, w, v\} =$ **SingularValueDecomposition**$[\{\{1, 2\}, \{1, 2\}\}]$.

Proof. Suppose $A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$. Then $B = \begin{bmatrix} a_{m1} & a_{m-1,1} & \cdots & a_{11} \\ \cdots & \cdots & \cdots & \cdots \\ a_{mn} & a_{m-1,n} & \cdots & a_{1n} \end{bmatrix}$. We can obtain from B the transpose A^T of A by swapping columns of B. Therefore, the question asked by the problem is reduced to the following two sub-questions: 1) Does a matrix and its transpose have the same singular values? 2) If we multiply a matrix by a unitary matrix, does the resulted matrix have the same singular values as the original one?

In view of the structure of SVD and the uniqueness of singular values, we can say *yes* to both questions. Hence A and B have the same singular values. $\square$

4.3.



Proof. We have the following MATLAB program.

```
function svd_plot(A)

%SVD_PLOT   find the SVD of a real 2-by-2 matrix A. It then
%           plots the right singular vectors in the unit circle on
%           the pre-image plane, as well as the scaled left singular
%           vectors in the appropriate ellipse on the image plane.
%
%           Exercise problem 4.3 of [1].
%
%           [1]
%           Lloyd N. Trefethen and David Bau III. Numerical linear algebra.
%           SIAM, 1997.
%
%           Z.Y., 09/19/2009.

t=0:0.01:(2*pi);

% Create the unit circe on the pre-image plane.
x1 = cos(t); y1 = sin(t);

% Create the ellipse on the image plane.
w = A*[x1; y1]; x2 = w(1,:); y2 = w(2,:);

% Obtain the SVD of A
[U, S, V] = svd(A);

% Create the right and left singular vectors
```

```
v1 = V(:,1); v2 = V(:,2);           % right singular vectors
u1 = U(:,1); u2 = U(:,2);           % left singular vectors
w1 = S(1,1)*u1; w2 = S(2,2)*u2;     % scaled left singular vectors

% Plot the unit circle and the right singular vectors on pre-image plane.
figure(1);
axis([-1.2 1.2 -1.2 1.2]);
plot(x1, y1);
hold on;
line_plot(v1, 'b');
hold on;
line_plot(v2, 'r');
grid on;
title('On pre-image plane');

% Plot the ellipse and the scaled left signualr vectors on image plane.
figure(2);
axis([-3 3 -3 3]);
plot(x2, y2);
hold on;
line_plot(w1,'b');
hold on;
line_plot(w2, 'r');
grid on;
title('On image plane');


end % svd_plot

%%%%%%%%%%%%%%%% Nested function %%%%%%%%%%%%%%%%%%
function line_plot(v, color)

%LINE_PLOT  plots the line segment between the origin and the
%           vector V on the 2-dimensional plane.
%
%           Example: line_plot([1 1], 'r').
%
%           Z.Y., 09/19/2009.

t = 0:0.01:1; x = v(1)*t; y = v(2)*t; plot(x, y, color);

end % line_plot
```

□

4.4.

Proof. Suppose A has SVD $U_1\Sigma_1V_1^*$ and B has SVD $U_2\Sigma_2V_2^*$.

If A and B are unitarily equivalent, there exists a unitary matrix Q, such that $A = QBQ^*$. Then $A = Q(U_2\Sigma_2V_2^*)Q^* = (QU_2)\Sigma_2(QV_2)^*$. So $(QU_2)\Sigma_2(QV_2)^*$ is an SVD of A. By the uniqueness of singular values, $\Sigma_2 = \Sigma_1$. In other words, A and B have the same singular values.

The converse is not necessarily true. As a counter example, note in Problem 4.1 (d), if we can let $A = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$ and $B = \begin{bmatrix} \sqrt{2} & 0 \\ 0 & 0 \end{bmatrix}$, then A and B have the same singular values. But they are not unitarily equivalent: otherwise A would be symmetric, which is clearly not true. □

4.5.

Proof. We give two proofs.

For the first proof, we can examine the proof of Theorem 4.1 and see that if we replace $\mathbb{C}^m$ with $\mathbb{R}^m$ everywhere, the same proof-by-induction works for real A. So when A is real, it has a real SVD.

For the second proof, we note A^*A is now real and symmetric. So it has a complete set of orthogonal (i.e. real and orthonormal) eigenvectors: v_1, v_2, $\cdots$, v_n. Define $V = [v_1, \cdots, v_n]$ and denote by λ_i the eigenvalue corresponding to v_i $(i = 1, \cdots, n)$. Since A^*A is nonnegative, all the eigenvalues are nonnegative. Without loss of generality, we assume $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Let r be the largest index such that $\lambda_r > 0$. Then $\lambda_{r+1} = \cdots = \lambda_n = 0$. Moreover, we note $\text{null}(A^*A) = \text{null}(A)$, since

$$x \in \text{null}(A^*A) \Longrightarrow A^*Ax = 0 \Longrightarrow (Ax)^*(Ax) = x^*A^*Ax = 0 \Longrightarrow Ax = 0 \Longrightarrow x \in \text{null}(A).$$

Coming back to Problem 4.5, we define $\sigma_i = \sqrt{\lambda_i}$ and $u_i = \frac{Av_i}{\sigma_i}$ $(i = 1, \cdots, r)$. Then $\{u_1, \cdots, u_r\}$ is an orthogonal set of vectors:

$$u_i^* u_j = \frac{(Av_i)^*(Av_j)}{\sigma_i\sigma_j} = \frac{v_i^*(A^*A)v_j}{\sqrt{\lambda_i\lambda_j}} = \frac{\lambda_j v_i^* v_j}{\sqrt{\lambda_i\lambda_j}} = \delta_{ij}.$$

We can expand the set $\{u_1, \cdots, u_r\}$ to be an orthogonal basis for $\mathbb{R}^m$: $\{u_1, \cdots, u_r, u_{r+1}, \cdots, u_m\}$. Define $U = [u_1, \cdots, u_m]$. We check U^*AV is a diagonal matrix (recall $\text{null}(A^*A) = \text{null}(A)$):

$$U^*AV = \begin{bmatrix} u_1^* \\ u_2^* \\ \cdots \\ u_m^* \end{bmatrix} A[v_1, \cdots, v_n] = (u_i^*Av_j), \text{ and } u_i^*Av_j = \begin{cases} 0 & \text{if } j > r \\ \sigma_i\delta_{ij} & \text{if } j \leq r. \end{cases}$$

This shows U^*AV is a (rectangular) diagonal matrix Σ. So A has a real SVD $A = U\Sigma V^*$.

Remark 4. *The second proof is actually an existence proof of SVD when A is real. It relies on the diagonalization result of real symmetric matrices.*

□

5 More on the SVD

5.1.

Proof. We note

$$AA^* = \begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 2 & 2 \end{bmatrix} = \begin{bmatrix} 5 & 4 \\ 4 & 4 \end{bmatrix}, \ \det(\lambda I - AA^*) = \det \begin{bmatrix} \lambda - 5 & -4 \\ -4 & \lambda - 4 \end{bmatrix} = \lambda^2 - 9\lambda + 4.$$

So the two eigenvalues of AA^* are $\lambda_{1,2} = \frac{9 \pm \sqrt{65}}{2}$. Therefore the largest and smallest singular values of A are, respectively, $\sigma_{\text{max}}(A) = \left(\frac{9+\sqrt{65}}{2}\right)^{1/2} \approx 2.920809626481889$ and $\sigma_{\text{min}}(A) = \left(\frac{9-\sqrt{65}}{2}\right)^{1/2} \approx 0.684741648982100$. □

5.2.

Proof. Suppose A has SVD $U\Sigma V^*$. For any $\varepsilon > 0$, define $A_\varepsilon = U(\Sigma + \varepsilon I_{m\times n})V^*$, where $I_{ij} = \delta_{ij}$. Since all the diagonal elements of Σ is nonnegative, all the diagonal elements of $(\Sigma + \varepsilon I_{m\times n})$ are positive. Since multiplication by unitary matrices preserves the rank of a matrix, A_ε is of full rank. By Theorem 5.3,

$$\|A - A_\varepsilon\|_2 = \|U(\varepsilon I_{m\times n})V^*\|_2 = \varepsilon.$$

Therefore $\lim_{\varepsilon\to 0} \|A - A_\varepsilon\|_2 = 0$ with A_ε of full rank. This proves the set of full-rank matrices is a dense subset of $\mathbb{C}^{m\times n}$. □

5.3. (a)

Proof. We will derive the expression U from the equation $AA^T = U\Sigma\Sigma U^T$. First of all, we note

$$AA^T = \begin{bmatrix} 125 & 75 \\ 75 & 125 \end{bmatrix} = 25 \begin{bmatrix} 5 & 3 \\ 3 & 5 \end{bmatrix}$$

The characteristic polynomial of $\begin{bmatrix} 5 & 3 \\ 3 & 5 \end{bmatrix}$ is

$$\det\left(\begin{bmatrix} 5-\lambda & 3 \\ 3 & 5-\lambda \end{bmatrix}\right) = \lambda^2 - 10\lambda + 16 = (\lambda - 8)(\lambda - 2).$$

So the eigenvalues of AA^T are $8 \cdot 25 = 200$ and $2 \cdot 25 = 50$. This implies the singular values of A are $10\sqrt{2}$ and $5\sqrt{2}$. Solving the equation

$$\begin{bmatrix} 5 & 3 \\ 3 & 5 \end{bmatrix} U = U \begin{bmatrix} 8 & 0 \\ 0 & 2 \end{bmatrix}$$

gives us the general expression of U:

$$U = \frac{1}{\sqrt{2}} \begin{bmatrix} a & b \\ a & -b \end{bmatrix},$$

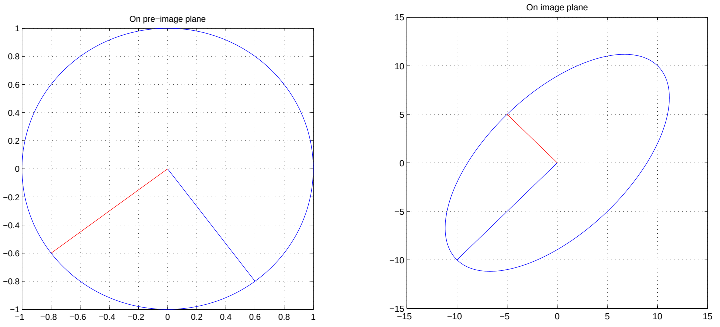
where $a, b \in \{1, -1\}$. Then

$$V = A^T U \Sigma^{-1} = \begin{bmatrix} -2 & -10 \\ 11 & 5 \end{bmatrix} \cdot \frac{1}{\sqrt{2}} \begin{bmatrix} a & b \\ a & -b \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{10\sqrt{2}} & 0 \\ 0 & \frac{1}{5\sqrt{2}} \end{bmatrix} = \begin{bmatrix} -\frac{3}{5}a & \frac{4}{5}b \\ \frac{4}{5}a & \frac{3}{5}b \end{bmatrix}.$$

To let U and V have minimal number of minus signs, we should take $a = b = 1$. □

(b)

Proof. From our calculation in part (a), the singular values of A are $10\sqrt{2}$ and $5\sqrt{2}$. The right singular vectors are $\pm \begin{bmatrix} -3/5 \\ 4/5 \end{bmatrix}$ and $\pm \begin{bmatrix} 4/5 \\ 3/5 \end{bmatrix}$. The left singular vectors are $\pm \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$ and $\pm \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$. The unit ball in $\mathbb{R}^2$ and its image under A, together with the singular vectors are shown in the following graph, plotted by the MATLAB program we used for Exercise Problem 4.3. 4.3. □



(c)

Proof. By Theorem 5.3, $\|A\|_2 = 10\sqrt{2}$, $\|A\|_F = \sqrt{200 + 50} = 5\sqrt{10}$. By Example 3.3, $\|A\|_1 = 16$. By Example 3.4, $\|A\|_\infty = 15$. □

(d)

Proof. Since $A = U\Sigma V^T$ and U, V are unitary,

$$A^{-1} = (U\Sigma V^T)^{-1} = V\Sigma^{-1}U^T = \begin{bmatrix} -\frac{3}{5} & \frac{4}{5} \\ \frac{4}{5} & \frac{3}{5} \end{bmatrix} \begin{bmatrix} \frac{1}{10\sqrt{2}} & 0 \\ 0 & \frac{1}{5\sqrt{2}} \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \frac{1}{100} \begin{bmatrix} 5 & -11 \\ 10 & -2 \end{bmatrix}.$$

□

(e)

Proof. The characteristic polynomial of A is

$$\det \begin{bmatrix} -2-\lambda & 11 \\ -10 & 5-\lambda \end{bmatrix} = \lambda^2 - 3\lambda + 100.$$

So the eigenvalues of A are $\lambda_{1,2} = (3 \pm \sqrt{391}i)/2$. □

(f)

Proof. We note $\det A = -2 \cdot 5 - 11 \cdot (-10) = 100$. $\lambda_1\lambda_2 = \frac{1}{4}(9+391) = 100$. $\sigma_1\sigma_2 = 10\sqrt{2} \cdot 5\sqrt{2} = 100$. □

(g)

Proof. By change-of-variable formula, the area of the llipsoid is $\det(A) = 100$. □

5.4.

Proof. We note

$$\begin{aligned} \begin{bmatrix} 0 & A^* \\ A & 0 \end{bmatrix} &= \begin{bmatrix} 0 & V\Sigma^*U^* \\ U\Sigma V^* & 0 \end{bmatrix} = \begin{bmatrix} 0 & I_{m\times m} \\ I_{m\times m} & 0 \end{bmatrix} \begin{bmatrix} U\Sigma V^* & 0 \\ 0 & V\Sigma^*U^* \end{bmatrix} \\ &= \begin{bmatrix} 0 & I_{m\times m} \\ I_{m\times m} & 0 \end{bmatrix} \begin{bmatrix} U & 0 \\ 0 & V \end{bmatrix} \begin{bmatrix} \Sigma & 0 \\ 0 & \Sigma^* \end{bmatrix} \begin{bmatrix} V^* & 0 \\ 0 & U^* \end{bmatrix}. \end{aligned}$$

The inverse of $\begin{bmatrix} 0 & I_{m\times m} \\ I_{m\times m} & 0 \end{bmatrix}$ is itself and the inverse of $\begin{bmatrix} U & 0 \\ 0 & V \end{bmatrix}$ is $\begin{bmatrix} U^* & 0 \\ 0 & V^* \end{bmatrix}$. So the inverse of

$$\begin{bmatrix} 0 & I_{m\times m} \\ I_{m\times m} & 0 \end{bmatrix} \begin{bmatrix} U & 0 \\ 0 & V \end{bmatrix}$$

is $\begin{bmatrix} U^* & 0 \\ 0 & V^* \end{bmatrix} \begin{bmatrix} 0 & I_{m\times m} \\ I_{m\times m} & 0 \end{bmatrix}$. Note

$$\begin{bmatrix} V^* & 0 \\ 0 & U^* \end{bmatrix} = \begin{bmatrix} 0 & I_{m\times m} \\ I_{m\times m} & 0 \end{bmatrix} \begin{bmatrix} U^* & 0 \\ 0 & V^* \end{bmatrix} \begin{bmatrix} 0 & I_{m\times m} \\ I_{m\times m} & 0 \end{bmatrix}.$$

Therefore, we have

$$\begin{aligned} \begin{bmatrix} 0 & A^* \\ A & 0 \end{bmatrix} &= \left(\begin{bmatrix} 0 & I_{m\times m} \\ I_{m\times m} & 0 \end{bmatrix} \begin{bmatrix} U & 0 \\ 0 & V \end{bmatrix} \right) \left(\begin{bmatrix} \Sigma & 0 \\ 0 & \Sigma^* \end{bmatrix} \begin{bmatrix} 0 & I_{m\times m} \\ I_{m\times m} & 0 \end{bmatrix} \right) \left(\begin{bmatrix} 0 & I_{m\times m} \\ I_{m\times m} & 0 \end{bmatrix} \begin{bmatrix} U & 0 \\ 0 & V \end{bmatrix} \right)^{-1} \\ &= \left(\begin{bmatrix} 0 & I_{m\times m} \\ I_{m\times m} & 0 \end{bmatrix} \begin{bmatrix} U & 0 \\ 0 & V \end{bmatrix} \right) \begin{bmatrix} 0 & \Sigma \\ \Sigma^* & 0 \end{bmatrix} \left(\begin{bmatrix} 0 & I_{m\times m} \\ I_{m\times m} & 0 \end{bmatrix} \begin{bmatrix} U & 0 \\ 0 & V \end{bmatrix} \right)^{-1}. \end{aligned}$$

Note Σ is a diagonal matrix with nonnegative diagonal elements, we must have $\Sigma = \Sigma^*$. It's easy to see

$$\begin{bmatrix} 0 & \Sigma \\ \Sigma & 0 \end{bmatrix} \begin{bmatrix} I_{m\times m} & I_{m\times m} \\ I_{m\times m} & -I_{m\times m} \end{bmatrix} = \begin{bmatrix} I_{m\times m} & I_{m\times m} \\ I_{m\times m} & -I_{m\times m} \end{bmatrix} \begin{bmatrix} \Sigma & 0 \\ 0 & -\Sigma \end{bmatrix}, \begin{bmatrix} I_{m\times m} & I_{m\times m} \\ I_{m\times m} & -I_{m\times m} \end{bmatrix} \begin{bmatrix} I_{m\times m} & I_{m\times m} \\ I_{m\times m} & -I_{m\times m} \end{bmatrix} = 2I_{2m\times 2m}.$$

So the formula

$$\begin{bmatrix} 0 & \Sigma \\ \Sigma & 0 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} I_{m\times m} & I_{m\times m} \\ I_{m\times m} & -I_{m\times m} \end{bmatrix} \cdot \begin{bmatrix} \Sigma & 0 \\ 0 & -\Sigma \end{bmatrix} \cdot \frac{1}{\sqrt{2}} \begin{bmatrix} I_{m\times m} & I_{m\times m} \\ I_{m\times m} & -I_{m\times m} \end{bmatrix}$$

gives the eigenvalue decomposition of $\begin{bmatrix} 0 & \Sigma \\ \Sigma & 0 \end{bmatrix}$. Combined, if we set

$$X = \begin{bmatrix} 0 & I_{m\times m} \\ I_{m\times m} & 0 \end{bmatrix} \begin{bmatrix} U & 0 \\ 0 & V \end{bmatrix} \cdot \frac{1}{\sqrt{2}} \begin{bmatrix} I_{m\times m} & I_{m\times m} \\ I_{m\times m} & -I_{m\times m} \end{bmatrix},$$

then $X \begin{bmatrix} \Sigma & 0 \\ 0 & -\Sigma \end{bmatrix} X^{-1}$ gives the eigenvalue decomposition of $\begin{bmatrix} 0 & A^* \\ A & 0 \end{bmatrix}$. □

6 Projectors

6.1.

Proof. Algebraically, we have by Theorem 6.1 $(I - 2P)^* = I^* - 2P^* = I - 2P$. So

$$(I-2P)^*(I-2P) = [(I-P)-P]^2 = (I-P)^2 - (I-P)P - P(I-P) + P^2 = I - P + P = I,$$

which shows $I - 2P$ is unitary. □

6.2.

Proof. It's easy to see $F^2 = I$. So $E^2 = \frac{1}{4}(I + 2F + F^2) = \frac{I+F}{2} = E$. This shows E is a projector. Note F has the matrix representation

$$F = \begin{pmatrix} 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 1 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 1 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \end{pmatrix}$$

So $F^* = F$ and consequently $E^* = E$. By Theorem (6.1), E is an orthogonal projector. The matrix representation of E is

$$E = \frac{1}{2} \begin{pmatrix} 1 & 0 & \cdots & 0 & 1 \\ 0 & 1 & \cdots & 1 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & 2 & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 1 & \cdots & 1 & 0 \\ 1 & 0 & \cdots & 0 & 1 \end{pmatrix}$$

□

6.3.

Proof. We suppose the SVD of A is $A = U\Sigma V^*$. Then A is of full rank if and only if Σ is of full rank. Since $A^*A = V\Sigma^*\Sigma V^*$, we can easily see A^*A is nonsingular if and only if Σ is of full rank. Therefore, A^*A is nonsingular if and only if A has full rank.

An alternative proof: If A^*A is singular, there exists $x \neq 0$ such that $A^*Ax = 0$, which implies $\|Ax\|_2^2 = x^*A^*Ax = 0$. So A is not of full rank. Conversely, if A is not of full rank, there must exist $x \neq 0$ such that $Ax = 0$. So $A^*Ax = 0$, which implies A^*A is singular. This argument appears in the proof of Theorem 11.1 (pp. 81). □

6.4. (a)

Proof. range(A) is the column space of A, which has an orthonormal basis $a_1 = \begin{pmatrix}1/\sqrt{2}\\0\\1/\sqrt{2}\end{pmatrix}$ and $a_2 = \begin{pmatrix}0\\1\\0\end{pmatrix}$. So the orthogonal projector P onto range(A) can be written as $Px = a_1^*xa_1 + a_2^*xa_2 = a_1a_1^*x + a_2a_2^*x = (a_1, a_2)\begin{pmatrix}a_1^*\\a_2^*\end{pmatrix}x = AA^*x$, for any $x \in \mathbb{C}^3$. Therefore, $P = AA^* = \begin{pmatrix}1/2 & 0 & 1/2\\0 & 1 & 0\\1/2 & 0 & 1/2\end{pmatrix}$ (the derivation so far is just the discussion after formula (6.7)). Consequently, the image under P of the vector $(1, 2, 3)^*$ is $(2, 2, 2)^*$. □

(b)

Proof. By formula (6.13),

$$P = B(B^*B)^{-1}B^* = \frac{1}{6}\begin{pmatrix}5 & 2 & 1\\2 & 2 & -2\\1 & -2 & 5\end{pmatrix}.$$

Accordingly, the image under P of the vector $(1, 2, 3)^*$ is $(2, 0, 2)^*$. □

6.5.

Proof. For any $v \in$ range(P), we have $Pv = v$. So $\|P\|_2 = \max_{x\in C^m\setminus\{0\}} \frac{\|Px\|_2}{\|x\|_2} \geq 1$. Suppose the SVD of P is $U\Sigma V^*$. Then for any $x \in \mathbb{C}^m$, we have

$$\|Px\|_2^2 = x^*P^*Px = x^*V\Sigma^*U^*U\Sigma V^*x = x^*V\Sigma^2V^*x.$$

Therefore

$$\|P\|_2 = \max_{x\in\mathbb{C}^m\setminus\{0\}} \frac{\|Px\|_2}{\|x\|_2} = \max_{x\in\mathbb{C}^m\setminus\{0\}} \frac{\|\Sigma V^*x\|_2}{\|V^*x\|_2} = \max_{x\in\mathbb{C}^m\setminus\{0\}} \frac{\|\Sigma x\|_2}{\|x\|_2} = \|\Sigma\|_2.$$

Therefore, $\|P\|_2 = 1 \iff \|\Sigma\|_2 = 1 \iff \Sigma = I$. But $P^2 = U\Sigma V^*U\Sigma V^* = P = U\Sigma V^*$, which is equivalent to $V^*U\Sigma = I$. So $\Sigma = I \iff U = V$. This is equivalent to P being an orthogonal projector. □

7 QR Factorization

7.1. (a)

Proof. We note the columns of A are already orthogonal, so a reduced QR factorization and a full QR factorization can be obtained by normalizing the columns of A

$$A = \begin{pmatrix}1 & 0\\0 & 1\\1 & 0\end{pmatrix} = \begin{pmatrix}\frac{1}{\sqrt{2}} & 0\\0 & 1\\\frac{1}{\sqrt{2}} & 0\end{pmatrix}\begin{pmatrix}\sqrt{2} & 0\\0 & 1\end{pmatrix} = \begin{pmatrix}\frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}}\\0 & 1 & 0\\\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}}\end{pmatrix}\begin{pmatrix}\sqrt{2} & 0\\0 & 1\\0 & 0\end{pmatrix}$$

□

(b)

Proof. We follow the Gram-Schmidt orthogonalization to produce the QR factorization of $B = (b_1, b_2)$. We note

$$(b_1, b_2) = (q_1, q_2)\begin{pmatrix}r_{11} & r_{12}\\0 & r_{22}\end{pmatrix}$$

So $r_{11} = \|b_1\|_2 = \sqrt{2}$ and $q_1 = b_1/\|b_1\|_2 = (1/\sqrt{2}, 0, 1/\sqrt{2})^T$. $r_{12} = q_1^*b_2 = (1/\sqrt{2}, 0, 1/\sqrt{2})\begin{pmatrix}2\\1\\0\end{pmatrix} = \sqrt{2}$, so

$$b_2 - r_{12}q_1 = \begin{pmatrix}2\\1\\0\end{pmatrix} - \sqrt{2}\begin{pmatrix}1/\sqrt{2}\\0\\1/\sqrt{2}\end{pmatrix} = \begin{pmatrix}1\\1\\-1\end{pmatrix}$$

and $r_{22} = \sqrt{3}$. Consequently, the reduced QR factorization of B is

$$B = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{3} \\ 0 & 1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{3} \end{pmatrix} \begin{pmatrix} \sqrt{2} & \sqrt{2} \\ 0 & \sqrt{3} \end{pmatrix}.$$

To find out the full QR factorization of B, we can set (see, for example, Lax [1], Chapter 7, Exercise 20 (vii))

$$q_3 = q_1 \times q_2 = \begin{pmatrix} -1/\sqrt{6} \\ 2/\sqrt{6} \\ 1/\sqrt{6} \end{pmatrix}.$$

Then

$$B = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{3} & -1/\sqrt{6} \\ 0 & 1/\sqrt{3} & 2/\sqrt{6} \\ 1/\sqrt{2} & -1/\sqrt{3} & 1/\sqrt{6} \end{pmatrix} \begin{pmatrix} \sqrt{2} & \sqrt{2} \\ 0 & \sqrt{3} \\ 0 & 0 \end{pmatrix}.$$

□

7.2.

Proof. In the matrix $\hat{R}$, we have $r_{ij} = 0$ if $i > j$ or $i \leq j$ but i and j are not simultaneously even or odd. □

7.3.

Proof. If A is not invertible, $\det A = 0$ and we have nothing to prove. So without loss of generality, we assume A is of full rank. Let $A = QR$ be the QR factorization of A, then $|\det A| = |\det Q \cdot \det R| = |\prod_{j=1}^m r_{jj}|$. According to the Gram-Schmidt orthogonalization, $r_{ij} = q_i^* a_j (i \neq j)$ and $|r_{jj}| = \|a_j - \sum_{i=1}^{j-1} r_{ij} q_i\|_2$. Therefore

$$|r_{jj}|^2 = \|a_j - \sum_{i=1}^{j-1} r_{ij} q_i\|_2^2 = (a_j^* - \sum_{i=1}^{j-1} r_{ij} q_i^*)(a_j - \sum_{i=1}^{j-1} r_{ij} q_i) = \|a_j\|_2^2 - 2\sum_{i=1}^{j-1} r_{ij}^2 + \sum_{i=1}^{j-1} r_{ij}^2 \leq \|a_j\|_2^2.$$

This implies $|\det A| = |\prod_{j=1}^m r_{jj}| \leq \prod_{j=1}^m \|a_j\|_2$. The geometric interpretation of this inequality for $m = 2$ is that the area of a parallelepiped is at most the product the lengths of its two neighboring sides, where the equality holds if and only if its sides are orthogonal to each other. Higher dimension has similar interpretation. □

Remark 5. *For an alternative proof, see Lax [1], Chapter 10, Theorem 11.*

7.4.

Proof. P as the intersection of $P^{(1)}$ and $P^{(2)}$ is a straight line. So P is orthogonal to both $x^{(1)} \times y^{(1)}$ and $x^{(2)} \times y^{(2)}$. In the full QR factorization of the 3×2 matrix $(x^{(1)}, y^{(1)})$, the third column $q_3^{(1)}$ of the Q matrix is a scalar multiple of $x^{(1)} \times y^{(1)}$; in the full QR factorization of the 3×2 matrix $(x^{(2)}, y^{(2)})$, the third column $q_3^{(2)}$ of the Q matrix is a scalar multiple of $x^{(2)} \times y^{(2)}$. Then a nonzero vector $v \in P$ can be chosen as $q_3^{(1)} \times q_3^{(2)}$, which is the third column of the Q matrix in the full QR factorization of the 3×2 matrix $(q_3^{(1)}, q_3^{(2)})$. Finding v is thus reduced to the computation of QR factorization of three 3×2 matrices. □

7.5. (a)

Proof. Suppose $A = (a_1, a_2, \cdots, a_n)$, $\hat{Q} = (q_1, q_2, \cdots, q_n)$ and

$$\hat{Q} = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ 0 & r_{22} & \cdots & r_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & r_{nn} \end{pmatrix}.$$

If A is not of full rank, then there exists k ($1 \leq k \leq n$) such that a_k can be represented by a linear combination of $a_1, \cdots, a_{k-1}$.[2] By the Gram-Schmidt orthogonalization, we necessarily have $r_{kk} = 0$. Conversely, if $r_{kk} = 0$ for some k with $1 \leq k \leq n$,[3] we must have

$$a_k = \sum_{i=1}^{k-1} r_{ik} q_i \in \langle a_1, \cdots, a_{k-1} \rangle.$$

That is, a_k is a linear combination of $a_1, \cdots, a_{k-1}$. So A is not of full rank. Combined, we conclude A has rank n if and only if all the diagonal entries of $\hat{R}$ are nonzero. □

(b)

Proof. At least k. By the condition, we have

$$\langle a_1, \cdots, a_k \rangle = \langle q_1, \cdots, q_k \rangle.$$

So the rank of A is at least k. The following example shows that the rank of A could be larger than k:

$$\hat{Q}\hat{R} = (q_1, q_2, q_3) \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} = (q_1, 0, q_2) = A.$$

Therefore $\text{rank}(A) = 2 > k = 1$. □

8 Gram-Schmidt Orthogonalization

8.1.

Proof. To obtain q_1, we use the formula $v_1 = a_1$ and $q_1 = \frac{v_1}{\|v_1\|}$. The computation of norm $\|\cdot\|$ involves m multiplications, $m-1$ additions and one square root. So the total number of flops needed for computing the Euclidean norm of a vector in $\mathbb{R}^m$ is $2m$. This implies the flops needed for obtaining q_1 is $2m + m = 3m$, where the second term m comes from division of v_1 by $\|v_1\|$. In summary, the flops for q_1 is $3m$.

Suppose we have obtained $q_1, q_2, \cdots, q_{j-1}$, we count the flops needed for obtaining q_j. Algorithm 8.1 uses the formula

$$v_j = P_{\perp q_{j-1}} \cdots P_{\perp q_2} P_{\perp q_1} a_j.$$

The projection $P_{\perp q_1} a_j = a_j - q_1 q_1^* a_j$ involves m multiplication and $(m-1)$ addition for $q_1^* a_j$, m multiplication for $q_1(q_1^* a_j)$, and m subtraction for $a_j - q_1(q_1^* a_j)$. So the total number of flops is $m+(m-1)+m+m = 4m-1$. This type of computation will be repeated $(j-1)$ times, so the flops needed to obtain v_j is $(4m-1)(j-1)$. As shown in the case of q_1, the normalization $q_j = \frac{v_j}{\|v_j\|}$ needs $3m$ flops. So the flops for q_j is $(4m-1)(j-1)+3m$.

Summing up flops in each step, the total number of flops for Algorithm 8.1 is $\sum_{j=1}^{n}[(4m-1)(j-1)+3m] = (4m-1) \cdot \frac{n(n-1)}{2} + 3mn$. □

8.2.

Proof. A translation of Algorithm 8.1 into MATLAB code is given below:

```
function [Q, R] = mgs(A)

%MGS computes a reduced QR factorization of an m x n matrix A with m no
%    less than n, using modified Gram-Schmidt orthogonalizatoin. The output
%    variables are an m x n orthonormal matrix Q and an n x n triangular
%    matrix.
```

[2] The case of $k = 1$ is understood as $a_1 = 0$.

[3] The case of $k = 1$ is understood as $a_1 = 0$.

```
%
%    Implementation is based on Algorithm 8.1 of [1], pp. 58.
%
%    Z.Y., 07/11/2010.
%
%    Reference:
%    [1] Lloyd N. Trefethen and David Bau III. Numerical Linear Algebra.
%    Exercise 8.2.

[m, n] = size(A);

if (m<n)
    error('number of rows must be no less than number of columns.');
end

Q = zeros(m,n);
R = zeros(n,n);

for i=1:n
    Q(:, i) = A(:, i);
end

for i=1:n
    R(i,i) = norm(Q(:,i),2);
    Q(:,i) = Q(:,i)/R(i,i);
    for j = (i+1):n
        R(i,j) = Q(:,i)'*Q(:,j);
        Q(:,j) = Q(:,j) - R(i,j)*Q(:,i);
    end
end

end %mgs
```

An alternative but equivalent, both theoretically and numerically, implementation is a translation of formula (8.7) into MATLAB code, which is given below:

```
function [Q, R] = mgs_alt(A)

%MGS_ALT computes a reduced QR factorization of an m x n matrix A with m no
%        less than n, using modified Gram-Schmidt orthogonalizatoin. The
%        output variables are an m x n orthonormal matrix Q and an n x n
%        triangular matrix.
%
%        Implementation is based on formula (8.7) of [1], pp. 58.
%
%        Z.Y., 07/11/2010.
%
%        Reference:
%        [1] Lloyd N. Trefethen and David Bau III. Numerical Linear Algebra.
%        Exercise 8.2.

[m, n] = size(A);

if (m<n)
```

```
    error('number of rows must be no less than number of columns.');
end

Q = zeros(m,n);
R = zeros(n,n);

R(1,1) = norm(A(:,1),2);
Q(:,1) = A(:,1)/R(1,1);

for j=2:n
    v = A(:,j);
    for i=1:(j-1)
        R(i,j) = Q(:,i)'*v;
        v = v - R(i,j)*Q(:,i);
    end
    R(j,j) = norm(v,2);
    Q(:,j) = v/R(j,j);
end

end %mgs_alt
```

□

8.3.

Proof. We have the following factorization of R_j:

$$\begin{aligned}
R_j &= \begin{pmatrix}
1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\
0 & 1 & \cdots & 0 & 0 & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & 0 & \cdots & \frac{1}{r_{jj}} & -\frac{r_{j(j+1)}}{r_{jj}} & \cdots & -\frac{r_{jn}}{r_{jj}} \\
0 & 0 & \cdots & 0 & 1 & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & 0 & \cdots & 0 & 0 & \cdots & 1
\end{pmatrix} \\
&= \begin{pmatrix}
1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\
0 & 1 & \cdots & 0 & 0 & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & 0 & \cdots & \frac{1}{r_{jj}} & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 & 1 & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & 0 & \cdots & 0 & 0 & \cdots & 1
\end{pmatrix}
\begin{pmatrix}
1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\
0 & 1 & \cdots & 0 & 0 & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & 0 & \cdots & 1 & -r_{j(j+1)} & \cdots & -r_{jn} \\
0 & 0 & \cdots & 0 & 1 & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & 0 & \cdots & 0 & 0 & \cdots & 1
\end{pmatrix}
\end{aligned}$$

From formula (8.10), we can easily see that the diagonal matrix corresponds to the line $q_i = v_i/r_{ii}$ in Algorithm 8.1, and the unit upper-triangular matrix corresponds to the line $v_j = v_j - r_{ij}q_i$ in Algorithm 8.1. □

9 MATLAB

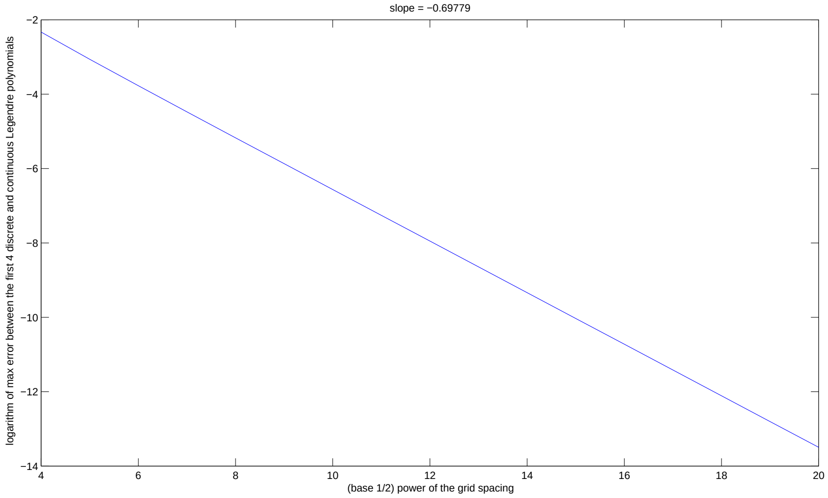
9.1.

Proof. The MATLAB code is given below:

```
function discrete_legendre_polynomials

%DISCRETE_LEGENDRE_POLYNOMIALS implements Experiment 1 of [1], Chapter 9.
```

```
%
%                              Z.Y., 07/11/2010.
%
% Reference
% [1] Lloyd N. Trefethen and David Bau III. Numerical Linear Algebra. SIAM,
%     1997.

% MATLAB program of Experiment 1
legendre(7);

% maximum error as a function of the (base 1/2) power of the grid spacing
p1 = 4;
p2 = 20;
errs = [];
for v = p1:p2
    errs = [errs, legendre(v)];
end

figure(3);
slope = (log(errs(end))-log(errs(1)))/(p2-p1);
plot(p1:p2, log(errs));
ylabel('logarithm of max error between the first 4 discrete and continuous Legendre polynomials');
xlabel('(base 1/2) power of the grid spacing');
title(['slope = ', num2str(slope)]);

end %discrete_legendre_polynomials

%%%%%%%%%%%%%%%% Nested Function %%%%%%%%%%%%%%%%%%%%%%%%%%

function error = legendre(v)

% Experiement 1
x = (-2^v:2^v)'/2^v;
A = [x.^0 x.^1 x.^2 x.^3];
[Q, R] = qr(A, 0);
scale = Q(2^(v+1)+1,:);
Q = Q*diag(1./scale);

figure(1);
plot(x, Q);

% error between the first 4 discrete and continuous Legendre polynomials,
% based on formula (7.1) of [1].
n = length(x);
L = [ones(n,1) x (1.5*x.^2-0.5) (2.5*x.^3-1.5*x)];

figure(2);
subplot(2,2,1);
plot(x, Q(:,1)-L(:,1));
subplot(2,2,2);
plot(x, Q(:,2)-L(:,2));
subplot(2,2,3);
plot(x, Q(:,3)-L(:,3));
```

```
subplot(2,2,4);
plot(x, Q(:,4)-L(:,4));

error = max(max(abs(L-Q)));

end %legendre
```



□

9.2.

Proof. The eigenvalue of A is 1; the determinant of A is 1; and the rank of A is m. To find the inverse of A, we denote by N the $m \times m$ matrix with 1 on the first superdiagonal and 0 everywhere else. Then N is nilpotent with power $(m-1)$, i.e. $N^{m-1} = 0$. Therefore $A^{-1} = (I+2N)^{-1} = I-2N+(2N)^2-\cdots+(-1)^{m-2}(2N)^{m-2}$. For example, when $m = 3$, we have

$$A = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix}, \; A^{-1} = \begin{pmatrix} 1 & -2 & 4 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix}.$$

By Theorem 5.3 and the fact that $\frac{1}{\sigma_m}$ is the largest singular value of A^{-1}, we have

$$\sigma_m = \frac{1}{\|A^{-1}\|_2}.$$

So an upper bound on σ_m is the inverse of a lower bound on $\|A^{-1}\|_2$. Since $A^{-1}e_m = ((-2)^{m-1}, (-2)^{m-2}, \cdots, 1)'$, we conclude $\|A^{-1}\|_2 \geq \|((-2)^{m-1}, (-2)^{m-2}, \cdots, 1)'\|_2 = \frac{\sqrt{4^m-1}}{\sqrt{3}}$. This gives us

$$\sigma_m \leq \frac{\sqrt{3}}{\sqrt{4^m-1}}.$$

For example, when $m = 3$, $\sigma \approx 0.1939$ while $\frac{\sqrt{3}}{\sqrt{4^m-1}} \approx 0.2182$; when $m = 5$, $\sigma \approx 0.0470$ while $\frac{\sqrt{3}}{\sqrt{4^m-1}} \approx 0.0542$. □

10 Householder Triangularization

10.1. (a)

Proof. The Householder reflector F has the general form of $I - 2qq^*$, where q is a unit vector. If λ is an eigenvalue and x is an associated eigenvector, we have $x - 2qq^*x = \lambda x$, or equivalently, $(1-\lambda)x = 2(q^*x)q$. So 1 is an eigenvalue and the space of associated eigenvectors is $\perp q = \{x : q^*x = 0\}$. Geometrically, this eigen-space is the hyperplane H in Figure 10.1, which is invariant under the Housholder reflection.

If $\lambda \neq 1$, we can conclude x is a scalar multiple of q. Suppose $x = \mu q$ with $\mu \neq 0$. Then $(1-\lambda)\mu = 2\mu$. So $\lambda = -1$ is an eigenvalue and the space of associated eigenvectors is $\langle q \rangle = \{\mu q : \mu \in \mathbb{C} \text{ or } \mathbb{R}\}$. Geometrically, this eigen-space is the straight line going through x and $\|x\|xe_1$ in Figure 10.1, which is invariance under the reflection with respect to H. □

Remark 6. *Beside direct computation, we can verify that the Housholder reflector satisfies the condition $F^2 = I$. So the eigenvalues must be either 1 or -1.*

(b)

Proof. By the fact that $F^2 = I$, we conclude $\det F = \pm 1$. Denote by $a_1, \cdots, a_m$ the eigenvalues of F, with the same multiplicity they have as roots of the characteristic polynomial of F, then $\sum_{i=1}^m a_i = \text{tr}F$ (see, for example, Lax [1], Chapter 6, Theorem 3). So $\sum_{i=1}^m a_i = m-2$. We already know from part (a) that each a_i is either 1 or -1. Define $p_+ = \#\{a_i : a_i = 1, 1 \le i \le m\}$ and $p_- = \#\{a_i : a_i = -1, 1 \le i \le m\}$. Solving the system of equations

$$\begin{cases} p_+ + p_- = m \\ p_+ - p_- = m-2, \end{cases}$$

we have $p_+ = m-1$, $p_- = 1$. Therefore $\det F = \prod_{i=1}^m a_i = -1$. □

Remark 7. *We don't really need to go through the above algebraic argument to see $\det F = -1$. Indeed, as easily seen from Figure 10.1, all the vectors on the hyperplane H have eigenvalue 1, while the vector perpendicular to H has eigenvalue -1. Since a vector perpendicular to H and a basis of H together form a basis for the whole space, we can conclude the eigenspace for eigenvalue 1 has dimension $(m-1)$ and the eigenspace for eigenvalue -1 has dimension 1. Hence $\det F = -1$.*

(c)

Proof. We note $F^* = F$. By Theorem 5.5, the singular values of F are all equal to 1. □

10.2. (a)

Proof. The MATLAB code is given below

```
function [W, R] = house(A)

%HOUSE computes an implicit representation of a full QR factorization A =
%       QR of an m x n matrix A with m >= n using Householder reflections.
%
%       Z.Y., 7/14/2010.
%
%       Reference
%       [1] Lloyd N. Trefethen and David Bau III. Numerical Linear Algebra.
%       SIAM, 1997. Algorithm 10.1. Exercise 10.2(a).

[m, n] = size(A);

if (m<n)
    error('number of rows cannot be smaller than number of columns.');
```

```
end

W = zeros(m,n);
for k=1:n
    x = A(k:m, k);
    if (x(1)>=0)
        sgn = 1;
    else
        sgn = -1;
    end
    v = sgn*norm(x,2)*eye(m-k+1,1)+x;
    v = v/norm(v,2);
    A(k:m, k:n) = A(k:m, k:n) - 2*v*v'*A(k:m, k:n);
    W(k:m,k) = v;
end

R = A(1:n,:);

end %house
```

□

(b)

Proof. The MATLAB code is given below

```
function Q = formQ(W)

%FORMQ takes the matrix W = (v1, ..., vn) produced by Algorithm 10.1 of
%      [1] as input and generates a corresponding m x m orthogonal matrix
%      Q.
%
%      Z.Y., 7/14/2010.
%
%      Reference
%      [1] Lloyd N. Trefethen and David Bau III. Numerical Linear Algebra.
%      SIAM, 1997. Algorithm 10.3. Exercise 10.2(b).

[m, n] = size(W);

if (m<n)
    error('number of rows cannot be smaller than number of columns.');
end

Q = eye(m,m);
for k = 1:m
    Q(:,k) = formQx(W,Q(:,k));
end

end %formQ

%%%%%%%%%%% Nested function %%%%%%%%%%%%%%%%

function y = formQx(W, x)
```

```
[m, n] = size(W);

if (m<n)
    error('number of rows cannot be smaller than number of columns.');
end

if (~isvector(x))
    error('x must be a vector.');
elseif (length(x) ~= m)
    error('length of x must agree with the number of rows for W.');
end

for k = n:-1:1
    x(k:m) = x(k:m)-2*W(k:m,k)*(W(k:m,k)'*x(k:m));
end

y = x;
end %formQx
```

□

10.3.

Proof. The Gram-Schmidt routine **mgs** of Exercise 8.2 produces the following Q and R:

$$Q = \begin{pmatrix} 0.1010 & 0.3162 & 0.5420 \\ 0.4041 & 0.3534 & 0.5162 \\ 0.7071 & 0.3906 & -0.5248 \\ 0.4041 & -0.5580 & 0.3871 \\ 0.4041 & -0.5580 & -0.1204 \end{pmatrix}, \ R = \begin{pmatrix} 9.8995 & 9.4954 & 9.6975 \\ 0 & 3.2919 & 3.0129 \\ 0 & 0 & 1.9701 \end{pmatrix}.$$

The Householder routines **house** and **formQ** of Exercise 10.2 produce the following Q and R:

$$Q = \begin{pmatrix} -0.1010 & -0.3162 & 0.5420 & -0.6842 & -0.3577 \\ -0.4041 & -0.3534 & 0.5162 & 0.3280 & 0.5812 \\ -0.7071 & -0.3906 & -0.5248 & 0.0094 & -0.2683 \\ -0.4041 & 0.5580 & 0.3871 & 0.3656 & -0.4918 \\ -0.4041 & 0.5580 & -0.1204 & -0.5390 & 0.4695 \end{pmatrix}, \ R = \begin{pmatrix} -9.8995 & -9.4954 & -9.6975 \\ 0 & -3.2919 & -3.0129 \\ 0 & 0 & 1.9701 \end{pmatrix}.$$

MATLAB's built-in command **qr** produces the following Q and R:

$$Q = \begin{pmatrix} -0.1010 & -0.3162 & 0.5420 \\ -0.4041 & -0.3534 & 0.5162 \\ -0.7071 & -0.3906 & -0.5248 \\ -0.4041 & 0.5580 & 0.3871 \\ -0.4041 & 0.5580 & -0.1204 \end{pmatrix}, \ R = \begin{pmatrix} -9.8995 & -9.4954 & -9.6975 \\ 0 & -3.2919 & -3.0129 \\ 0 & 0 & 1.9701 \end{pmatrix}.$$

To see all these results agree, recall $QR = (q_1, \cdots, q_n) \begin{pmatrix} r_1 \\ \cdots \\ r_n \end{pmatrix}$. So, if a column in Q changes its sign, the corresponding row in R should also change its sign. □

10.4. (a)

Proof.

$$F \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -c & s \\ s & c \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -cx + sy \\ sx + cy \end{pmatrix}.$$

So the midpoint of $\begin{pmatrix}x\\y\end{pmatrix}$ and $F\begin{pmatrix}x\\y\end{pmatrix}$ is $\begin{pmatrix}\frac{1-c}{2}x+\frac{s}{2}y\\ \frac{s}{2}+\frac{c+1}{2}y\end{pmatrix}$, which is perpendicular to the vector connecting $\begin{pmatrix}x\\y\end{pmatrix}$ and $F\begin{pmatrix}x\\y\end{pmatrix}$:

$$\frac{1}{2}\left[\begin{pmatrix}x\\y\end{pmatrix}+F\begin{pmatrix}x\\y\end{pmatrix}\right]^*\left[\begin{pmatrix}x\\y\end{pmatrix}-F\begin{pmatrix}x\\y\end{pmatrix}\right]=\frac{1}{2}\left[(x,y)\begin{pmatrix}x\\y\end{pmatrix}-(x,y)F^*F\begin{pmatrix}x\\y\end{pmatrix}\right]=0.$$

This shows F is a reflection.

$$J\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}\cos\theta & \sin\theta\\ -\sin\theta & \cos\theta\end{pmatrix}\begin{pmatrix}r\cos\alpha\\ r\sin\alpha\end{pmatrix}=r\begin{pmatrix}\cos\theta\cos\alpha+\sin\theta\sin\alpha\\ -\sin\theta\cos\alpha+\cos\theta\sin\alpha\end{pmatrix}=r\begin{pmatrix}\cos(\alpha-\theta)\\ \sin(\alpha-\theta)\end{pmatrix}.$$

So J represents a clockwise rotation of angle θ. □

11 Least Squares Problems

11.1.

Proof. If $m=n$, we have nothing to prove. So without loss of generality, we assume $m>n$. Suppose the reduced QR factorization of A is $\hat{Q}\hat{R}=\begin{pmatrix}\hat{Q}_1\\ \hat{Q}_2\end{pmatrix}\hat{R}$ where $\hat{R}$ is an $n\times n$ diagonal matrix and $\hat{Q}_1$ is also of dimension $n\times n$. Then $A_1=\hat{Q}_1\hat{R}$ and we have

$$A^+=(A^*A)^{-1}A^*=(\hat{R}^*\hat{Q}^*\hat{Q}\hat{R})^{-1}(\hat{R}^*\hat{Q}^*)=\hat{R}^{-1}\hat{Q}^*=A_1^{-1}\hat{Q}_1\hat{Q}^*.$$

Therefore $\|A^+\|_2\le\|A_1^{-1}\|_2\|\hat{Q}_1\hat{Q}^*\|_2$. To finish the proof, it suffices to prove $\|\hat{Q}_1\hat{Q}^*\|_2\le 1$. Indeed, since the columns of $\hat{Q}$ are orthonormal, we can always find an $m\times(m-n)$ matrix H such that $(\hat{Q},H)$ is an orthonormal matrix. Write H in the form of $\begin{pmatrix}H_1\\ H_2\end{pmatrix}$ where H_1 is of size $n\times(m-n)$ and H_2 is of size $(m-n)\times(m-n)$. Then, for any $b\in\mathbb{R}^m$ or $\mathbb{C}^m$, we have

$$\|\hat{Q}_1\hat{Q}^*b\|_2\le\left\|\begin{pmatrix}\hat{Q}_1 & H_1\\ \hat{Q}_2 & H_2\end{pmatrix}\begin{pmatrix}\hat{Q}^*b\\ 0\end{pmatrix}\right\|_2=\left\|\begin{pmatrix}\hat{Q}^*b\\ 0\end{pmatrix}\right\|_2\le\left\|\begin{pmatrix}\hat{Q}^*b\\ H^*b\end{pmatrix}\right\|_2=\left\|\left(\hat{Q},H\right)^*b\right\|_2=\|b\|_2.$$

This implies $\|\hat{Q}_1\hat{Q}^*\|_2\le 1$. Combined, we can conclude $\|A^+\|_2\le\|A_1^{-1}\|_2$. □

Remark 8. *We make the following interesting observation. Suppose Q is an $m\times m$ orthogonal matrix and we write it in the form of*

$$Q=\begin{pmatrix}Q_{11} & Q_{12}\\ Q_{21} & Q_{22}\end{pmatrix},$$

where Q_{11} is of size $n\times n$, Q_{12} is of size $n\times(m-n)$, Q_{21} is of size $(m-n)\times n$, and Q_{22} is of size $(m-n)\times(m-n)$. In general, we cannot claim Q_{11} is orthogonal and conclude $\|Q_{11}\|_2=1$, but we can conclude Q_{11} is a contraction under the 2-norm, i.e. $\|Q_{11}\|_2\le 1$. This is because for any $z\in\mathbb{R}^n$ or $\mathbb{C}^n$,

$$\|Q_{11}z\|_2\le\left\|\begin{pmatrix}Q_{11} & Q_{12}\\ Q_{21} & Q_{22}\end{pmatrix}\begin{pmatrix}z\\ 0\end{pmatrix}\right\|=\|z\|_2,$$

where the special property that an orthogonal matrix preserves 2-norm is used.

11.2. (a)

Proof. The problem is to find coefficient a_1, a_2 and a_3 such that $\int_1^2 \left[a_1 e^x + a_2 \sin x + a_3 \Gamma(x) - x^{-1}\right]^2 dx$ is minimized. Let $F(a_1, a_2, a_3) = \int_1^2 \left[a_1 e^x + a_2 \sin x + a_3 \Gamma(x) - x^{-1}\right]^2 dx$. Then the conditions for extremum

$$\begin{cases} \frac{\partial}{\partial a_1} F(a_1, a_2, a_3) = 0 \\ \frac{\partial}{\partial a_2} F(a_1, a_2, a_3) = 0 \\ \frac{\partial}{\partial a_3} F(a_1, a_2, a_3) = 0 \end{cases}$$

translate into the following system of linear equations

$$\begin{pmatrix} \int_1^2 e^{2x} dx & \int_1^2 e^x \sin x dx & \int_1^2 \Gamma(x) e^x dx \\ \int_1^2 e^x \sin x dx & \int_1^2 \sin^2 x dx & \int_1^2 \Gamma(x) \sin x dx \\ \int_1^2 e^x \Gamma(x) dx & \int_1^2 \sin x \Gamma(x) dx & \int_1^2 \Gamma^2(x) dx \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} \int_1^2 \frac{e^x}{x} dx \\ \int_1^2 \frac{\sin x}{x} dx \\ \int_1^2 \frac{\Gamma(x)}{x} dx \end{pmatrix}.$$

Using the numerical integration routine of MATLAB (**quad**), we can compute the matrix in the above equation as well as the column vector in the right side of the equation. Solving it gives us the (theoretically) accurate solution of optimal fitting.

```
function [acc_coeff, acc_L2err, apprx_coeff, apprx_L2err] = L2Approx(l,r)

%L2APPROX approximates the function x^(-1) in the space L^2[1, 2] by a linear
%         combination of exp(x), sin(x) and gamma(x).
%
%         Z.Y., 7/18/2010.
%
%         Reference
%         [1] Lloyd Trefethen and David Bau III. Numerical Linear Algebra,
%         SIAM, 1997. Exercise 11.2.
%
%         Example: L2Approx(1,2);

% Accurate solution: compute coefficients
tol = 1.0e-15;
f11 = @(x)exp(2*x);
f12 = @(x)exp(x).*sin(x);
f13 = @(x)gamma(x).*exp(x);
f22 = @(x)sin(x).*sin(x);
f23 = @(x)gamma(x).*sin(x);
f33 = @(x)gamma(x).*gamma(x);
g1 = @(x)exp(x)./x;
g2 = @(x)sin(x)./x;
g3 = @(x)gamma(x)./x;

accA = zeros(3,3); accb = zeros(3,1);
for i=1:3
    accb(i) = eval(['quad(g',num2str(i),', l, r,tol);']);
    for j=i:3
        accA(i,j) = eval(['quad(f',num2str(i),num2str(j),', l, r,tol);']);
        accA(j,i) = accA(i,j);
    end
end

acc_coeff = accA\accb;
acc_L2err = quad(@(x)L2Err(x,acc_coeff), l, r, tol);
```

```
figure(1);
step = 0.001;
x = l:step:r;
y = exp(x)*acc_coeff(1) + sin(x)*acc_coeff(2) + gamma(x)*acc_coeff(3);
z = 1./x;
plot(x, z, 'b', x, y, 'r');
legend('1/x', [num2str(acc_coeff(1)), 'exp(x)+', ...
               num2str(acc_coeff(2)), 'sin(x)+', ...
               num2str(acc_coeff(3)), 'gamma(x)']);
grid on;
title(['Approximation of 1/x by linear combination of exp(x), sin(x), ',...
       'gamma(x) in L2-norm over [', num2str(l),',',num2str(r),']', ...
       ', L2 err: ', num2str(acc_L2err)]);

% Approximate solution: least square problem
apprx_step = 0.00001;
xx = l:apprx_step:r;
apprxA = [exp(xx)', sin(xx)', gamma(xx)'];
apprxb = 1./xx';
apprx_coeff = apprxA\apprxb;
apprx_L2err = norm(apprxA*apprx_coeff - apprxb, 2)/length(apprxb);

figure(2);
plot(xx, 1./xx, 'b', xx, apprxA*apprx_coeff, 'r');
legend('1/x', [num2str(apprx_coeff(1)), 'exp(x)+', ...
               num2str(apprx_coeff(2)), 'sin(x)+', ...
               num2str(apprx_coeff(3)), 'gamma(x)']);
grid on;
title(['Approximation of 1/x by linear combination of exp(x), sin(x), ',...
       'gamma(x) in discrete L2-norm over [', num2str(l),',',num2str(r),...
       '], discrete L2 err: ', num2str(apprx_L2err)]);


end %appox

%%%%%%%%%%%%%%%%%%%%% Nested function %%%%%%%%%%%%%%%%%%%%%
function err = L2Err(x, a)

a1 = a(1); a2 = a(2); a3 = a(3);
err = (a1*exp(x) + a2*sin(x) + a3*gamma(x) - 1./x).^2;
end %L2_err
```

□

Remark 9. *The help information of MATLAB, under the entry* **pseudoinverses**, *explains three equivalent ways of solving a rank-deficient system: If A is m-by-n with $m > n$ and full rank n, then each of the three statements*

```
x = A\b
x = pinv(A)*b
x = inv(A'*A)*A'*b
```

theoretically computes the same least squares solution x, although the backslash operator does it faster.

11.3.

Proof. The MATLAB code for this exercise is given below:

```
function x = poly_apprx

%POLY_APPRX approximates the function x^(-1) in the space L^2[1, 2] by a linear
%          combination of exp(x), sin(x) and gamma(x).
%
%          Z.Y., 7/18/2010.
%
%          Reference
%          [1] Lloyd Trefethen and David Bau III. Numerical Linear Algebra,
%          SIAM, 1997. Exercise 11.3.

format long;
m = 50; n = 12;
t = linspace(0,1,m);

A = fliplr(vander(t));
A = A(:,1:12);
b = cos(4*t)';

%(a) Formation and solution of the normal equations, using MATLAB's \
R = chol(A'*A);
x1 = R\(R'\(A'*b));

%(b) QR factorization computed by mgs (modified Gram-Schmidt, Exercise 8.2)
[Q, R] = mgs(A);
x2 = R\(Q'*b);

%(c) QR factorization computed by house (Householder triangularization,
%Exercise 10.2)
[W, R] = house(A);
Q = formQ(W);
Q = Q(:,1:n);
x3 = R\(Q'*b);

%(d) QR factorization computed by MATLAB's qr (also Householder
%triangularization)
[Q, R] = qr(A);
x4 = R\(Q'*b);

%(e) x = A\b in MATLAB (also based on QR factorization)
x5 = A\b;

%(f) SVD, using MATLAB's svd
[U, S, V] = svd(A,0);
x6 = V*(S\(U'*b));

x = [x1, x2, x3, x4, x5, x6];
end %poly_apprx
```

The results of these six lists of twelve coefficients are recorded in the following tables, from which we can see that results from method (c)-(f) are consistent, result from method (b) is somewhat different, and result from method (a) is unstable.

coefficients	x_1	x_2	x_3
a_0	1.00000001465169	1.00000000350410	1.00000000099659
a_1	-0.00000449714573	-0.00000117405761	-0.00000042274146
a_2	-7.99982732153282	-7.99995281040164	-7.99998123572225
a_3	-0.00259730734173	-0.00074001641075	-0.00031876281895
a_4	10.68694453795226	10.67267129913343	10.66943079325323
a_5	-0.09313280176144	-0.02850450420331	-0.01382027756329
a_6	-5.42151357259457	-5.60529257307527	-5.64707565382729
a_7	-0.48952347964704	-0.15207694321151	-0.07531598015849
a_8	2.18420198493399	1.78455754119606	1.69360691536360
a_9	-0.35588588385285	-0.06108461047809	0.00603214174793
a_{10}	-0.22300544585926	-0.34618754322297	-0.37424171636699
a_{11}	0.06070014261595	0.08296770979626	0.08804057827562

coefficients	x_4	x_5	x_6
a_0	1.00000000099661	1.00000000099661	1.00000000099661
a_1	-0.00000042274331	-0.00000042274367	-0.00000042274330
a_2	-7.99998123567923	-7.99998123566715	-7.99998123567958
a_3	-0.00031876330399	-0.00031876345909	-0.00031876329890
a_4	10.66943079635676	10.66943079740723	10.66943079631425
a_5	-0.01382028980042	-0.01382029406980	-0.01382028959570
a_6	-5.64707562270147	-5.64707561163632	-5.64707562330371
a_7	-0.07531603222675	-0.07531605097010	-0.07531603110511
a_8	1.69360697232565	1.69360699300149	1.69360697099570
a_9	0.00603210250426	0.00603208818858	0.00603210347811
a_{10}	-0.37424170091224	-0.37424169526212	-0.37424170131392
a_{11}	0.08804057562238	0.08804057465256	0.08804057569379

□

References

[1] Peter Lax. *Linear algebra and its applications*, 2nd Edition. John Wiley & Sons, Inc., New York, 2007. 1, 7, 5, 10

[2] James R. Munkres. *Analysis on manifolds.* Westview Press, 1997. 1

[3] Kosaku Yosida. *Functional analysis*, 6th Edition. Springer, 2003.

2