Selected Problems ⋆ Trefethen & Bau, Numerical Linear Algebra

Nguyen Quan Ba Hong[1]
Doan Tran Nguyen Tung[2]
Nguyen An Thinh[3]
Students at Faculty of Math and Computer Science,
Ho Chi Minh University of Science, Vietnam
email. nguyenquanbahong@gmail.com
email. dtrngtung@live.com
email. anthinh297@gmail.com
blog. http://hongnguyenquanba.wordpress.com [4]

Tuesday 19th December, 2017

[1] Student ID: 1411103

[2] Student ID: 1411352

[3] Student ID: 1411389

[4] Visit my site http://hongnguyenquanba.wordpress.com to get more.

Contents

1 Fundamentals **5**

1.1 Lecture 1: Matrix-Vector Multiplications 5

1.2 Lecture 2: Orthogonal Vectors and Matrices 11

1.3 Lecture 3: Norms 17

1.4 Lecture 4: The Singular Value Decomposition 21

1.5 Lecture 5: More on the SVD 23

2 QR Factorization and Least Squares **30**

2.1 Lecture 6: Projectors 30

2.2 Lecture 7: QR Factorization 35

2.3 Lecture 8: Gram-Schmidt Orthogonalization 39

2.4 Lecture 9: MATLAB 40

2.5 Lecture 10: Householder Triangularization 41

2.6 Lecture 11: Least Squares Problems 46

3 Conditioning and Stability **54**

3.1 Lecture 12: Conditioning and Condition Numbers 54

3.2 Lecture 13: Floating Point Arithmetic 56

3.3 Lecture 14: Stability 56

3.4 Lecture 15: More on Stability 57

3.5 Lecture 16: Stability of Householder Triangularization 58

3.6 Lecture 17: Stability of Back Substitution 59

3.7 Lecture 18: Conditioning of Least Squares Problems 60

3.8 Lecture 19: Stability of Least Squares Algorithms 61

4 Systems of Equations **63**

4.1 Lecture 20: Gaussian Elimination 63

4.2 Lecture 21: Pivoting 65

4.3 Lecture 22: Stability of Gaussian Elimination 67

4.4 Lecture 23: Cholesky Factorization 67

5 Eigenvalues **69**

5.1 Lecture 24: Eigenvalue Problems 69

5.1.1 Gershgorin Circle Theorem 71

5.1.2 An Applications of Gerschgorin's Theorem 73

5.2 Lecture 25: Overview of Eigenvalue Algorithms 74

5.3 Lecture 26: Reduction to Hessenberg or Tridiagonal Form 75

5.4 Lecture 27: Rayleigh Quotient, Inverse Iteration 76

5.5 Lecture 28: QR Algorithm without Shifts 78
5.6 Lecture 29: QR Algorithm with Shifts 79
5.7 Lecture 30: Other Eigenvalue Algorithms 80
5.8 Lecture 31: Computing the SVD 81

6 Iterative Methods 83

6.1 Lecture 32: Overview of Iterative Methods 83
6.2 Lecture 33: The Arnoldi Iteration 84
6.3 Lecture 34: How Arnoldi Locates Eigenvalues 85
6.4 Lecture 35: GMRES 86
6.5 Lecture 36: The Lanczos Iteration 89
6.6 Lecture 37: From Lanczos to Gauss Quadrature 91
6.7 Lecture 38: Conjugate Gradients 92
6.8 Lecture 39: Biorthogonalization Methods 94
6.9 Lecture 40: Preconditioning 96

List of Figures

2.1 *Exercise 9.3, [1].* 41

2.2 *The L^2 norm of error on the interval* $[1, 2]$. 48

2.3 *The L^2 norm of error on the interval* $\left[10^{-13}, 1\right]$. 50

2.4 *The L^2 norm of error on the interval* $\left[10^{-14}, 1\right]$. 50

2.5 *The L^2 norm of error on the interval* $\left[10^{-15}, 1\right]$. 51

2.6 *The L^2 norm of error on the interval* $\left[10^{-16}, 1\right]$. 52

3.1 *Degree 10 polynomial interpolant to eleven data points. The axis scales are not given, as these have no effect on the picture.* 55

5.1 *The Rayleigh quotient $r(x)$ is a continuous function on the unit sphere $\|x\| = 1$ in $\mathbb{R}^n$, and the stationary points of $r(x)$ are the normalized eigenvectors of A. In this example with $m = 3$, there are three orthogonal stationary points (as well as their antipodes).* 78

5.2 *Operation counts for three bidiagonalization algorithms applied to $m \times n$ matrices,. Three-step bidiagonalization provides a pleasingly smooth interpolant between the other two methods, though the improvement is hardly large.* 81

6.1 *GMRES convergence curve for the matrix of of Figure 35.4. The convergence has slowed down greatly. When an iterative method stagnates like this, it is time to look for a better preconditioner.* . 87

6.2 *Eigenvalues of the 200×200 matrix A of `m = 200; A = 2*eye(m) + 0.5*randn(m)/sqrt(m)`. The dashed curve is the circle of radius $\frac{1}{2}$ with center $z = 2$ in $\mathbb{C}$. The eigenvalues are approximately uniformly distributed within this disk.* 88

6.3 *Eigenvalues of the 200×200 matrix, like that of `m = 200; A = 2*eye(m) + 0.5*randn(m)/sqrt(m)` except with a modified diagonal. Now the eigenvalues surround the origin on one side.* . . 88

6.4 *Ritz values for the first 20 steps of the Lanczos iteration applied to the same matrix. The convergence to the eigenvalues 2.5 and 3.0 is geometric. Little useful convergence to individual eigenvalues occurs in the $[0, 2]$ part of the spectrum. Instead, the Ritz values in $[0, 2]$ approximate Chebyshev points in that interval, marked by dots on the right-hand boundary.* 90

6.5 *CG convergence curves for the* 500×500 *sparse matrices A described in the text. For* $\tau = 0.01$*, the system is solved about 700 times faster by CG than by Cholesky factorization. For* $\tau = 0.2$*, the matrix is not positive definite and there is no convergence.* . . 93
6.6 *Comparison of GMRES and BCG for the* 500×500 *matrix labeled* $\tau = 0.01$ *in Figure 38.1, but with the signs of the entries randomized.* 95

Chapter 1

Fundamentals

1.1 Lecture 1: Matrix-Vector Multiplications

Problem 1.1 (Exercise 1.1, [1]). *Let B be a 4×4 matrix to which we apply the following operations:*

1. *double column 1,*
2. *halve row 3,*
3. *add row 3 to row 1,*
4. *interchange columns 1 and 4,*
5. *subtract row 2 from each of the other rows,*
6. *replace column 4 by column 3,*
7. *delete column 1 (so that the column dimension is reduced by 1).*

Then

1. *Write the result as a product of eight matrices.*
2. *Write it again as a product ABC (same B) of three matrices.*

Problem 1.2 (Exercise 1.2, [1]). *Suppose masses m_1, m_2, m_3, m_4 are located at positions x_1, x_2, x_3, x_4 in a line and connected by springs constants k_{12}, k_{23}, k_{34} whose natural lengths of extension are l_{12}, l_{23}, l_{34}. Let f_1, f_2, f_3, f_4 denote the rightward forces on the masses, e.g., $f_1 = k_{12}(x_2 - x_1 - l_{12})$.*

1. *Write the 4×4 matrix equation relating the column vectors f and x. Let K denote the matrix in this equation.*
2. *What are the dimensions of the entries of K in the physics sense (e.g., mass times time, distance divided by mass, etc.)?*
3. *What are the dimensions of $\det(K)$, again in the physics sense?*

4. *Suppose K is given numerical values based on the units meters, kilograms, and seconds. Now the system is rewritten with a matrix K' based on centimeters, grams, and seconds. What is the relationship of K' to K? What is the relationship of* $\det(K')$ *to* $\det(K)$*?*

Problem 1.3 (Exercise 1.3, [1]). *Generalizing Example 1.3, [1], we say that a square or rectangular matrix R with entries r_{ij} is upper-triangular if $r_{ij} = 0$ for $i > j$. By considering what space is spanned by the first n columns of R and using*

$$e_j = \sum_{i=1}^{m} z_{ij} a_i \tag{1.1}$$

show that if R is a nonsingular $m \times m$ upper-triangular matrix, then R^{-1} is also upper-triangular. (The analogous result also holds for lower-triangular matrices.)

Problem 1.4 (Exercise 1.4, [1]). *Let $f_1, \dots, f_8$ be a set of functions defined on the interval $[1, 8]$ with the property that for any numbers $d_1, \dots, d_8$, there exists a set of coefficients $c_1, \dots, c_8$ such that*

$$\sum_{j=1}^{8} c_j f_j(i) = d_i, \quad i = 1, \dots, 8 \tag{1.2}$$

1. *Show by appealing to the theorems of Lecture 1, [1], that $d_1, \dots, d_8$ determine $c_1, \dots, c_8$ uniquely.*
2. *Let A be the 8×8 matrix representing the linear mapping from data $d_1, \dots, d_8$ to coefficients $c_1, \dots, c_8$. What is the i, j entry of A^{-1}?*

Problem 1.5. *Let $A = (a_{ij})_{m \times n} \in \mathbb{R}^{m \times n}$ be a nonzero matrix.*
Prove that

$$column\_rank(A) = row\_rank(A) \tag{1.3}$$

Solution. We write A as

$$\left[\begin{array}{c|c|c} a_1 & \cdots & a_n \end{array}\right] \tag{1.4}$$

where $a_1, \dots, a_n$ are the columns of A.

Let $column\_rank(A) = r$ and $\{b_1, \dots, b_r\}$ be the basis of the column space of A and

$$B = (b_{ij}) = \left[\begin{array}{c|c|c} b_1 & \cdots & b_r \end{array}\right] \in \mathbb{R}^{m \times r} \tag{1.5}$$

Since $\{b_1, \dots, b_r\}$ is the basis of the column space of A, we can write each column of A as a linear combination of $b_1, \dots, b_r$

$$a_i = \sum_{j=1}^{r} c_{ji} \left[\begin{array}{c} b_j \end{array}\right] \tag{1.6}$$

Let

$$C = (c_{ij}) = \left[\begin{array}{c|c|c} c_1 & \cdots & c_n \end{array}\right] \in R^{r \times n} \tag{1.7}$$

be a matrix with c_{ij} being the coefficients defined in (1.7) and $c_1, \dots, c_n$ are the columns of C.

Due to (1.7), we can write

$$A = BC \tag{1.8}$$

$$\Leftrightarrow \left[\begin{array}{c|c|c} a_1 & \cdots & a_n \end{array}\right] = \left[\begin{array}{c|c|c} b_1 & \cdots & b_r \end{array}\right] \left[\begin{array}{c|c|c} c_1 & \cdots & c_n \end{array}\right] \tag{1.9}$$

$$\Leftrightarrow \left[\begin{array}{c} {a_1}' \\ \hline \vdots \\ \hline {a_n}' \end{array}\right] = \left[\begin{array}{c} {b_1}' \\ \hline \vdots \\ \hline {b_m}' \end{array}\right] \left[\begin{array}{c} {c_1}' \\ \hline \vdots \\ \hline {c_r}' \end{array}\right] \tag{1.10}$$

with $\{{a_1}', \dots, {a_m}'\}, \{{b_1}', \dots, {b_m}'\}, \{{c_1}', \dots, {c_r}'\}$ are the rows of A, B, C respectively.

We can write

$${a_i}' = \sum_{j=1}^{r} b_{ij} \left[\quad {c_j}' \quad \right] \tag{1.11}$$

which means each row of A can be written as a linear combination of the rows of C, this leads to

$$row\_rank\,(A) \le r \tag{1.12}$$

Suppose on the contrary that

$$row\_rank\,(A) < r \tag{1.13}$$

Let $\{{e_1}', \dots, {e_{r'}}'\}$ be the basis of the row space of A,

$$E = (e_{ij}) = \left[\begin{array}{c|c|c} e_1 & \cdots & e_n \end{array}\right] = \left[\begin{array}{c} {e_1}' \\ \hline \vdots \\ \hline {e_{r'}}' \end{array}\right] \tag{1.14}$$

We can write

$${a_i}' = \sum_{j=1}^{r} d_{ij} \left[\quad {e_j}' \quad \right] \tag{1.15}$$

Let

$$D = (d_{ij}) = \left[\begin{array}{c|c|c} d_1 & \cdots & d_{r'} \end{array}\right] = \left[\begin{array}{c} {d_1}' \\ \hline \vdots \\ \hline {d_m}' \end{array}\right] \tag{1.16}$$

with d_{ij} being the coefficients in (1.59).

Similarly to (1.8), we can write

$$A = DE \tag{1.17}$$

$$\Leftrightarrow \left[\begin{array}{c} {a_1}' \\ \hline \\ \hline {a_n}' \end{array}\right] = \left[\begin{array}{c} {d_1}' \\ \hline \\ \hline {d_m}' \end{array}\right] \left[\begin{array}{c} {e_1}' \\ \hline \\ \hline {e_{r'}}' \end{array}\right] \tag{1.18}$$

$$\Leftrightarrow \left[\begin{array}{c|c|c} a_1 & \cdots & a_n \end{array}\right] = \left[\begin{array}{c|c|c} d_1 & \cdots & d_{r'} \end{array}\right] \left[\begin{array}{c|c|c} e_1 & \cdots & e_n \end{array}\right] \tag{1.19}$$

(1.17) leads to

$$a_i = \sum_{j=1}^{r'} e_{ji} \left[\begin{array}{c} d_j \end{array}\right] \tag{1.20}$$

which means each column of A can be written as a linear combination of the columns of D and

$$column\_rank\,(A) \leq r' < r \tag{1.21}$$

This contradiction finishes our proof. □

Problem 1.6. *Prove that all eigenvalues of any real symmetric matrix are real.*

Solution. Let $A \in \mathbb{R}^{n\times n}$ be a nonzero real symmetric matrix and λ be an eigenvalue of A, then there exists a nonzero vector $X \in C^n$ such that

$$AX = \lambda X \tag{1.22}$$

Take transpose two hands of (1.22)

$$X^T A^T = \lambda X^T \tag{1.23}$$

Take complex conjugate of two hands of (1.22)

$$A\bar{X} = \bar{\lambda}\bar{X} \tag{1.24}$$

Combining (1.23) and (1.24), we deduce

$$X^T A\bar{X} = X^T \bar{\lambda}\bar{X} = \bar{\lambda}\left(X^T\bar{X}\right) \tag{1.25}$$

On the other hand, since A is symmetric, we have

$$X^T A\bar{X} = \left(X^T A^T\right)\bar{X} = \left(\lambda X^T\right)\bar{X} \tag{1.26}$$

Combining (1.25) and (1.26), we deduce that

$$\left(\lambda - \bar{\lambda}\right)\left(X^T\bar{X}\right) = 0 \tag{1.27}$$

We write down vector X explicitly as

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \tag{1.28}$$

Then

$$X^T \bar{X} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_n \end{bmatrix} = \sum_{i=1}^{n} |x_n|^2 \tag{1.29}$$

Since X is a nonzero vector, (1.29) deduces

$$X^T \bar{X} > 0 \tag{1.30}$$

Hence, $\lambda = \bar{\lambda}$, equivalently, λ is real. □

Problem 1.7. *Let $A = (a_{ij}) \in \mathbb{R}^{m \times n}$ and $\langle \cdot, \cdot \rangle$ be the standard inner product in $\mathbb{R}^n$.*
Prove that A is symmetric if and only if

$$\langle x, Ay \rangle = \langle Ax, y \rangle, \forall x, y \in \mathbb{R}^n \tag{1.31}$$

Solution. Suppose A is symmetric and $x, y \in \mathbb{R}^n$, we have

$$\langle x, Ay \rangle = x^T Ay = x^T A^T y = (Ax)^T y = \langle Ax, y \rangle \tag{1.32}$$

Conversely, suppose that

$$\langle x, Ay \rangle = \langle Ax, y \rangle, \forall x, y \in \mathbb{R}^n \tag{1.33}$$

Transform as above, we have

$$x^T Ay = x^T A^T y, \forall x, y \in \mathbb{R}^n \tag{1.34}$$

Since (1.34) holds for every $y \in \mathbb{R}^n$, using the familiar property of linear system of equations, we deduce that

$$x^T A = x^T A^T, \forall x \in \mathbb{R}^n \tag{1.35}$$

Since (1.35) holds for every $x \in \mathbb{R}^n$, we also deduce that

$$A = A^T \tag{1.36}$$

This completes our proof. □

Problem 1.8. *Let $A, B, C \in R^{n \times n}$ for which A and C are invertible and I be a block matrix defined by*

$$I = \left[\begin{array}{c|c} A & B \\ \hline 0 & C \end{array}\right] \in \mathbb{R}^{2n \times 2n} \tag{1.37}$$

1. *Prove that I is invertible.*

2. *Find I^{-1}.*

SOLUTION. We consider the following block matrix

$$X = \left[\begin{array}{c|c} A^{-1} & -A^{-1}BC^{-1} \\ \hline 0 & C^{-1} \end{array}\right] \in R^{2n\times 2n} \tag{1.38}$$

Since A and C are invertible, X makes sense. Now, we only need to very that X is invertible matrix of I.

$$XI = \left[\begin{array}{c|c} A^{-1} & -A^{-1}BC^{-1} \\ \hline 0 & C^{-1} \end{array}\right]\left[\begin{array}{c|c} A & B \\ \hline 0 & C \end{array}\right] \tag{1.39}$$

$$= \left[\begin{array}{c|c} A^{-1}A & A^{-1}B - A^{-1}BC^{-1}C \\ \hline 0 & C^{-1}C \end{array}\right] \tag{1.40}$$

$$= \left[\begin{array}{c|c} I_n & 0 \\ \hline 0 & I_n \end{array}\right] \tag{1.41}$$

$$= I_{2n} \tag{1.42}$$

Checking $IX = I_{2n}$ is similar. Therefore, I is invertible and $I^{-1} = X$. □

Problem 1.9. *Prove or give counter-examples to the following assertions.*

1. *Product of two diagonal matrices is a diagonal matrix.*

2. *Product of two upper-triangular matrices is a upper-triangular matrix.*

3. *Product of two symmetric matrices is a symmetric matrix.*

SOLUTION.
1. We consider two arbitrary diagonal matrices

$$A = diag\,(a_1, \ldots, a_n)\,, B = diag\,(b_1, \ldots, b_n) \tag{1.43}$$

Multiplying A by B as usual, we obtain easily

$$AB = diag\,(a_1b_1, \ldots, a_nb_n) \tag{1.44}$$

Hence, product of two diagonal matrices is a diagonal matrix. □

2. Let $A = (a_{ij}) \in \mathbb{R}^{n\times n}, B = (b_{ij}) \in \mathbb{R}^{n\times n}$ be two upper-triangular matrices and $C = (c_{ij}) = AB \in \mathbb{R}^{n\times n}$.

Since A and B are upper-triangular,

$$a_{ij} = b_{ij} = 0, \forall 1 \le j < i \le n \tag{1.45}$$

Hence, for all couples $1 \le j < i \le n$

$$c_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj} \tag{1.46}$$

$$= \sum_{k=1}^{j} a_{ik}b_{kj} + \sum_{k=j+1}^{n} a_{ik}b_{kj} \tag{1.47}$$

$$= 0 + 0 \tag{1.48}$$
$$= 0 \tag{1.49}$$

Therefore, C is a upper-triangular matrix. □

3. Let $A, B \in \mathbb{R}^{n\times n}$ be two symmetric matrices and $C = AB$.

The necessary and sufficient condition for which C is symmetric is

$$\begin{aligned} C^T = C &\Leftrightarrow (AB)^T = AB && (1.50)\\ &\Leftrightarrow B^T A^T = AB && (1.51)\\ &\Leftrightarrow BA = AB && (1.52)\end{aligned}$$

Hence, if A and B are not commutative, C will not be symmetric.

Counter-example 1.9.1. We can take an easy example

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}, B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \tag{1.53}$$

Then

$$AB = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \tag{1.54}$$
$$BA = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \tag{1.55}$$

Hence, $AB \neq BA$.

This counter-example collapses the assertion. □

1.2 Lecture 2: Orthogonal Vectors and Matrices

Problem 1.10 (Exercise 2.1, [1]). *Show that if a matrix A is both triangular and unitary, then it is diagonal.*

SOLUTION. Suppose A is upper-triangular. The case A is lower-triangular is handled similarly.

Since A is unitary, we have

$$\begin{aligned} [A^*A]_{ij} &= \sum_{k=1}^{n} [A^*]_{ik}[A]_{kj} && (1.56)\\ &= \sum_{k=1}^{n} \bar{a}_{ki}a_{kj} && (1.57)\end{aligned}$$

Combining with $A^*A = I$, we deduce

$$\sum_{k=1}^{n} \bar{a}_{ki}a_{kj} = \delta_{ij}, 1 \leq i, j \leq n \tag{1.58}$$

Assuming that $i > j$, (2.138) becomes

$$\sum_{k=1}^{j} \bar{a}_{ki} a_{kj} = \delta_{ij}, 1 \le j \le i \le n \tag{1.59}$$

Take $j = 1$ in (1.59), we obtain

$$\bar{a}_{1i} a_{11} = \delta_{i1}, 1 \le i \le n \tag{1.60}$$

Take $i = 1$ in (5.5), we obtain $|a_{11}| = 1$. Hence, $a_{11} \neq 0$. Combining this with (5.5), we deduce that

$$a_{1i} = 0, 1 < i \le n \tag{1.61}$$

Now, we will use *strong induction* on N to handle the situation. Suppose that

$$a_{mi} = 0, m < i \le n, \forall m < N \le n \tag{1.62}$$

Now we prove (1.194) holds for $m = N$.

Indeed, taking $j = N$ in (1.59), we obtain

$$\sum_{k=1}^{N} \bar{a}_{ki} a_{kN} = \delta_{iN}, 1 \le N \le i \le n \tag{1.63}$$

Using the induction hypotheses (1.194), (1.195) becomes

$$\bar{a}_{Ni} a_{NN} = \delta_{iN}, N \le i \le n \tag{1.64}$$

We now handle this as the case $N = 1$. Taking $i = N$ in (2.144), we obtain $|a_{NN}| = 1$. Hence $a_{NN} \neq 0$. Combining this with (2.144), we deduce

$$a_{Ni} = 0, N < i \le n \tag{1.65}$$

So, (1.194) holds for $N \le n$. This means

$$a_{ij} = 0, \forall i < j \tag{1.66}$$

Therefore, A is a diagonal matrix. □

Problem 1.11 (Exercise 2.2, [1]). *The Pythagorean theorem asserts that for a set of n orthogonal vectors $\{x_i\}$,*

$$\left\| \sum_{i=1}^{n} x_i \right\|^2 = \sum_{i=1}^{n} \|x_i\|^2 \tag{1.67}$$

1. *Prove this in the case $n = 2$ by an explicit computation of $\|x_1 + x_2\|^2$.*

2. *Show that this computation also establishes the general case, by induction.*

Solution.

1. For $n = 2$, we compute straightforward,

$$\begin{aligned}\|x_1 + x_2\|^2 &= (x_1 + x_2)^* (x_1 + x_2) && (1.68)\\ &= x_1^* x_1 + x_1^* x_2 + x_2^* x_1 + x_2^* x_2 && (1.69)\\ &= x_1^* x_1 + x_2^* x_2 && (1.70)\\ &= \|x_1\|^2 + \|x_2\|^2 && (1.71)\end{aligned}$$

   where we have used

$$x_1^* x_2 = x_2^* x_1 = 0 \tag{1.72}$$

   since x_1 and x_2 are orthogonal.

2. The general case is similar. We compute straightforward

$$\begin{aligned}\left\|\sum_{i=1}^n x_i\right\|^2 &= \left(\sum_{i=1}^n x_i\right)^* \left(\sum_{i=1}^n x_i\right) && (1.73)\\ &= \sum_{i=1}^n x_i^* x_i + \sum_{i,j=1,i\neq j}^n x_i^* x_j && (1.74)\\ &= \sum_{i=1}^n \|x_i\|^2 && (1.75)\end{aligned}$$

   where we have used

$$x_i^* x_j = 0, \forall i \neq j \tag{1.76}$$

   since $\{x_i\}_{i=1}^n$ is a set of orthogonal vectors.

Done. □

Problem 1.12 (Exercise 2.3, [1]). *Let $A \in \mathbb{C}^{m\times m}$ be hermitian. An eigenvector of A is a nonzero vector $x \in \mathbb{C}^m$ such that $Ax = \lambda x$ for some $\lambda \in \mathbb{C}$, the corresponding eigenvalue.*

1. *Prove that all eigenvalues of A are real.*
2. *Prove that if x and y are eigenvectors corresponding to distinct eigenvalues, then x and y are orthogonal.*

Solution.

1. Let λ be an arbitrary eigenvalues of A. There exists a nonzero vector $x \in \mathbb{C}^m$ for which

$$Ax = \lambda x \tag{1.77}$$

   Taking the hermitian conjugate of both sides of (1.138), we obtain

$$x^* A = \bar{\lambda} x^* \tag{1.78}$$

   where we have used $A = A^*$ since A is hermitian.

We now consider the quantity x^*Ax. We can compute it by two ways.

$$x^*Ax = (x^*A)\,x \tag{1.79}$$
$$= \bar{\lambda}x^*x = \bar{\lambda}\|x\|^2 \tag{1.80}$$

and

$$x^*Ax = x^*\,(Ax) \tag{1.81}$$
$$= \lambda x^*x \tag{1.82}$$
$$= \lambda\|x\|^2 \tag{1.83}$$

Hence

$$\left(\bar{\lambda} - \lambda\right)\|x\|^2 = 0 \tag{1.84}$$

Combining this with x is a nonzero vector, we have

$$\bar{\lambda} = \lambda \tag{1.85}$$

i.e., λ is real. Since λ is taken arbitrarily, we deduce that all eigenvalues of A are real.

2. Denote $\lambda_1 \neq \lambda_2$ are two eigenvalues of A associated with x and y, respectively. We consider the quantity x^*Ay by two ways.

$$x^*Ay = x^*\,(Ay) \tag{1.86}$$
$$= \lambda_2 x^*y \tag{1.87}$$

and

$$x^*Ay = (Ax)^*y \tag{1.88}$$
$$= (\lambda_1 x)^*y \tag{1.89}$$
$$= \bar{\lambda}_1 x^*y \tag{1.90}$$
$$= \lambda_1 x^*y \tag{1.91}$$

where we have used the fact that λ_1 is real.

Combing these, we deduce that

$$(\lambda_1 - \lambda_2)\,x^*y = 0 \tag{1.92}$$

Hence, $x^*y = 0$, i.e., x and y are orthogonal.

Done. □

Problem 1.13 (Exercise 2.4, [1]). *What can be said about the eigenvalues of a unitary matrix?*

Solution. Let λ be an arbitrary eigenvalues of a given unitary matrix A. There exists a nonzero vector x for which

$$Ax = \lambda x \tag{1.93}$$

Multiplying A^* to both sides of (4.4), we obtain

$$x = \lambda A^* x \tag{1.94}$$

where we have used $A^* A = I$ since A is a unitary matrix.

We now compute the quantity $x^* A^* x$ by two ways.

$$\begin{align} x^* A^* x &= x^* \left(A^* x \right) \tag{1.95} \\ &= \frac{1}{\lambda} x^* x \tag{1.96} \\ &= \frac{1}{\lambda} \|x\|^2 \tag{1.97} \end{align}$$

and

$$\begin{align} x^* A^* x &= \left(x^* A^* \right) x \tag{1.98} \\ &= \left(Ax \right)^* x \tag{1.99} \\ &= \left(\lambda x \right)^* x \tag{1.100} \\ &= \bar{\lambda} x^* x \tag{1.101} \\ &= \bar{\lambda} \|x\|^2 \tag{1.102} \end{align}$$

We deduce that

$$\left(\frac{1}{\lambda} - \bar{\lambda} \right) \|x\|^2 = 0 \tag{1.103}$$

Hence,

$$|\lambda| = 1 \tag{1.104}$$

We conclude that all eigenvalues of an arbitrary unitary matrix must have unit magnitude. □

Problem 1.14 (Exercise 2.5, [1]). *Let $S \in \mathbb{C}^{m \times m}$ be* ***skew-hermitian****, i.e., $S^* = -S$.*

1. *Show by using Problem 1.12 that the eigenvalues of S are pure imaginary.*
2. *Show that $I - S$ is nonsingular.*
3. *Show that the matrix $Q = (I - S)^{-1} (I + S)$, known as the* ***Cayley transform*** *of S, is unitary. (This is a matrix analogue of a linear fractional transformation $\frac{1+s}{1-s}$, which maps the left half of the complex s-plane conformally onto the unit disk.*

Solution

1. By simple observation

   $$(iS)^* = -iS^* = iS \tag{1.105}$$

   we deduce that iS is hermitian. Hence, using Problem 3, we have all eigenvalues of iS are real. Therefore, all eigenvalues of S are pure imaginary.

2. Let λ be an arbitrary eigenvalue of $I - S$. There exists a nonzero vector x for which

$$(I - S)\,x = \lambda x \tag{1.106}$$

equivalently,

$$Sx = (1 - \lambda)\,x \tag{1.107}$$

Since all eigenvalues of S are pure imaginary, we have

$$1 - \lambda = i\alpha \tag{1.108}$$

where α is a nonzero real number. Hence,

$$\lambda = 1 - i\alpha \neq 0 \tag{1.109}$$

So, all eigenvalues of $I - S$ are nonzero. Therefore, the $\det(I - S)$ which is the product of these eigenvalues is nonzero, i.e., $I - S$ is nonsingular.

3. We compute straightforward

$$\begin{aligned}
Q^*Q &= \left[(I - S)^{-1}(I + S)\right]^*(I - S)^{-1}(I + S) && (1.110)\\
&= (I + S)^*\left[(I - S)^{-1}\right]^*(I - S)^{-1}(I + S) && (1.111)\\
&= (I - S)\left[(I - S)^*\right]^{-1}(I - S)^{-1}(I + S) && (1.112)\\
&= (I - S)\,(I + S)^{-1}(I - S)^{-1}(I + S) && (1.113)\\
&= (I - S)\,(I + S)^{-1}(I + S)\,(I - S)^{-1} && (1.114)\\
&= (I - S)\,(I - S)^{-1} && (1.115)\\
&= I && (1.116)
\end{aligned}$$

where we have used the fact that$(I - S)^{-1}$ and $(I + S)$ are commutative. Indeed,

$$\begin{aligned}
&(I - S)^{-1}(I + S) = (I + S)\,(I - S)^{-1} && (1.117)\\
&\Leftrightarrow (I + S)\,(I - S) = (I - S)\,(I + S) && (1.118)\\
&\Leftrightarrow I - S^2 = I - S^2 && (1.119)
\end{aligned}$$

Hence, Q is unitary.

Done. □

Problem 1.15 (Exercise 2.6, [1]). *If u and v are m-vectors, the matrix $A = I + uv^*$ is known as a **rank-one perturbation of the identity**. Show that if A is nonsingular, then its inverse has the form $A^{-1} = I + \alpha uv^*$ for some scalar α, and give an expression for α. For what u and v is A singular? If it is singular, what is null(A)?*

Problem 1.16 (Exercise 2.7, [1]). *A **Hadamard matrix** is a matrix whose entries are all ±1 and whose transpose is equal to its inverse times a constant* *factor. It is known that if A is a Hadamard matrix of dimension $m > 2$, then m is a multiple of 4. It is not known, however, whether there is a Hadamard matrix for every such m, though examples have been found for all cases $m \leq 424$.*

Show that the following recursive description provides a Hadamard matrix of each dimension $m = 2^k$, $k = 0, 1, 2, \ldots$:

$$H_0 = [1] \tag{1.120}$$

$$H_{k+1} = \begin{bmatrix} H_k & H_k \\ H_k & -H_k \end{bmatrix} \tag{1.121}$$

1.3 Lecture 3: Norms

Problem 1.17 (Exercise 3.1, [1]). *Prove that if W is an arbitrary nonsingular matrix, the function $\|\cdot\|_W$ defined by*

$$\|x\|_W = \|Wx\| \tag{1.122}$$

is a vector norm.

Problem 1.18 (Exercise 3.2, [1]). *Let $\|\cdot\|$ denote any norm on $\mathbb{C}^m$ and also the induced matrix norm on $\mathbb{C}^{m\times m}$. Show that*

$$\rho(A) \leq \|A\| \tag{1.123}$$

where $\rho(A)$ is the ***spectral radius*** *of A, i.e., the largest absolute value $|\lambda|$ of an eigenvalue of A.*

Solution. Let λ be an arbitrary eigenvalue of A, and x be the unit nonzero vector of A associated with λ

$$Ax = \lambda x \tag{1.124}$$

We have

$$\|Ax\| = \|\lambda x\| = |\lambda| \|x\| = |\lambda| \tag{1.125}$$

and

$$\|Ax\| \leq \|A\| \|x\| = \|A\| \tag{1.126}$$

Combining (5.10) and (1.138), we deduce that

$$|\lambda| \leq \|A\| \tag{1.127}$$

holds for all eigenvalues of A. In particular,

$$\rho(A) \leq \|A\| \tag{1.128}$$

holds. □

Problem 1.19 (Exercise 3.3, [1]). *Vector and matrix p-norms are related by various inequalities, often involving the dimensions m or n. For each of the following, verify the inequality and give an example of a nonzero vector or* *matrix (for general m, n) for which equality is achieved. In this problem x is an m-vector and A is an $m \times n$ matrix.*

$$\|x\|_\infty \leq \|x\|_2 \tag{1.129}$$

$$\|x\|_2 \leq \sqrt{m}\|x\|_\infty \tag{1.130}$$

$$\|A\|_\infty \leq \sqrt{n}\|A\|_2 \tag{1.131}$$

$$\|A\|_2 \leq \sqrt{m}\|A\|_\infty \tag{1.132}$$

Problem 1.20 (Exercise 3.4, [1]). *Let A be an $m \times n$ matrix and let B be a submatrix of A, that is, a $\mu \times \nu$ matrix ($\mu \leq m, \nu \leq n$) obtained by selecting certain rows and columns of A.*

1. *Explain how B can be obtained by multiplying A by certain row and column "deletion matrices" as in step 7 of Exercise 1.1, [1].*
2. *Using this product, show that*

$$\|B\|_p \leq \|A\|_p \tag{1.133}$$

for any p with $1 \leq p \leq \infty$.

Problem 1.21 (Exercise 3.5, [1]). *Example 3.6, [1], shows that if E is an outer product $E = uv^*$, then*

$$\|E\|_2 = \|u\|_2\|v\|_2 \tag{1.134}$$

Is the same true for the Frobenius norm, i.e.,

$$\|E\|_F = \|u\|_F\|v\|_F \tag{1.135}$$

Prove it or give a counterexample.

Solution. Let λ be an arbitrary eigenvalue of A, and x be the unit nonzero vector of A associated with λ

$$Ax = \lambda x \tag{1.136}$$

We have

$$\|Ax\| = \|\lambda x\| = |\lambda| \, \|x\| = |\lambda| \tag{1.137}$$

and

$$\|Ax\| \leq \|A\| \, \|x\| = \|A\| \tag{1.138}$$

Combining (5.10) and (1.138), we deduce that

$$|\lambda| \leq \|A\| \tag{1.139}$$

holds for all eigenvalues of A. In particular,

$$\rho(A) \leq \|A\| \tag{1.140}$$

holds. □

Problem 1.22 (Exercise 3.6, [1]). *Let $\|\cdot\|$ denote any norm on $\mathbb{C}^m$. The corresponding dual norm $\|\cdot\|'$ is defined by the formula*

$$\|x\|' = \sup_{\|y\|=1} |y^* x| \tag{1.141}$$

1. *Prove that $\|\cdot\|'$ is a norm.*
2. *Let $x, y \in \mathbb{C}^m$ with $\|x\| = \|y\| = 1$ be given. Show that there exists a rank-one matrix $B = yz^*$ such that $Bx = y$ and $\|B\| = 1$, where $\|B\|$ is the matrix norm of B induced by the vector norm $\|\cdot\|$. You may use the following lemma, without proof.*

Lemma 1.22.1. *Given $x \in \mathbb{C}^m$, there exists a nonzero $z \in \mathbb{C}^m$ such that*

$$|z^* x| = \|z\|' \|x\| \tag{1.142}$$

SOLUTION.

1. We prove that $\|\cdot\|'$ satisfies three properties of a norm in turns.

   (a) We have

   $$\|x\|' = \sup_{\|y\|=1} |y^* x| \geq 0, \forall x \tag{1.143}$$

   and

   $$\|0\|' = \sup_{\|y\|=1} |y^* 0| = \sup_{\|y\|=1} 0 = 0 \tag{1.144}$$

   Conversely, we suppose

   $$\|x\|' = \sup_{\|y\|=1} |y^* x| = 0 \tag{1.145}$$

   Choosing $y = \dfrac{x}{\|x\|}$ yields

   $$0 = \left| \left(\frac{x}{\|x\|} \right)^* x \right| = \|x\| \tag{1.146}$$

   hence $x = 0$ since $\|\cdot\|$ is a norm.

   (b) We prove triangle inequality holds for $\|\cdot\|'$.

   $$\begin{aligned} \|x + y\|' &= \sup_{\|z\|=1} |z^* (x + y)| && (1.147) \\ &\leq \sup_{\|z\|=1} (|z^* x| + |z^* y|) && (1.148) \\ &\leq \sup_{\|z\|=1} |z^* x| + \sup_{\|z\|=1} |z^* y| && (1.149) \\ &= \|x\|' + \|y\|' && (1.150) \end{aligned}$$

   (c)

   $$\begin{aligned} \|\alpha x\|' &= \sup_{\|z\|=1} |z^* (\alpha x)| && (1.151) \\ &= |\alpha| \sup_{\|z\|=1} |z^* x| && (1.152) \\ &= |\alpha| \|x\|' && (1.153) \end{aligned}$$

Therefore $\|\cdot\|'$ is a norm.

2. Using the lemma, there exists $z_0 \neq 0$ for which

$$|z_0^* x| = \|z_0\|' \|x\| = \|z_0\|' \tag{1.154}$$

Put $z = \dfrac{z_0}{x^* z_0}$, we have

$$\begin{aligned} z^* x &= \left(\frac{z_0}{x^* z_0}\right)^* x && (1.155)\\ &= \frac{z_0^*}{z_0^* x} x && (1.156)\\ &= 1 && (1.157)\end{aligned}$$

Using (1.154), we have

$$\begin{aligned} \|z\|' &= \left\|\frac{z_0}{x^* z_0}\right\|' && (1.158)\\ &= \frac{1}{|x^* z_0|} \|z_0\|' && (1.159)\\ &= \frac{1}{|x^* z_0|} |z_0^* x| && (1.160)\\ &= 1 && (1.161)\end{aligned}$$

Now, we put $B = yz^*$. Using (2.127) yields

$$Bx = yz^* x = y(z^* x) = y \tag{1.162}$$

Using (1.161) yields

$$\begin{aligned} \|B\| &= \sup_{\|t\|=1} \|Bt\| && (1.163)\\ &= \sup_{\|t\|=1} \|y(z^* t)\| && (1.164)\\ &= \sup_{\|t\|=1} (|z^* t| \|y\|) && (1.165)\\ &= \|y\| \sup_{\|t\|=1} |z^* t| && (1.166)\\ &= \|y\| \|z\|' && (1.167)\\ &= \|y\| && (1.168)\\ &= 1 && (1.169)\end{aligned}$$

Finally, we easily prove $rank(B) = 1$. Using familiar rank inequality, we have

$$\begin{aligned} rank(B) &= rank(yz^*) && (1.170)\\ &\leq \min\{rank(y), rank(z^*)\} && (1.171)\\ &= 1 && (1.172)\end{aligned}$$

Hence, $rank(B) = 1$. And B satisfies all conditions of our problem.

Done. □

Problem 1.23. *Prove that the triangle inequality holds for the Frobenius norm.*

Solution. The *Hilbert-Schmidt* or *Frobenius norm* is defined by

$$\|A\|_F = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2 \right)^{\dfrac{1}{2}} \tag{1.173}$$

Observe that this is the same as the 2-norm of the matrix when viewed as an mn-dimensional vector. The 2-norm obviously satisfies triangle inequality. Hence, so Frobenius norm does. □

Problem 1.24. *Let $\|A\|_F$ be the Frobenius norm of matrix $A \in \mathbb{C}^{n\times n}$. Prove that*

$$\|A\|_F = \sqrt{tr\,(A^*A)} = \sqrt{tr\,(AA^*)} \tag{1.174}$$

Solution. We have

$$\begin{aligned} tr\,(A^*A) &= \sum_{i=1}^{n} [A^*A]_{ii} & (1.175)\\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} \overline{a_{ji}} a_{ji} & (1.176)\\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} |a_{ji}|^2 & (1.177)\\ &= \|A\|_F^2 & (1.178) \end{aligned}$$

and

$$tr\,(A^*A) = tr\,(AA^*) \tag{1.179}$$

therefore (1.174) holds. □

1.4 Lecture 4: The Singular Value Decomposition

Problem 1.25 (Exercise 4.1, [1]). *Determine SVDs of the following matrices (by hand calculation).*

$$\begin{bmatrix} 3 & 0 \\ 0 & -2 \end{bmatrix}, \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}, \begin{bmatrix} 0 & 2 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \tag{1.180}$$

Solution. Using svd command in MATLAB, we type

```
[U,S,V] = svd(A)
```

and obtain the following results.

1.

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \tag{1.181}$$

2.

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \tag{1.182}$$

3.

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \tag{1.183}$$

4.

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{2} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix} \tag{1.184}$$

5.

$$\begin{bmatrix} -\frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix} \tag{1.185}$$

Problem 1.26 (Exercise 4.2, [1]). *Suppose A is an $m \times n$ matrix and B is the $n \times m$ matrix obtained by rotating A ninety degrees clockwise on paper (not exactly a standard mathematical transformation!). Do A and B have the same singular values? Prove that the answer is yes or give a counterexample.*

Problem 1.27 (Exercise 4.3, [1]). *Write a* MATLAB *program (see Lecture 9, [1]) which, given a real 2×2 matrix, A plots the right singular vectors v_1 and v_2 in the unit circle and also the left singular vectors u_1 and u_2 in the appropriate ellipse, as in Figure 4.1. Apply your program to the matrix*

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix} \tag{1.186}$$

and also to the 2×2 matrices of Exercise 4.1, [1].

Problem 1.28 (Exercise 4.4, [1]). *Two matrices $A, B \in \mathbb{C}^{m \times m}$ are **unitarily equivalent** f $A = QBQ^*$ for some unitary $Q \in \mathbb{C}^{m \times m}$. Is it true or false that A and B are unitarily equivalent if and only if they have the same singular values?*

Problem 1.29 (Exercise 4.5, [1]). *Theorem 4.1, [1], asserts that every $A \in \mathbb{C}^{m \times n}$ has an SVD $A = U \sum V^*$. Show that if A is real, then it has a real SVD ($U \in \mathbb{R}^{m \times m}, V \in \mathbb{R}^{m \times m}$).*

1.5 Lecture 5: More on the SVD

Problem 1.30 (Exercise 5.1, [1]). *In Example 3.1, [1], we considered the matrix*

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix} \tag{1.187}$$

and asserted, among other things, that its 2-norm is approximately 2.9208*. Using the SVD, work out (on paper) the exact values of* $\sigma_{\min}(A)$ *and* $\sigma_{\max}(A)$ *for this matrix.*

Problem 1.31 (Exercise 5.2, [1]). *Using the SVD, prove that any matrix in* $\mathbb{C}^{m\times n}$ *is the limit of a sequence of matrices of full rank. In other words, prove that the set of full-rank matrices is a dense subset of* $\mathbb{C}^{m\times n}$*. Use the 2-norm for your proof. (The norm doesn't matter, since all norms on a finite-dimensional space are equivalent.)*

Problem 1.32 (Exercise 5.3, [1]). *Consider the matrix*

$$A = \begin{bmatrix} -2 & 11 \\ -10 & 5 \end{bmatrix} \tag{1.188}$$

1. *Determine, on paper, a real SVD of* A *in the form* $A = U\Sigma V^T$*. The SVD is not unique, so find the one that has the minimal number of minus signs in* U *and* V*.*

2. *List the singular values, left singular vectors, and right singular vectors of* A*. Draw a careful, labeled picture of the unit ball in* $\mathbb{R}^2$ *and its image under* A*, together with the singular vectors, with the coordinates of their vertices marked.*

3. *What are the 1-norm, 2-norm,* ∞*-norm and the Frobenius norm of* A*?*

4. *Find* A^{-1} *not directly but via the SVD.*

5. *Find the eigenvalues* λ_1 *and* λ_2 *of* A*.*

6. *Verify that* $\det A = \lambda_1\lambda_2$ *and* $|\det A| = \sigma_1\sigma_2$*.*

7. *What is the area of the ellipsoid onto which* A *maps the unit ball of* $\mathbb{R}^2$*.*

Solution.

1. To find the SVD of the real matrix A, we can find the singular values of A by finding square root of eigenvalues of A^TA or AA^T and the singular vectors by finding the orthonormal eigenvectors of A^T and AA^T respectively.

   Using the following Matlab script

   ```
   A = [-2 11;-10 5];
   eig(A*A')
   ```

yields that the eigenvalues of AA^T are 200 and 50. Hence,

$$\sigma_1 = 10\sqrt{2} \tag{1.189}$$
$$\sigma_2 = 5\sqrt{2} \tag{1.190}$$

We have

$$A^T A = \begin{bmatrix} 104 & -72 \\ -72 & 146 \end{bmatrix} \tag{1.191}$$

To find the eigenvectors of $A^T A$, we solve the linear systems

$$\begin{bmatrix} 104 & -72 \\ -72 & 146 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 200 \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{1.192}$$
$$\begin{bmatrix} 104 & -72 \\ -72 & 146 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = 50 \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \tag{1.193}$$

We obtain

$$x_1 = -\frac{3}{4} x_2 \tag{1.194}$$
$$y_1 = \frac{4}{3} y_2 \tag{1.195}$$

So we choose an orthonormal vectors with the minimal number of minus signs. There clearly is clear one or three minus signs here.

$$v_1 = \begin{bmatrix} -\frac{3}{5} \\ \frac{4}{5} \end{bmatrix} \tag{1.196}$$
$$v_2 = \begin{bmatrix} \frac{4}{5} \\ \frac{3}{5} \end{bmatrix} \tag{1.197}$$

Then

$$u_1 = \frac{1}{\sigma_1} A v_1 \tag{1.198}$$
$$= \frac{1}{10\sqrt{2}} \begin{bmatrix} -2 & 11 \\ -10 & 5 \end{bmatrix} \begin{bmatrix} -\frac{3}{5} \\ \frac{4}{5} \end{bmatrix} \tag{1.199}$$
$$= \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \tag{1.200}$$

$$u_2 = \frac{1}{\sigma_2} A v_2 \tag{1.201}$$
$$= \frac{1}{5\sqrt{2}} \begin{bmatrix} -2 & 11 \\ -10 & 5 \end{bmatrix} \begin{bmatrix} \frac{4}{5} \\ \frac{3}{5} \end{bmatrix} \tag{1.202}$$

$$= \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix} \tag{1.203}$$

Therefore, we have a real SVD of A which has the minimal number of minus signs in U and V as follow

$$A = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 10\sqrt{2} & 0 \\ 0 & 5\sqrt{2} \end{bmatrix} \begin{bmatrix} -\frac{3}{5} & \frac{4}{5} \\ \frac{4}{5} & \frac{3}{5} \end{bmatrix} \tag{1.204}$$

We note that there is two minus signs in this SVD. It is the smallest number of minus signs in SVDs of A. Because there will be more or equal to three minus signs if we choose

$$v_2 = \begin{bmatrix} -\frac{4}{5} \\ -\frac{3}{5} \end{bmatrix} \tag{1.205}$$

in (1.194) and (1.195). Therefore, (4.1) is the best choice.

2. (a) Singular values

$$\sigma_1 = 10\sqrt{2} \tag{1.206}$$

$$\sigma_2 = 5\sqrt{2} \tag{1.207}$$

(b) Left singular vectors

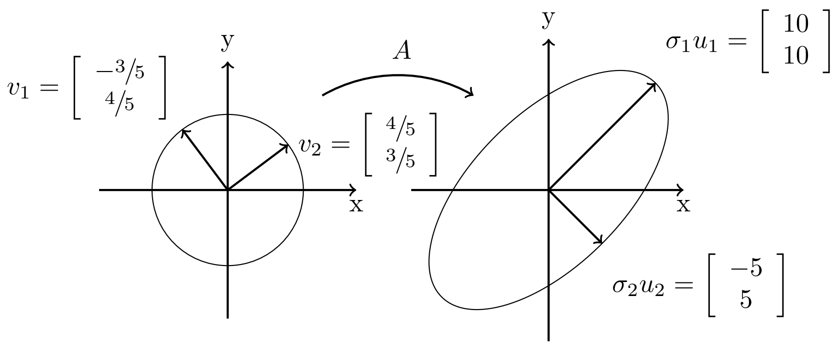
$$u_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \tag{1.208}$$

$$u_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix} \tag{1.209}$$

(c) Right singular vectors

$$v_1 = \begin{bmatrix} -\frac{3}{5} \\ \frac{4}{5} \end{bmatrix} \tag{1.210}$$

$$v_2 = \begin{bmatrix} \frac{4}{5} \\ \frac{3}{5} \end{bmatrix} \tag{1.211}$$


$v_1 = \begin{bmatrix} -3/5 \\ 4/5 \end{bmatrix}$
y
x
A
$v_2 = \begin{bmatrix} 4/5 \\ 3/5 \end{bmatrix}$
y
x
$\sigma_1 u_1 = \begin{bmatrix} 10 \\ 10 \end{bmatrix}$
$\sigma_2 u_2 = \begin{bmatrix} -5 \\ 5 \end{bmatrix}$

3. We denote a_1, a_2 the columns of A and a_1^*, a_2^* the rows of A, we have

$$\|A\|_1 = \max_{i=1,2} \|a_i\|_1 = \|a_2\|_1 = 16 \tag{1.212}$$

$$\|A\|_2 = \max_{i=1,2} \sigma_i = \sigma_1 = 10\sqrt{2} \tag{1.213}$$

$$\|A\|_\infty = \max_{i=1,2} \|a_i^*\|_1 = \|a_2^*\|_1 = 15 \tag{1.214}$$

$$\|A\|_F = \sqrt{\sigma_1^2 + \sigma_1^2} = 5\sqrt{10} \tag{1.215}$$

4. We have the SVD

$$A = U\Sigma V^T \tag{1.216}$$

Therefore,

$$\begin{aligned} A^{-1} &= \left(U\Sigma V^T\right)^{-1} && (1.217)\\ &= V^{-T}\Sigma^{-1}U^{-1} && (1.218)\\ &= V\Sigma^{-1}U^T && (1.219)\\ &= \begin{bmatrix} \frac{1}{20} & -\frac{11}{100} \\ \frac{1}{10} & -\frac{1}{50} \end{bmatrix} && (1.220) \end{aligned}$$

5. Use the following MATLAB code

```
E = eig(A)
```

yields

$$\lambda_{1,2} = \frac{3}{2} \pm \frac{\sqrt{391}}{2} i \tag{1.221}$$

6. Use the following MATLAB code to verify $\det A = \lambda_1\lambda_2$ and $|\det A| = \sigma_1\sigma_2$.

```
det(A)-E(1)*E(2)
abs(det(A))-10*sqrt(2)*5*sqrt(2)
```

7. The lengths of the semi-axes of the ellipsoid onto which A maps the unit ball in $\mathbb{R}^2$ are $\sigma_1 = 10\sqrt{2}, \sigma_2 = 5\sqrt{2}$. Therefore, the area of the ellipsoid is

$$S_{ellipsoid} = \pi\sigma_1\sigma_2 = 100\pi \tag{1.222}$$

Done. □

Problem 1.33 (Exercise 5.4, [1]). *Suppose $A \in \mathbb{C}^{m\times m}$ has an SVD $A = U\Sigma V^*$. Find an eigenvalue decomposition of the $2m \times 2m$ hermitian martrix*

$$M = \begin{bmatrix} 0 & A^* \\ A & 0 \end{bmatrix} \tag{1.223}$$

Solution 1. Let $\sigma_i, i = 1, 2, \ldots, m$ be singular values of A. We have

$$Av_i = \sigma_i u_i, \qquad i = 1, 2, \ldots, m \tag{1.224}$$

$$A^*u_i = \sigma_i v_i, \qquad i = 1, 2, \ldots, m \tag{1.225}$$

Hence,

$$\begin{bmatrix} 0 & A^* \\ A & 0 \end{bmatrix}\begin{bmatrix} v_i \\ u_i \end{bmatrix} = \sigma_i \begin{bmatrix} v_i \\ u_i \end{bmatrix} \tag{1.226}$$

and

$$\begin{bmatrix} 0 & A^* \\ A & 0 \end{bmatrix}\begin{bmatrix} v_i \\ -u_i \end{bmatrix} = -\sigma_i \begin{bmatrix} v_i \\ -u_i \end{bmatrix} \tag{1.227}$$

Due to (5.12) and (5.13), we have a set of eigenvectors of M

$$\frac{1}{\sqrt{2}}\begin{bmatrix} v_i \\ \pm u_i \end{bmatrix}, i = 1, 2, \ldots, m \tag{1.228}$$

corresponding to eigenvalues $\pm\sigma_i$.

We now denote Q the matrix whose columns are the eigenvectors (1.228) of M

$$Q = \frac{1}{\sqrt{2}}\begin{bmatrix} V & V \\ U & -U \end{bmatrix} \tag{1.229}$$

Then

$$Q^* = \begin{bmatrix} V^* & U^* \\ V^* & -U^* \end{bmatrix} \tag{1.230}$$

and

$$QQ^* = \frac{1}{\sqrt{2}}\begin{bmatrix} V & V \\ U & -U \end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix} V^* & U^* \\ V^* & -U^* \end{bmatrix} \tag{1.231}$$

$$= \frac{1}{2}\begin{bmatrix} 2I_m & 0 \\ 0 & 2I_m \end{bmatrix} \tag{1.232}$$

$$= I_{2m} \tag{1.233}$$

Thus, Q is invertible and we have an eigenvalue decomposition of M as follow

$$M = \frac{1}{\sqrt{2}} \begin{bmatrix} V & V \\ U & -U \end{bmatrix} \begin{bmatrix} \Sigma & 0 \\ 0 & -\Sigma \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} V & V \\ U & -U \end{bmatrix}^{-1} \tag{1.234}$$

□

Solution 2. We consider a SVD of A

$$A = U\Sigma V^* \tag{1.235}$$

We denote v_i the columns of V, u_i the columns of U and σ_i the singular values of A. We want to find eigenvalues and eigenvectors of M. The equation

$$\begin{bmatrix} 0 & A^* \\ A & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \lambda \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{1.236}$$

yields

$$A^* x_2 = \lambda x_1 \tag{1.237}$$
$$A x_1 = \lambda x_2 \tag{1.238}$$

Hence,

$$AA^* x_2 = \lambda A x_1 \tag{1.239}$$
$$= \lambda^2 x_2 \tag{1.240}$$

We deduce that x_2 is a left singular vector of A.

Similarly,

$$A^* A x_1 = \lambda A^* x_2 \tag{1.241}$$
$$= \lambda^2 x_1 \tag{1.242}$$

we deduce that x_1 is a right singular vector of A. Thus, we obtain $2m$ eigenvectors of M

$$\frac{1}{\sqrt{2}} \begin{bmatrix} v_i \\ \pm u_i \end{bmatrix}, \qquad i = 1, 2, \ldots, m \tag{1.243}$$

corresponding to the eigenvaluess $\lambda = \pm\sigma_i$. Finally, we get the eigenvalue decomposition

$$M = \frac{1}{\sqrt{2}} \begin{bmatrix} V & V \\ U & -U \end{bmatrix} \begin{bmatrix} \Sigma & 0 \\ 0 & -\Sigma \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} V & V \\ U & -U \end{bmatrix}^{-1} \tag{1.244}$$

□

Problem 1.34. *Let A be a Hermitian matrix. Prove that there exists an unitary matrix Q such that*

$$A = Q\Sigma Q^* \tag{1.245}$$

Solution. We consider a SVD of A

$$A = U\Sigma V^* \tag{1.246}$$

Since A is hermitian, all of its singular values of real, therefore

$$\begin{aligned} A^* &= V\Sigma^* U^* && (1.247)\\ &= V\Sigma U^* && (1.248) \end{aligned}$$

We also have

$$\begin{aligned} A^2 &= AA^* && (1.249)\\ &= U\Sigma V^* V\Sigma U^* && (1.250)\\ &= U\Sigma^2 U^{-1} && (1.251) \end{aligned}$$

Since Σ is a diagonal matrix with all positive entries, we can find A by taking the square of A^2.

$$\begin{aligned} A &= U\Sigma U^{-1} && (1.252)\\ &= U\Sigma U^* && (1.253) \end{aligned}$$

Done. □

Chapter 2

QR Factorization and Least Squares

2.1 Lecture 6: Projectors

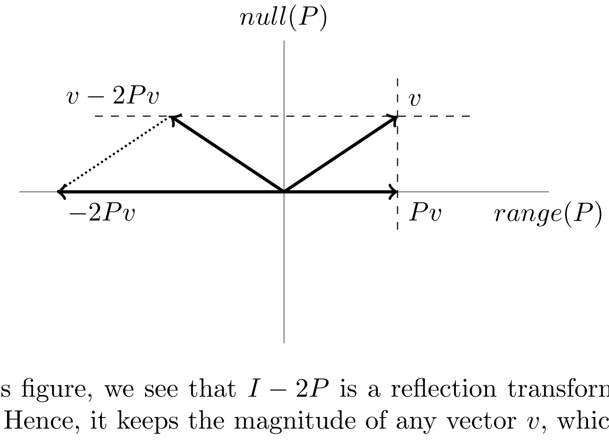
Problem 2.1 (Exercise 6.1, [1]). *If P is an orthogonal projector, then $I-2P$ is unitary. Prove this algebraically, and give a geometric interpretation.*

Solution. Since P is an orthogonal projector, we have $P^2 = P$ and $P = P^*$. Hence

$$\begin{aligned}
(I-2P)^*(I-2P) &= (I-2P^*)(I-2P) && (2.1)\\
&= I - 2P - 2P^* + 4P^*P && (2.2)\\
&= I - 2P - 2P + 4P^2 && (2.3)\\
&= I - 4P + 4P && (2.4)\\
&= I && (2.5)
\end{aligned}$$

Therefore, $I-2P$ is unitary.

Geometric interpretation.



From this figure, we see that $I-2P$ is a reflection transform with respect to $null(P)$. Hence, it keeps the magnitude of any vector v, which is consistent which the fact that $I - 2P$ is unitary. □

Problem 2.2 (Exercise 6.2, [1]). *Let E be the $m \times m$ matrix that extracts the "even parts" of an m-vector: $Ex = \dfrac{x + Fx}{2}$, where F is the $m \times m$ matrix that flips $(x_1, \dots, x_m)^*$ to $(x_m, \dots, x_1)^*$. Is E an orthogonal projector, an oblique projector, or not a projector at all? What are its entries?*

Solution. The matrix F that flips $(x_1, \dots, x_m)^*$ to $(x_m, \dots, x_1)^*$ has the form

$$F = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 1 & 0 \\ \vdots & \vdots & \iddots & \vdots & \vdots \\ 0 & 1 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \end{bmatrix} \tag{2.6}$$

Hence,

$$F^* = F \tag{2.7}$$

$$F^2 = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \iddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \tag{2.8}$$

$$= I_m \tag{2.9}$$

We have

$$Ex = \frac{x + Fx}{2}, \forall x \in \mathbb{C}^m \tag{2.10}$$

or equivalently,

$$E = \frac{I_m + F}{2} \tag{2.11}$$

Hence,

$$E^2 = \left(\frac{I_m + F}{2}\right)^2 \tag{2.12}$$

$$= \frac{I_m + 2F + F^2}{4} \tag{2.13}$$

$$= \frac{I_m + 2F + I_m}{4} \tag{2.14}$$

$$= \frac{I_m + F}{2} \tag{2.15}$$

$$= E \tag{2.16}$$

i.e., E is a projector.

In addition, we also have

$$
\begin{aligned}
E^* &= \left(\frac{I_m + F}{2}\right)^* \tag{2.17}\\
&= \frac{I_m^* + F^*}{2} \tag{2.18}\\
&= \frac{I_m + F}{2} \tag{2.19}\\
&= E \tag{2.20}
\end{aligned}
$$

i.e., E is an orthogonal projector.

Finally, it is straightforward to compute E explicitly.

$$
\begin{aligned}
E &= \frac{I_m + F}{2} \tag{2.21}\\
&= \frac{1}{2}\left(\begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \iddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 1 & 0 \\ \vdots & \vdots & \iddots & \vdots & \vdots \\ 0 & 1 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \end{bmatrix}\right) \tag{2.22}\\
&= \begin{cases} \begin{bmatrix} \frac{1}{2} & 0 & \cdots & \cdots & 0 & \frac{1}{2} \\ 0 & \ddots & \ddots & \iddots & \iddots & 0 \\ \vdots & \ddots & \frac{1}{2} & \frac{1}{2} & \iddots & \vdots \\ \vdots & \iddots & \frac{1}{2} & \frac{1}{2} & \ddots & \vdots \\ 0 & \iddots & \iddots & \ddots & \ddots & 0 \\ \frac{1}{2} & 0 & \cdots & \cdots & 0 & \frac{1}{2} \end{bmatrix}, & \text{if } m \text{ is even} \\ \begin{bmatrix} \frac{1}{2} & 0 & \cdots & 0 & \frac{1}{2} \\ 0 & \ddots & \cdots & \iddots & 0 \\ \vdots & \vdots & 1 & \vdots & \vdots \\ 0 & \iddots & \cdots & \ddots & 0 \\ \frac{1}{2} & 0 & \cdots & 0 & \frac{1}{2} \end{bmatrix}, & \text{if } m \text{ is odd} \end{cases} \tag{2.23}
\end{aligned}
$$

Done. □

Problem 2.3 (Exercise 6.3, [1]). *Given $A \in \mathbb{C}^{m\times n}$ with $m \geq n$, show that A^*A is nonsingular if and only if A has full rank.*

Solution. Since the nonzero singular values of A are the square roots of the nonzero eigenvalues of A^*A, we have that the event A^*A is nonsingular, i.e., $rank\,(A^*A) = n$ and A^*A has n nonzero eigenvalues, is equivalent to the event A has n nonzero singular values, i.e., $rank\,(A) = n$ and A has n nonzero singular values, i.e., A has full rank. □

Problem 2.4 (Exercise 6.4, [1]). *Consider the matrices*

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \tag{2.24}$$

$$B = \begin{bmatrix} 1 & 2 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \tag{2.25}$$

Answer the following questions by hand calculation.

1. *What is the orthogonal projector P onto range (A), and what is the image under P of the vector $(1,2,3)^*$?*
2. *Same questions for B.*

Solution.

1. The orthogonal projector P_A onto *range* (A) is

$$P_A = A(A^*A)^{-1}A^* \tag{2.26}$$

$$= \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix} \tag{2.27}$$

The image under P_A of the vector $v = (1,2,3)^*$ is

$$P_A v = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \tag{2.28}$$

$$= \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix} \tag{2.29}$$

2. Similarly, the orthogonal projector P_B onto *range* (B) is

$$P_B = B(B^*B)^{-1}B^* \tag{2.30}$$

$$= \begin{bmatrix} \frac{5}{6} & \frac{1}{3} & \frac{1}{6} \\ \frac{1}{3} & \frac{1}{3} & -\frac{1}{3} \\ \frac{1}{6} & -\frac{1}{3} & \frac{5}{6} \end{bmatrix} \tag{2.31}$$

The image under P_B of the vector $v = (1,2,3)^*$ is

$$P_B v = \begin{bmatrix} \frac{5}{6} & \frac{1}{3} & \frac{1}{6} \\ \frac{1}{3} & \frac{1}{3} & -\frac{1}{3} \\ \frac{1}{6} & -\frac{1}{3} & \frac{5}{6} \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \tag{2.32}$$

$$= \begin{bmatrix} 2 \\ 0 \\ 2 \end{bmatrix} \tag{2.33}$$

Done. □

Problem 2.5 (Exercise 6.5, [1]). *Let $P \in \mathbb{C}^{m\times m}$ be a nonzero projector. Show that $\|P\|_2 \geq 1$, with equality if and only if P is an orthogonal projector.*

Solution. Since P is an nonzero projector, we have

$$\|P\|_2 = \left\|P^2\right\|_2 \leq \|P\|_2^2 \tag{2.34}$$

Dividing both side of (5.16) by $\|P\|_2 \neq 0$ yields

$$\|P\|_2 \geq 1 \tag{2.35}$$

We now consider the equality of (4.1).

1. ($\Leftarrow$) Suppose P is an orthogonal projector, we have

$$P^2 = P \tag{2.36}$$
$$P = P^* \tag{2.37}$$

Since P is hermitian, the singular values of P are the absolute values of the eigenvalues of P (Theorem 5.5, [1]). Hence,

$$\sigma_i = |\lambda_i| \in \{0, 1\}, \forall i \tag{2.38}$$

Since P is a nonzero projector, there is at least one $\sigma_i = 1$, i.e., (4.3) gives

$$\|P\|_2 = \sigma_1 = 1 \tag{2.39}$$

2. ($\Rightarrow$) We prove by contradiction. Suppose on the contrary that P is an oblique projector, we need to prove $\|P\|_2 > 1$. It suffices to choose an y with $\|y\|_2 \leq 1$ for which

$$\|Py\|_2 > \|y\|_2 \tag{2.40}$$

Since P is an oblique projector, we can choose $u \in \mathit{range}(P)$ and $v \in \mathit{null}(P)$ such that

$$\|u\|_2 = 1 \tag{2.41}$$
$$\|v\|_2 = 1 \tag{2.42}$$
$$\langle u, v\rangle \neq 0 \tag{2.43}$$

(Note that (2.122) can not hold if P is an orthogonal projector).

We now choose

$$z = u - \langle u, v\rangle v \tag{2.44}$$

By (2.122), we then have

$$\begin{aligned}
\|z\|_2^2 &= \|u - \langle u, v\rangle v\|_2^2 && (2.45)\\
&= u^2 - 2\langle u, v\rangle \langle u, v\rangle + \langle u, v\rangle^2 v^2 && (2.46)\\
&= 1 - \langle u, v\rangle^2 < 1 && (2.47)
\end{aligned}$$

i.e., $\|z\|_2 < 1$, whereas

$$\|Pz\|_2 = \|u\|_2 = 1 \tag{2.48}$$

Hence,

$$\begin{aligned}
\|P\|_2 &= \sup_{x\in\mathbb{C}^m, x\neq 0} \frac{\|Px\|_2}{\|x\|_2} && (2.49)\\
&\geq \frac{\|Pz\|_2}{\|z\|_2} && (2.50)\\
&= \frac{1}{\|z\|_2} > 1 && (2.51)
\end{aligned}$$

This contradicts our assumption $\|P\|_2 = 1$. Therefore, P is an orthogonal projector.

Done. □

2.2 Lecture 7: QR Factorization

Problem 2.6 (Exercise 7.1, [1]). *Consider again the matrices A and B of Exercise 6.4, [1].*

$$\begin{aligned}
A &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} && (2.52)\\
B &= \begin{bmatrix} 1 & 2 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} && (2.53)
\end{aligned}$$

1. *Using any method you like, determine (on paper) a reduced QR factorization $A = \hat{Q}\hat{R}$ and a full QR factorization $A = QR$.*

2. *Again using any method you like, determine reduced and full QR factorizations $B = \hat{Q}\hat{R}$ and $B = QR$.*

Solution. Denote a_i and b_i be the columns of A and B respectively.

1. **Reduced QR factorization.**

$$\begin{aligned}
A &= \hat{Q}\hat{R} && (2.54)\\
r_{11} &= \|a_1\|_2 = \sqrt{2} && (2.55)
\end{aligned}$$

$$q_1 = \frac{a_1}{r_{11}} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix} \tag{2.56}$$

$$r_{12} = q_1^* a_2 = 0 \tag{2.57}$$

$$r_{22} = \|a_2 - r_{12} q_1\|_2 = 1 \tag{2.58}$$

$$q_2 = \frac{a_2 - r_{12} q_1}{r_{22}} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \tag{2.59}$$

Hence,

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \tag{2.60}$$

$$= \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 \\ 0 & 1 \\ \frac{1}{\sqrt{2}} & 0 \end{bmatrix} \begin{bmatrix} \sqrt{2} & 0 \\ 0 & 1 \end{bmatrix} \tag{2.61}$$

Full QR factorization. We take $a_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ and find a unit vector q_3 orthogonal to q_1 and q_2.

$$q_3 = \frac{a_3 - q_1^* a_3 q_1 - q_2^* a_3 q_2}{\|a_3 - q_1^* a_3 q_1 - q_2^* a_3 q_2\|_2} \tag{2.62}$$

$$= \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix} \tag{2.63}$$

Hence,

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \tag{2.64}$$

$$= \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \sqrt{2} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \tag{2.65}$$

2. **Reduced QR factorization.**

$$r_{11} = \|b_1\|_2 = \sqrt{2} \tag{2.66}$$

$$q_1 = \frac{b_1}{r_{11}} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix} \tag{2.67}$$

$$r_{12} = q_1^* b_2 = \sqrt{2} \tag{2.68}$$

$$r_{22} = \|b_2 - r_{12} q_1\|_2 = \sqrt{3} \tag{2.69}$$

$$q_2 = \frac{b_2 - r_{12} q_1}{r_{22}} = \begin{bmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} \end{bmatrix} \tag{2.70}$$

Hence,

$$B = \begin{bmatrix} 1 & 2 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \tag{2.71}$$

$$= \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \end{bmatrix} \begin{bmatrix} \sqrt{2} & \sqrt{2} \\ 0 & \sqrt{3} \end{bmatrix} \tag{2.72}$$

Full QR factorization. We take $b_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ and find a unit vector q_3 orthogonal to q_1 and q_2.

$$q_3 = \frac{b_3 - q_1^* b_3 q_1 - q_2^* b_3 q_2}{\|b_3 - q_1^* b_3 q_1 - q_2^* b_3 q_2\|_2} \tag{2.73}$$

$$= \begin{bmatrix} -\frac{1}{\sqrt{6}} \\ \frac{\sqrt{6}}{3} \\ \frac{1}{\sqrt{6}} \end{bmatrix} \tag{2.74}$$

Hence,

$$B = \begin{bmatrix} 1 & 2 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \tag{2.75}$$

$$= \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{6}} \\ 0 & \frac{1}{\sqrt{3}} & \frac{\sqrt{6}}{3} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} \end{bmatrix} \begin{bmatrix} \sqrt{2} & \sqrt{2} \\ 0 & \sqrt{3} \\ 0 & 0 \end{bmatrix} \tag{2.76}$$

Done. □

Problem 2.7 (Exercise 7.2, [1]). *Let A be a matrix with the property that columns $1, 3, 5, 7, \dots$ are orthogonal to columns $2, 4, 6, 8, \dots$. In a reduced QR factorization $A = \hat{Q}\hat{R}$, what special structure does $\hat{R}$ possess?*

Problem 2.8 (Exercise 7.3, [1]). *Let A be an $m \times m$ matrix, and let a_j be its jth column. Give an algebraic proof of* ***Hadamard's inequality****.*

$$|\det A| \leq \prod_{j=1}^{m} \|a_j\|_2 \tag{2.77}$$

Also give a geometric interpretation of this result, making use of the fact that the determinant equals the volume of a parallelepiped.

Problem 2.9 (Exercise 7.4, [1]). *Let $x^{(1)}, y^{(1)}, x^{(2)}$ and $y^{(2)}$ be nonzero vectors in $\mathbb{R}^3$ with the property that $x^{(1)}$ and $y^{(1)}$ are linearly independent and so are $x^{(2)}$ and $y^{(2)}$. Consider the two planes in $\mathbb{R}^3$,*

$$P^{(1)} = \left\langle x^{(1)}, y^{(1)} \right\rangle \tag{2.78}$$

$$P^{(2)} = \left\langle x^{(2)}, y^{(2)} \right\rangle \tag{2.79}$$

Suppose we wish to find a nonzero vector $v \in \mathbb{R}^3$ that lies in the intersection $P = P^{(1)} \cap P^{(2)}$. Devise a method for solving this problem by reducing it to the computation of QR factorizations of three 3×2 matrices.

Problem 2.10 (Exercise 7.5, [1]). *Let A be an $m \times n$ matrix ($m \geq n$) and let $A = \hat{Q}\hat{R}$ be a reduced QR factorization.*

1. *Show that A has rank n if and only if all the diagonal entries of $\hat{R}$ are nonzero.*
2. *Suppose $\hat{R}$ has k nonzero diagonal entries for some k with $0 \leq k < n$. What does this imply about the rank of A? Exactly k? At most k? Give a precise answer, and prove it.*

Solution 1.

1. We have the following *rank inequality.*

Lemma 2.10.1 (Rank Inequality). *For $A \in \mathbb{C}^{m\times n}, B \in \mathbb{C}^{n\times p}$*

$$rank\left(AB\right) \leq \min\left\{rank\left(A\right), rank\left(B\right)\right\} \tag{2.80}$$

Using Lemma 2.10.1 for $\hat{Q}$ and $\hat{R}$ yields

$$\begin{aligned} rank\left(A\right) &= rank\left(\hat{Q}\hat{R}\right) && (2.81)\\ &\leq \min\left\{rank\left(\hat{Q}\right), rank\left(\hat{R}\right)\right\} && (2.82)\\ &\leq n && (2.83)\end{aligned}$$

where the last inequality is deduced from $\hat{R} \in C^{n\times n}$.

Thus,

$$rank\,(A) = n \Leftrightarrow rank\left(\hat{R}\right) = n \tag{2.84}$$

In addition, $\hat{R}$ is an upper-triangular matrix. Therefore, (2.84) means A has rank n if and only if all the diagonal entries of $\hat{R}$ are nonzero.

Solution 2. Since $rank\left(\hat{Q}\right) = n$, we have

$$rank\,(A) = rank\left(\hat{Q}\hat{R}\right) = rank\left(\hat{R}\right) \tag{2.85}$$

Therefore,

$$\begin{aligned} & rank\,(A) = n && (2.86)\\ \Leftrightarrow\ & rank\left(\hat{R}\right) = n && (2.87)\\ \Leftrightarrow\ & \det\left(\hat{R}\right) = \prod_{i=1}^{n} r_{ii} \neq 0 && (2.88)\\ \Leftrightarrow\ & r_{ii} \neq 0, \quad i = 1, 2, \ldots, n && (2.89)\end{aligned}$$

2. Since $\hat{R}$ is an upper-triangular matrix, if it has k nonzero diagonal entries, then it will have at least k linearly indenpendent row vectors. Hence,

$$rank\,(A) = rank\left(\hat{R}\right) \geq k \tag{2.90}$$

Example 2.10.2.

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \tag{2.91}$$

then a reduced QR factorization of A is

$$\hat{Q} = I_2 \tag{2.92}$$

$$\hat{R} = A \tag{2.93}$$

We have $k = 0$ but $rank\,(A) = 1 > 0$.

Done. □

2.3 Lecture 8: Gram-Schmidt Orthogonalization

Problem 2.11 (Exercise 8.1, [1]). *Let A be an $m \times n$ matrix. Determine the exact numbers of floating point additions, subtractions, multiplications, and divisions involved in computing the factorization $A = \hat{Q}\hat{R}$ by Algorithm 8.1, [1].*

Problem 2.12 (Exercise 8.2, [1]). *Write a* Matlab *function `[Q,R] =`*

`mgs(A)` (see Lecture 9, [1]) that computes a reduces QR factorization $A = \hat{Q}\hat{R}$ of an $m \times n$ matrix A with $m \geq n$ using modified Gram-Schmidt orthogonalization. The output variables are a matrix $Q \in \mathbb{C}^{m\times n}$ with orthonormal columns and a triangular matrix $R \in \mathbb{C}^{n\times n}$.

Problem 2.13 (Exercise 8.3, [1]). *Each upper-triangular matrix R_j of p. 61, [1] can be interpreted as the product of a diagonal matrix and a unit upper-triangular matrix (i.e., an upper-triangular matrix with 1 on the diagonal). Explain exactly what these factors are, and which line of Algorithm 8.1, [1], corresponds to each.*

2.4 Lecture 9: MATLAB

Problem 2.14 (Exercise 9.1, [1]).

1. *Run the six-line* MATLAB *program of Experiment 1, [1], to produce a plot of approximate Legendre polynomials.*

2. *For $k = 0, 1, 2, 3$, plot the difference on the 257-point grid between these approximations and the exact polynomials*

$$P_0(x) = 1 \tag{2.94}$$

$$P_1(x) = x \tag{2.95}$$

$$P_2(x) = \frac{3}{2}x^2 - \frac{1}{2} \tag{2.96}$$

$$P_3(x) = \frac{5}{2}x^3 - \frac{3}{2}x \tag{2.97}$$

   How big are the errors, and how are they distributed?

3. *Compare these results with what you get with grid spacings $\Delta = 2^{-\nu}$ for other values of ν. What power of Δx appears to control the convergence?*

Problem 2.15 (Exercise 9.2, [1]). *In Experiment 2, the singular values of A match the diagonal elements of a QR factor R approximately. Consider now a very different example. Suppose $Q = I$ and $A = R$, the $m \times m$ matrix (a **Toeplitz matrix**) with 1 on the main diagonal, 2 on the first superdiagonal, and 0 everywhere else.*

1. *What are the eigenvalues, determinant, and rank of A?*

2. *What is A^{-1}?*

3. *Give a nontrivial upper bound on σ_m, the mth singular value of A. You are welcome to use*MATLAB *for inspiration, but the bound you give should be justified analytically. (Hint: Use part (2).)*

This problem illustrates that you cannot always infer much about the singular values of a matrix from its eigenvalues or from the diagonal entries of a QR factor R.

Problem 2.16 (Exercise 9.3, [1])

1. *Write a*MATLAB *program that sets up a* 15×40 *matrix with entries 0 everywhere except for the values 1 in the position indicated in the picture below. The upper-leftmost 1 is in position* $(2, 2)$, *and the lower-rightmost 1 is in position* $(13, 39)$. *This picture was produced with the command* `spy(A)`.

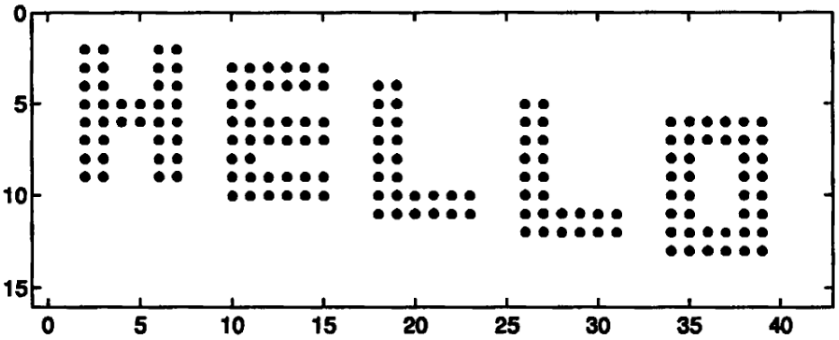


Figure 2.1: *Exercise 9.3, [1].*

2. *Call* `svd` *to compute the singular values of A, and prints the results. Plot these numbers using both* `plot` *and* `semilogy`. *What is the mathematically exact rank of A? How does this show up in the computed singular values?*

3. *For each i from 1 to rank(A), construct the rank-i matrix B that is the best approximation to A in the 2-norm. Use the command* `pcolor(B)` *with* `colormap(gray)` *to create images of these various approximations.*

2.5 Lecture 10: Householder Triangularization

Problem 2.17 (Exercise 10.1, [1]). *Determine the (1) eigenvalues, (2) determinant and (3) singular values of a Householder reflector. For the eigenvalues, give a geometric argument as well as an algebraic proof.*

In the complex case, there is a circle of possible reflections, i.e., the vector x can be reflected to $z \|x\| e_1$, where z is any scalar with $|z| = 1$. And in the real case, there are two alternatives, with $z = \pm 1$ in $z \|x\| e_1$. Thus, we just need to handle this problem in the complex case generally.

Now, in the complex case, suppose we are constructing unitary matrices Q_k in Householder method. Each Q_k is chosen to be a unitary matrix of the form

$$Q_k = \begin{bmatrix} I & 0 \\ 0 & F \end{bmatrix} \tag{2.98}$$

where I is the $(k-1) \times (k-1)$ identity and F is an $(m-k+1) \times (m-k+1)$ unitary matrix.

Suppose, at the beginning of step k, the entries $k, \ldots, m$ of the kth column are given by the vector $x \in \mathbb{C}^{m-k+1}$. The Householder reflector F effects the following map.

$$F : x \mapsto Fx = \|x\| e_1 \tag{2.99}$$

Explicitly, the Householder reflector F is

$$F = I - 2\frac{vv^*}{v^*v} \tag{2.100}$$

where $v = z \|x\| e_1 - x$.

Solution. We compute eigenvalues, determinant and singular values of F in turns. First of all, we can easily choose z such that $v \neq 0$.
1. If $v^*u = 0$, i.e., u is perpendicular to v, then

$$\begin{align} Fu &= u - 2v\frac{v^*u}{v^*v} \tag{2.101}\\ &= u \tag{2.102}\end{align}$$

Hence, F has $m-k$ eigenvalues 1 since the subspace of $\mathbb{C}^{m-k+1}$ that is orthogonal to $v \neq 0$ has dimension $m - k$.

In addition, we also have v is an eigenvector of F itself. Indeed,

$$\begin{align} Fv &= v - 2v\frac{v^*v}{v^*v} \tag{2.103}\\ &= v - 2v \tag{2.104}\\ &= -v \tag{2.105}\end{align}$$

Thus, the last eigenvalue of F is -1.

Geometric interpretation. The reflection of v is $-v$, and reflection of any vector perpendicular to v is v itself.

2. The determinant of F is

$$\begin{align} \det F &= \prod_{i=1}^{m-k+1} \lambda_i \tag{2.106}\\ &= (-1)\, 1^{m-k} \tag{2.107}\\ &= -1 \tag{2.108}\end{align}$$

3. Since F is unitary, i.e., $F^*F = I$, using the Theorem 5.4, [1], we deduce that all singular values of A equal to 1. □

Problem 2.18 (Exercise 10.2, [1]).

1. *Write a* Matlab *function* `[W,R] = house(A)` *that computes an implicit representation of a full QR factorization $A = QR$ of an $m \times n$ matrix A with $m \geq n$ using Householder reflections. The output variables are a lower-triangular matrix $W \in \mathbb{C}^{m\times n}$ whose columns are the vectors v_k defining the successive Householder reflections, and a triangular matrix $R \in \mathbb{C}^{n,n}$.*

2. *Write a* MATLAB *function* `Q = formQ(W)` *that takes the matrix W produced by* `house` *as input and generates a corresponding $m \times m$ orthogonal matrix Q.*

Problem 2.19 (Exercise 10.3, [1]). *Let Z be the matrix*

$$Z = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 7 \\ 4 & 2 & 3 \\ 4 & 2 & 2 \end{bmatrix} \tag{2.109}$$

Compute three reduced QR factorizations of Z in MATLAB*: by the Gram-Schmidt routine* `mgs` *of Exercise 8.2, [1], by the Householder routines* `house` *and* `formQ` *of Exercise 10.2, [1], and by* MATLAB*'s built-in command* `[Q,R] = qr(Z,0)`*. Compare these three and comment on any differences you see.*

Problem 2.20 (Exercise 10.4, [1]). *Consider the 2×2 orthogonal matrices*

$$F = \begin{bmatrix} -c & s \\ s & c \end{bmatrix} \tag{2.110}$$

$$J = \begin{bmatrix} c & s \\ -s & c \end{bmatrix} \tag{2.111}$$

where $s = \sin\theta_0$ and $c = \cos\theta_0$ for some θ_0. The first matrix has $\det F = -1$ and is a reflector - the special case of a Householder reflector in dimension 2. The second has $\det J = 1$ and effects a rotation instead of a reflection. Such a matrix is called a ***Givens rotations****.*

1. *Describe exactly what geometric effects left-multiplications by F and J have on the plane $\mathbb{R}^2$. (J rotates the plane by the angle θ, for example, but is the rotation clockwise or counterclockwise?)*

2. *Describe an algorithm for QR factorization that is analogous to Algorithm 10.1, [1]-p.73, but based on Givens rotations instead of Householder reflections.*

3. *Show that your algorithm involves six flops per entry operated on rather than four, so that the asymptotic operation count is 50% greater than*

$$\text{Work for Householder orthogonalization: } \sim 2mn^2 - \frac{2}{3}n^3 \text{ flops.}$$

SOLUTION 1. We now solve **1** and **2** through trigonometric view points since it is a good idea to use trigonometric in describing geometric effects in [1].
1. We consider $\mathbb{R}^2$ in polar coordinate for simplicity.

$$F \begin{bmatrix} r\sin\alpha \\ r\cos\alpha \end{bmatrix} = \begin{bmatrix} -c & s \\ s & c \end{bmatrix} \begin{bmatrix} r\sin\alpha \\ r\cos\alpha \end{bmatrix} \tag{2.112}$$

$$= r \begin{bmatrix} \sin\theta_0 \cos\alpha - \cos\theta_0 \sin\alpha \\ \sin\theta_0 \sin\alpha + \cos\theta_0 \cos\alpha \end{bmatrix} \tag{2.113}$$

$$= \begin{bmatrix} r\sin(\theta_0 - \alpha) \\ r\cos(\theta_0 - \alpha) \end{bmatrix} \tag{2.114}$$

Geometric interpretation of F. The left-multiplication by F will reflect the plane $\mathbb{R}^2$ across the hyperplane

$$\theta = \frac{\theta_0}{2} \tag{2.115}$$

We handle J similarly.

$$J\begin{bmatrix} r\sin\alpha \\ r\cos\alpha \end{bmatrix} = \begin{bmatrix} c & s \\ -s & c \end{bmatrix}\begin{bmatrix} r\sin\alpha \\ r\cos\alpha \end{bmatrix} \tag{2.116}$$

$$= r\begin{bmatrix} \cos\theta_0\sin\alpha + \sin\theta_0\cos\alpha \\ \cos\theta_0\cos\alpha - \sin\theta_0\sin\alpha \end{bmatrix} \tag{2.117}$$

$$= \begin{bmatrix} r\sin(\alpha + \theta_0) \\ r\cos(\alpha + \theta_0) \end{bmatrix} \tag{2.118}$$

Geometric interpretation of J. The left-multiplication by F will rotate the plane $\mathbb{R}^2$ by the angle θ_0. In particular, if $\theta_0 > 0$, this rotation is counter-clockwise and if $\theta_0 < 0$, this rotation is clockwise. Of course, if $\theta_0 = 0$ this rotation is identity operator.

2. We focus on performing elimination under a single column of A, which we then repeat for each column. For Householder, this is done by a single Householder rotation. Since we are using 2×2 rotations, we have to eliminate under a column one number at a time. Given 2-component vector $x = \begin{bmatrix} r\sin\alpha \\ r\cos\alpha \end{bmatrix}$, we need

$$J : x \mapsto Jx = \begin{bmatrix} \|x\|_2 \\ 0 \end{bmatrix} \tag{2.119}$$

We now use (2.119) to find givens rotation J explicitly.

$$\begin{bmatrix} r \\ 0 \end{bmatrix} = J\begin{bmatrix} r\sin\alpha \\ r\cos\alpha \end{bmatrix} \tag{2.120}$$

$$= \begin{bmatrix} r\sin(\alpha + \theta_0) \\ r\cos(\alpha + \theta_0) \end{bmatrix} \tag{2.121}$$

Thus, we obtain a system of equations

$$r\sin(\alpha + \theta_0) = r \tag{2.122}$$

$$r\cos(\alpha + \theta_0) = 0 \tag{2.123}$$

If $r = 0$, take J is an arbitrary givens rotations, i.e., θ_0 can be taken arbitrarily. If $r \neq 0$, we can choose

$$\theta_0 = \frac{\pi}{2} - \alpha \tag{2.124}$$

to make (2.122)-(2.123) hold.

Describe an algorithm for QR factorization for given rotations. Then we just do this working bottom-up from the column: rotate the bottom two rows to introduce one zero, the the next two rows to introduce a second zero, etc.

SOLUTION 2. We now solve **2** and **3** as usual. This approach is good for counting flops in **3**.

2. Instead of trigonometric notation, we can take $x = \begin{bmatrix} a \\ b \end{bmatrix}$ as usual. Then (2.119) becomes

$$\begin{aligned} \begin{bmatrix} \sqrt{a^2+b^2} \\ 0 \end{bmatrix} &= J \begin{bmatrix} s \\ b \end{bmatrix} && (2.125) \\ &= \begin{bmatrix} a\cos\theta_0 + b\sin\theta_0 \\ -a\sin\theta_0 + b\cos\theta_0 \end{bmatrix} && (2.126) \end{aligned}$$

Thus, we obtain a system of equations

$$\begin{aligned} a\cos\theta_0 + b\sin\theta_0 &= \sqrt{a^2+b^2} && (2.127) \\ -a\sin\theta_0 + b\cos\theta_0 &= 0 && (2.128) \end{aligned}$$

Solving (2.127)-(2.128) out yields

$$\begin{aligned} \cos\theta_0 &= \frac{a}{\sqrt{a^2+b^2}} && (2.129) \\ \sin\theta_0 &= \frac{b}{\sqrt{a^2+b^2}} && (2.130) \end{aligned}$$

Some easy computation yields

$$\begin{aligned} \tan^2\theta_0 &= \frac{1}{\cos^2\theta_0} - 1 && (2.131) \\ &= \frac{a^2+b^2}{a^2} - 1 && (2.132) \\ &= \frac{b^2}{a^2} && (2.133) \end{aligned}$$

Hence, we can choose

$$\theta_0 = \arctan\frac{b}{a} \tag{2.134}$$

and the description of an algorithm for QR factorization for givens rotations is the same as the above argument.

3. To multiply J by a single 2-component vector requires mulplications and 2 addition, or 6 flops, as indicated below

$$J \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} a \bullet \cos\left(\arctan\frac{b}{a}\right) + b \bullet \sin\left(\arctan\frac{b}{a}\right) \\ -a \bullet \sin\left(\arctan\frac{b}{a}\right) + b \bullet \cos\left(\arctan\frac{b}{a}\right) \end{bmatrix} \tag{2.135}$$

That is 6 flops per column vector element of the matrix. Whereas Householder requires 4 flops per column vector element, see [1]-p.59. Therefore, 6 flops of givens rotations is 50% more than 4 flops of Householder. □

2.6 Lecture 11: Least Squares Problems

Problem 2.21 (Exercise 11.1, [1]). *Suppose the $m \times n$ matrix A has the form*

$$A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} \tag{2.136}$$

where A_1 is a nonsingular matrix of dimension $n \times n$ and A_2 is an arbitrary matrix of dimension $(m-n) \times n$. Prove that

$$\left\|A^+\right\|_2 \leq \left\|A_1^{-1}\right\|_2 \tag{2.137}$$

Solution. Under the orthogonal projector $P \in \mathbb{C}^{m \times m}$ that maps $\mathbb{C}^m$ onto range (A), any $x \in \mathbb{C}^m$ can be decomposed uniquely as $x = Px + (x - Px)$, where $Px \in \text{range}\,(A)$ and $x - Px \in \text{null}\,(A)$.

Now, we have

$$(x - Px)^* Ay = 0, \forall y \in \mathbb{C}^n \tag{2.138}$$

since $Ay \in \text{range}\,(A)$.

Applying the Hermitian operator on both side of (2.138) yields

$$y^* A^* (x - Px) = 0, \forall y \in \mathbb{C}^n \tag{2.139}$$

Hence,

$$A^* (x - Px) = 0 \tag{2.140}$$

Using (5.5), we estimate the 2-norm of the pseudoinverse matrix A^+ of A as follow

$$\begin{aligned}
\left\|A^+\right\|_2 &= \sup_{x \in \mathbb{C}^m, x \neq 0} \frac{\left\|A^+ x\right\|_2}{\|x\|_2} && (2.141) \\
&= \sup_{x \in \mathbb{C}^m, x \neq 0} \frac{\left\|(A^*A)^{-1} A^* (Px + (x - Px))\right\|_2}{\|Px + (x - Px)\|_2} && (2.142) \\
&\leq \sup_{x \in \mathbb{C}^m, x \neq 0} \frac{\left\|(A^*A)^{-1} A^* Px\right\|_2}{\|Px\|_2} && (2.143) \\
&= \sup_{y \in \mathbb{C}^m, y \neq 0} \frac{\left\|(A^*A)^{-1} A^* Ay\right\|_2}{\|Ay\|_2} && (2.144) \\
&= \sup_{y \in \mathbb{C}^m, y \neq 0} \frac{\|y\|_2}{\|Ay\|_2} && (2.145)
\end{aligned}$$

$$
\begin{aligned}
&\leq \sup_{y\in\mathbb{C}^m, y\neq 0} \frac{\|y\|_2}{\sqrt{\|A_1y\|_2^2+\|A_2y\|_2^2}} && (2.146)\\
&\leq \sup_{y\in\mathbb{C}^m, y\neq 0} \frac{\|y\|_2}{\|A_1y\|_2} && (2.147)\\
&= \sup_{y\in\mathbb{C}^m, y\neq 0} \frac{\left\|A_1^{-1}y\right\|_2}{\|y\|_2} && (2.148)\\
&= \left\|A_1^{-1}\right\|_2^2 && (2.149)
\end{aligned}
$$

where the equality (2.144) is deduced from the fact that there exists an nonzero vector $y \in \mathbb{C}^n$ such that $Px = Ay$ since $Px \in \text{range}(A)$.

We also have used the following estimate

$$
\begin{aligned}
\|Px + (x-Px)\|_2^2 &= \|Px\|_2^2 + \|x-Px\|_2^2 && (2.150)\\
&\geq \|Px\|_2^2 && (2.151)
\end{aligned}
$$

The estimate (5.12)-(5.13) is deduce from applying the Pythagorean theorem for two orthogonal vector $Px\perp(x-Px)$. Done. □

Problem 2.22 (Exercise 11.2, [1]).

1. *How closely, as measured in the L^2 norm on the interval $[1,2]$, can the function $f(x) = x^{-1}$ be fitted by a linear combination of the function $e^x, \sin x$ and $\Gamma(x)$? ($\Gamma(x)$ is the gamma function, a built-in function in* MATLAB.*) Write a program that determines the answer to at least two digits of relative accuracy using the discretization of $[1,2]$ and a discrete least squares problem. Write down your estimate of the answer and also of the coefficients of the optimal linear combination, and produce a plot of the optimal approximation.*

2. *Now repeat, but with $[1,2]$ replaced by $[0,1]$, You may find the following fact helpful: if $g(x) = \dfrac{1}{\Gamma(x)}$ then $g'(0) = 1$.*

SOLUTION.

1. Using the following MATLAB code

```
close all
clear all
clc
format long

% Trefethen & Bau, Exercise 11.2.
N = 300;
for n = 2:N
   % set up least square problem.
   A = zeros(n,3);
   x = linspace(1,2,n);
   A(:,1) = exp(x');
   A(:,2) = sin(x');
```

```
    A(:,3) = gamma(x');
    b = 1./x';
    % solve least square problem.
    [Q,R] = qr(A);
    sol = R\(Q'*b);
    % compute L^2 norm of error.
    error(n) = sqrt(quad(@(z)F(z,sol(1),sol(2),sol(3)),1,2));
    % compute discrete L^2 norm of error.
    discrete_error(n) = sqrt((norm(b-sol(1)*A(:,1)- ...
                        sol(2)*A(:,2)-sol(3)*A(:,3)))^2/n);
end
plot(2:N,error(2:N),2:N,discrete_error(2:N));
xlabel('Number of Data Points');
ylabel('Errors');
title('Errors of Least Square Problems');
legend('L^2 norm','discrete L^2 norm');
print('-r300','-djpeg');
```

and subroutine

```
function y = F(z,c1,c2,c3)
y = (1./z-c1*exp(z)-c2*sin(z)-c3*gamma(z)).^2;
```

yields

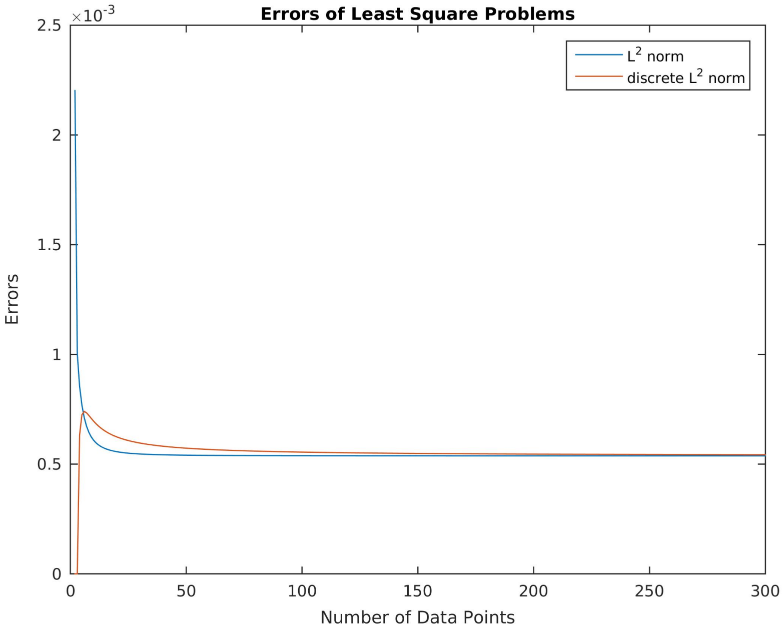


Figure 2.2: *The L^2 norm of error on the interval* $[1,2]$.

2. Using the following MATLAB code

```
close all
clear all
clc
format long

% Trefethen & Bau, Exercise 11.2.
N = 200;
tol = 1e-13;
for n = 2:N
   % set up least square problem.
   A = zeros(n,3);
   x = linspace(tol,1,n);
   A(:,1) = exp(x');
   A(:,2) = sin(x');
   A(:,3) = gamma(x');
   b = 1./x';
   % solve least square problem.
   [Q,R] = qr(A);
   sol = R\(Q'*b);
   % compute L^2 norm of error.
   error(n) = sqrt(quad(@(z)F(z,sol(1),sol(2),sol(3)),tol,1));
   % compute discrete L^2 norm of error.
   discrete_error(n) = sqrt((norm(b-sol(1)*A(:,1)- ...
                       sol(2)*A(:,2)-sol(3)*A(:,3)))^2/n);
end
plot(2:N,error(2:N),2:N,discrete_error(2:N));
xlabel('Number of Data Points');
ylabel('Errors');
title('Errors of Least Square Problems');
legend('L^2 norm','discrete L^2 norm');
print('-r300','-djpeg');
```

and subroutine

```
function y = F(z,c1,c2,c3)
y = (1./z-c1*exp(z)-c2*sin(z)-c3*gamma(z)).^2;
```

yields

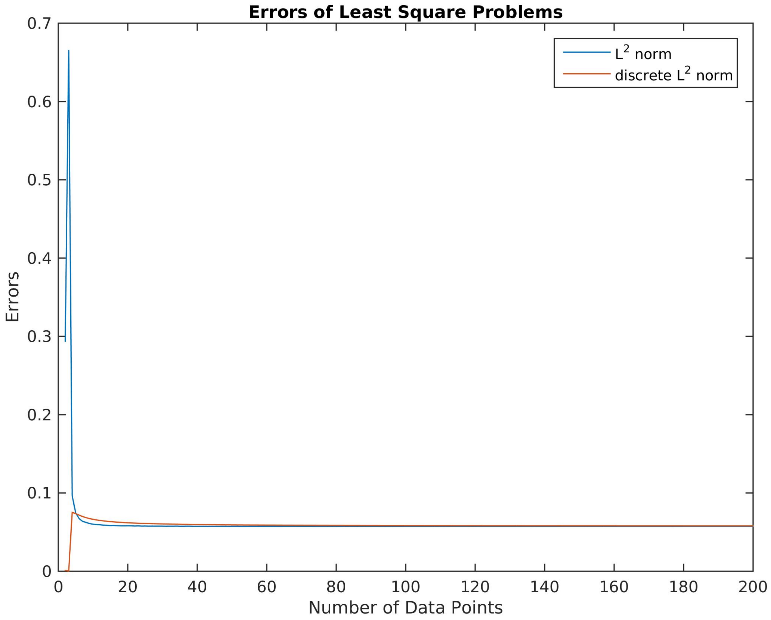


Figure 2.3: *The* L^2 *norm of error on the interval* $[10^{-13}, 1]$.

If you use `tol=1e-14`, this MATLAB code returns

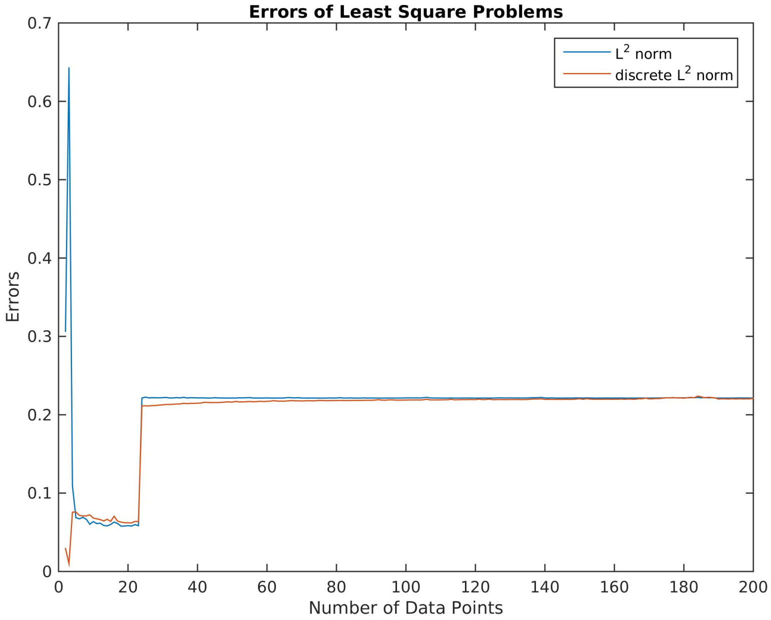


Figure 2.4: *The* L^2 *norm of error on the interval* $[10^{-14}, 1]$.

If you use `tol=1e-15`, this MATLAB code returns

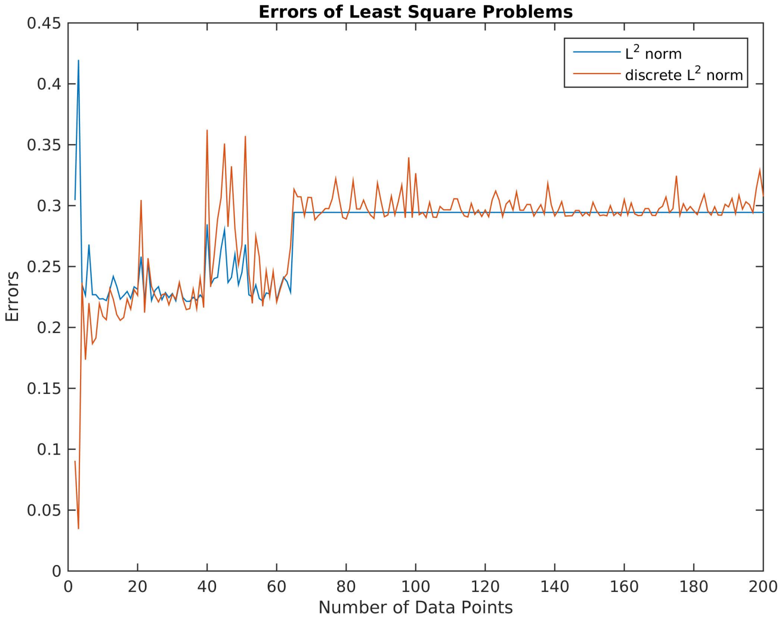


Figure 2.5: *The L^2 norm of error on the interval* $\left[10^{-15}, 1\right]$.

If you use `tol=1e-16`, this MATLAB code returns

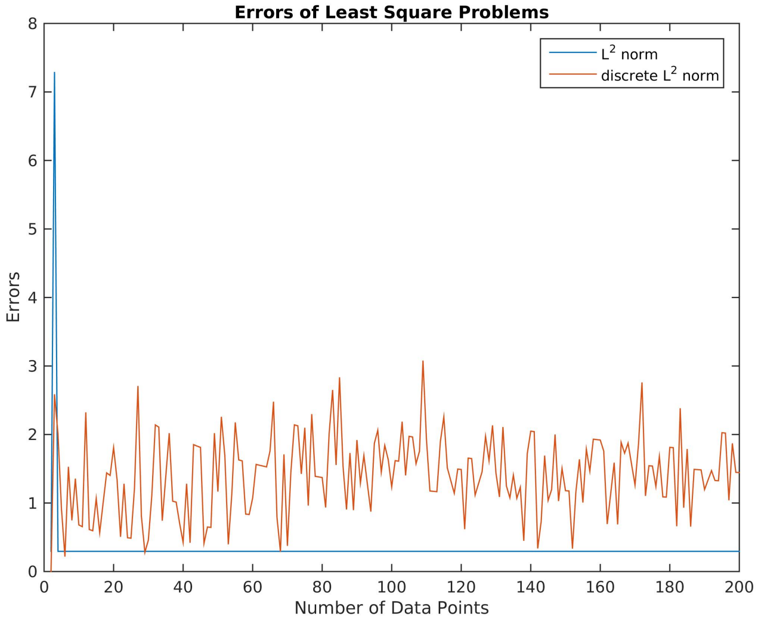


Figure 2.6: *The L^2 norm of error on the interval $\left[10^{-16}, 1\right]$.*

Done. □

Problem 2.23 (Exercise 11.3, [1]). *Take $m = 50, n = 12$. Using* MATLAB*'s* `linspace`*, define t to be the m-vector corresponding to linearly spaced grid points from 0 to 1. Using* MATLAB*'s* `vander` *and* `fliplr`*, define A to be the $m \times n$ matrix associated with least squares fitting on this grid by a polynomial of degree $n-1$. Take b to the function $\cos(4t)$ evaluated on the grid. How, calculate and print (to sixteen-digit precision) the least squares coefficients vector x by six methods:*

1. *Formation and solution of the normal equations, using* MATLAB*'s*`\`.

2. *QR factorization computed by* `mgs` *(modified Gram-Schmidt, Exercise 8.2, [1]).*

3. *QR factorization computed by* `house` *(Householder triangularization, Exercise 10.2, [2]).*

4. *QR factorization computed by* MATLAB*'s* `qr` *(also Householder triangularization).*

5. `x = A\b` *in* MATLAB *(also based on QR factorization).*

6. *SVD, using* MATLAB*'s* `svd`*.*

7. *The calculations above will produce six lists of twelve coefficients. In each list, shade with red pen the digits that appear to be wrong (affected by*

rounding error). Comment on what differences you observe. Do the normal equations exhibit instability? You do not have to explain your observations.

Chapter 3

Conditioning and Stability

3.1 Lecture 12: Conditioning and Condition Numbers

Problem 3.1 (Exercise 12.1, [1]). *Suppose A is a 202×202 matrix with $\|A\|_2 = 100$ and $\|A\|_F = 101$. Give the sharpest possible lower bound on the 2-norm condition number $\kappa(A)$.*

Problem 3.2 (Exercise 12.2, [1]). *In Example 11.1, [1], we remarked that polynomial interpolation in equispaced points is ill-conditioned. To illustrate this phenomenon, let $x_1, \ldots, x_n$ and $y_1, \ldots, y_m$ be n and m equispaced points from -1 to 1, respectively.*

1. *Derive a formula for the $m \times n$ matrix A that maps an n-vector of data at $\{x_j\}$ to an m-vector of sampled values $\{p(y_j)\}$, where p is the degree $n-1$ polynomial interpolant of the data (see Example 1.1, [1]).*

2. *Write a program to calculate A and plot $\|A\|_\infty$ on a semilog scale for $n = 1, 2, \ldots, 30, m = 2n-1$. In the continuous limit $m \to \infty$, the number $\|A\|_\infty$ are known as the **Lebesgue constants** for equispaced interpolation, which are asymptotic to*

$$\frac{2^n}{e(n-1)\log n} \tag{3.1}$$

as $n \to \infty$.

3. *For $n = 1, 2, \ldots, 30$ and $m = 2n-1$, what is the ∞-norm condition number κ of the problem of interpolating the constant function 1? Use*

$$\kappa = \frac{\|J(x)\| \|x\|}{\|f(x)\|} \tag{3.2}$$

4. *How close is your result for $n = 11$ to the bound implicit in the following Figure.*

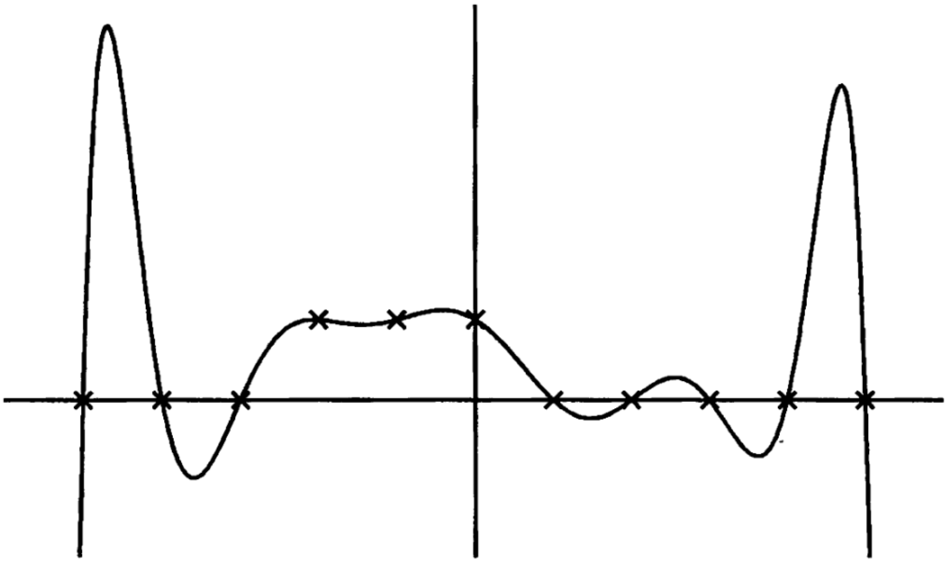


Figure 3.1: *Degree 10 polynomial interpolant to eleven data points. The axis scales are not given, as these have no effect on the picture.*

Problem 3.3 (Exercise 12.3, [1]). *The goal of this problem is to explore some properties of random matrices. Your job is to be a laboratory scientist, performing experiments that lead to conjectures and more refined experiments. Do not try to prove any thing. Do produce well-designed plots, which are worth a thousand numbers.*

*Define a **random matrix** to be an $m \times m$ matrix whose entries are independent samples from the real normal distribution with mean zero and standard derivation $\frac{1}{\sqrt{m}}$. (In* MATLAB, `A = randn(m,m)/sqrt(m)`*.) The factor $\sqrt{m}$ is introduced to make the limiting behavior clean as $m \to \infty$.*

1. *What do the eigenvalues of a random matrix look like? What happens, say, if you take 100 random matrices and superimpose all their eigenvalues in a single plot? If you do this for $m = 8, 16, 32, 64, \ldots$, what pattern is suggested? How does the spectral radius $\rho(A)$ (Exercise 3.2, [1]) behave as $m \to \infty$?*

2. *What about norms? How does the 2-norm of a random matrix behave as $m \to \infty$? Of course, we must have $\rho(A) \leq \|A\|$ (Exercise 3.2, [1]). Does this inequality appear to approach an equality as $m \to \infty$?*

3. *What about condition numbers-or more simply, the smallest singular value $\sigma_{\min}$? Even for fixed m this equation is interesting. What proportions of random matrices in $\mathbb{R}^{m\times m}$ seem to have $\sigma_{\min} \leq \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots$? In other words, what does the tail of the probability distribution of smallest singular values look like? How does the scale of all this change with m?*

4. *How do the answer to (1)-(3) change if we consider random triangular*

instead of full matrices, i.e., upper-triangular matrices whose entries are samples from the same distribution as above?

3.2 Lecture 13: Floating Point Arithmetic

Problem 3.4 (Exercise 13.1, [1]). *Between an adjacent pair of nonzero IEEE single precision real numbers, how many IEEE double precision numbers are there?*

Problem 3.5 (Exercise 13.2, [1]). *The floating point system $\mathbb{F}$ defined by*

$$x = \pm \left(m/\beta^t\right) \beta^e \tag{3.3}$$

includes many integers, but not all of them.

1. *Give an exact formula for the smallest positive integer n that does not belong to $\mathbb{F}$.*

2. *In particular, what are the values of n for IEEE single and double precision arithmetic?*

3. *Figure out a way to verify this result for your own computer. Specifically, design and run a program that produces evidence that $n-3, n-2,$ and $n-1$ belong to $\mathbb{F}$ but n does not. What about $n+1, n+2, n+3$?*

Problem 3.6 (Exercise 13.3, [1]). *Consider the polynomial*

$$\begin{aligned} p(x) &= (x-2)^9 && (3.4)\\ &= x^9 - 18x^8 + 144x^7 - 672x^6 + 2016x^5 - 4032x^4 && (3.5)\\ &\quad + 5376x^3 - 4608x^2 + 2304x - 512 && (3.6) \end{aligned}$$

1. *Plot $p(x)$ for $x = -1.920, -1.919, -1.918, \ldots, 2.080,$ evaluating p via its coefficients $1, -18, 144, \ldots$.*

2. *Produce the same plot again, now evaluating p via the expression $(x-2)^9$.*

Problem 3.7 (Exercise 13.4, [1]). *The polynomial $p(x) = x^5 - 2x^4 - 3x^3 + 3x^2 - 2x - 1$ has three real zeros. Applying Newton's method to p with initial guess $x_0 = 0$ produces a series of estimates $x_1, x_2, x_3, \ldots$ that converge rapidly to a zero $x_* \approx -0.315$.*

1. *Compute $x_1, \ldots, x_6$ in floating point arithmetic with $\epsilon_{machine} \approx 10^{-16}$. How many digits do you estimate are correct in each of these numbers?*

2. *Compute $x_1, \ldots, x_6$ again exactly with the aid of a symbolic algebra system such as* MAPLE *or*MATHEMATICA. *Each x_j is a rational number. How many digits are there in the numerator and the denominator for each j?*

3.3 Lecture 14: Stability

Problem 3.8 (Exercise 14.1, [1]). *True of False?*

1. $\sin x = O(1)$ *as* $x \to \infty$.

2. $\sin x = O(1)$ *as* $x \to 0$.

3. $\log x = O\left(x^{\frac{1}{100}}\right)$ *as* $x \to \infty$.

4. $n! = O\left(\left(\frac{n}{e}\right)^n\right)$ *as* $n \to \infty$.

5. $A = O\left(V^{\frac{2}{3}}\right)$ *as* $V \to \infty$*, where* A *and* V *are the surface area and volume of a sphere measured in square miles and cubic microns, respectively.*

6. $fl(\pi) - \pi = O(\epsilon_{machine})$. *(We dot not mention that the limit is* $\epsilon_{machine} \to 0$*, since that is implicit for all expressions* $O(\epsilon_{machine})$ *in this book.*

7. $fl(n\pi) - n\pi = O(\epsilon_{machine})$*, uniformly for all integers* n*. (Here* $n\pi$ *represents the exact mathematical quantity, not the result of a floating point calculation.*

Problem 3.9 (Exercise 14.2, [1]).

1. *Show that*

$$(1 + O(\epsilon_{machine}))(1 + O(\epsilon_{machine})) = 1 + O(\epsilon_{machine}) \tag{3.7}$$

The precise meaning of this statement is that if f *is a function satisfying*

$$f(\epsilon_{machine}) = (1 + O(\epsilon_{machine}))(1 + O(\epsilon_{machine})) \text{ as } \epsilon_{machine} \to 0 \tag{3.8}$$

then f *also satisfies*

$$f(\epsilon_{machine}) = 1 + O(\epsilon_{machine}) \text{ as } \epsilon_{machine} \to 0 \tag{3.9}$$

2. *Show that*

$$(1 + O(\epsilon_{machine}))^{-1} = 1 + O(\epsilon_{machine}) \tag{3.10}$$

3.4 Lecture 15: More on Stability

Problem 3.10 (Exercise 15.1, [1]). *Each of the following problems describes an algorithm implemented on a computer satisfying the axioms*
Axiom 3.10.1.

$$\forall x \in \mathbb{R}, \exists \epsilon, |\epsilon| \le \epsilon_{machine}, fl(x) = x(1+\epsilon) \tag{3.11}$$

Fundamental Axiom of Floating Point Arithmetic 3.10.2.

$$\forall x \in F, \forall y \in F, \exists \epsilon, |\epsilon| \le \epsilon_{machine}, x \circledast y = (x * y)(1+\epsilon) \tag{3.12}$$

For each one, state whether the algorithm is ***backward stable, stable but not backward stable*** *or* ***unstable****, and prove it or at least give a reasonably convincing argument. Be sure to follow the definitions as given in the text.*

1. *Data:* $x \in \mathbb{C}$. *Solution:* $2x$, *computed as* $x \oplus x$.
2. *Data:* $x \in \mathbb{C}$. *Solution:* x^2, *computed as* $x \otimes x$.
3. *Data:* $x \in \mathbb{C}\backslash \{0\}$. *Solution: 1, computed as* $x \bigcirc\!\!\!\!\!\div\; x$. *(A machine satisfying*

$$x \circledast y = fl\,(x * y) \tag{3.13}$$

   will give exactly the right answer, but our definitions are based on the weaker condition (3.12).*)*

4. *Data:* $x \in \mathbb{C}$. *Solution: 0, computed as* $x\ominus$. *(Again, a real machine may do better than our definitions based on* (3.12).*)*
5. *Data: none. Solution: e, computed by summing* $\sum_{k=0}^{\infty} \frac{1}{k!}$ *from left to right using* $\otimes$ *and* $\oplus$, *stopping when a summand is reached of magnitude* $< \epsilon_{machine}$.
6. *Data: none. Solution: e, computed by the same algorithm as above except with the series summed from right to left.*
7. *Data: none. Solution:* π, *computed by doing an exhaustive search to find the smallest floating point number* x *in the interval* $[3,4]$ *such that*

$$s\,(x) \otimes s\,(x') \leq 0 \tag{3.14}$$

   Here $s\,(x)$ *is an algorithm that calculates* $\sin x$ *stably in the given interval, and* x' *denotes the next floating point number after* x *in the floating point system.*

Problem 3.11 (Exercise 15.2, [1]). *Consider an algorithm for the problem of computing the (full) SVD of a matrix. The data for this problem is a matrix* A, *and the solution is three matrices* U *(unitary),* $\sum$ *(diagonal), and* V *(unitary) such that* $A = U \sum V^*$. *(We are speaking here of explicit matrices* U *and* V, *not implicit representations as products of reflectors.)*

1. *Explain what it would mean for this algorithm to be backward stable.*
2. *In fact, for a simple reason, this algorithm cannot be backward stable. Explain.*
3. *Fortunately, the standard algorithms for computing the SVD (Lecture 31, [1]) are stable. Explain what stability means for such an algorithm.*

3.5 Lecture 16: Stability of Householder Triangularization

Problem 3.12 (Exercise 16.1, [1]).

1. *Let unitary matrices* $Q_1, \ldots, Q_k \in \mathbb{C}^{m,m}$ *be fixed and consider the problem of computing, for* $A \in \mathbb{C}^{m \times n}$, *the product* $B = Q_k \ldots Q_1 A$. *Let the computation be carried out from right to left by straightforward floating point operations on a computer satisfying* (3.11) *and* (3.12). *Show that this algorithm is backward stable. (Here* A *is thought of as data that can be perturbed; the matrices* Q_j *are fixed and not to be perturbed.)*

2. *Give an example to show that this result no longer holds if the unitary matrices Q_j are replaced by arbitrary matrices $X_j \in \mathbb{C}^{m\times m}$.*

Problem 3.13 (Exercise 16.2, [1]). *The idea of this exercise is to carry out an experiment analogous to the one described in this lecture, but for the SVD instead of QR factorization.*

1. *Write a* MATLAB *program that constructs a 50×50 matrix `A = U*S*V'`, where U and V are random orthogonal matrices and S is a diagonal matrix whose diagonal entries are random uniformly distributed numbers in $[0, 1]$, sorted into nonincreasing order. Have your program compute `[U2,S2,V2] = svd(A)` and the norms of `U-U2`, `V-V2`, `S-S2`, and `A-U2*S2*V2'`. Do this for five matrices A and comment on the results. (Hint: Plots of `diag(U2'*U)` and `diag(V2'*V)` may be informative.*

2. *Fix the signs in your computed SVD so that the difficulties of (1) go away. Run the program again for five random matrices and comment on the various norms. Do they have a connection with `cond(A)`?*

3. *Replace the diagonal entries of S by their sixth powers and repeat (2). Do you see significant differences between the results of this exercise and those of the experiments for QR factorization?*

3.6 Lecture 17: Stability of Back Substitution

Problem 3.14 (Exercise 17.1, [1]). *For any particular choice of norm $\|\cdot\|$, the bound*

$$\frac{|\delta r_{ij}|}{|r_{ij}|} \leq m\epsilon_{machine} + O\left(\epsilon_{machine}^2\right) \tag{3.15}$$

implies a more quantitative normwise bound than

$$\frac{\|\delta R\|}{\|R\|} = O\left(\epsilon_{machine}\right) \tag{3.16}$$

Derive such bounds for the norms $\|\cdot\|_1$, $\|\cdot\|_2$ and $\|\cdot\|_\infty$.

Problem 3.15 (Exercise 17.2, [1]). *Let $L \in \mathbb{C}^{m\times m}$ be a unit lower-triangular matrix (i.e., with diagonal entries equal to 1). For convenience, write L in the form*

$$sL = \begin{bmatrix} 1 & & & & \\ -l_{2,1} & 1 & & & \\ -l_{3,1} & -l_{3,2} & 1 & & \\ \vdots & \vdots & & \ddots & \\ -l_{m,1} & -l_{m,2} & -l_{m,3} & \cdots & 1 \end{bmatrix} \tag{3.17}$$

and define $M = L^{-1}$.

1. *Derive a formula for m_{ij} (which may involve other entries of M). Which entries of L does m_{ij} depend on?*

2. *Suppose the subdiagonal entries of L are independent random numbers ± 1 with equal probability. Fix k and define*

$$\mu_1 = m_{kk} \tag{3.18}$$

$$\mu_2 = m_{k+1,k} \tag{3.19}$$

$$\mu_3 = m_{k+2,k} \tag{3.20}$$

$$\cdots \tag{3.21}$$

   Write down a system of recurrences relations with random coefficients for the numbers μ_j.

3. *Experiments show that random triangular matrices with entries ± 1 are exponentially ill-conditioned in the sense that if κ_m denote the 2-norm condition number of a matrix of this kind of dimension m, then*

$$\lim_{m\to\infty} \sqrt[m]{\kappa_m} = C \tag{3.22}$$

   for some constant $1 < C < 1.5$. (The limit process can be made precise in various ways, but we shall not go into the technicalities; think of it as holding "with probability 1.") Perform numerical experiments involving random matrices of various dimensions to estimate C to 10% accuracy of better.

4. *Larger scale experiments become feasible if the random matrices of (3) are replaced by the random sequences $\mu_1, \mu_2, \mu_3, \ldots$ of (2). Explain (without proof) why the constant C can also be obtained by considering these sequences, and carry out numerical experiments to estimate C to 1% accuracy or better.*

3.7 Lecture 18: Conditioning of Least Squares Problems

Problem 3.16 (Exercise 18.1, [1]). *Consider the example*

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1.0001 \\ 1 & 1.0001 \end{bmatrix} \tag{3.23}$$

$$b = \begin{bmatrix} 2 \\ 0.0001 \\ 4.0001 \end{bmatrix} \tag{3.24}$$

1. *What are the matrices A^+ and P for this example? Give exact answers.*

2. *Find the exact solutions x and $y = Ax$ to the least squares problem $Ax \approx b$.*

3. *What are $\kappa(A), \theta$, and η? From here on, numerical answers are acceptable.*

4. *What are the four condition numbers of Theorem 18.1, [1]?*

5. *Give examples of perturbations δb and δA that approximately attain these four condition numbers.*

Problem 3.17 (Exercise 18.2, [1]). *Social scientists depend on the technique of* ***regression****, in which a vector of observations of some quantity is approximated in the least squares sense by a linear combination of other vectors. The coefficients of the fit are then interpreted as representing, say, the effects on annual income of IQ, years of education, parents' years of education, and parents' income.*

One might think that the more variables one included in such a model, the more information one would obtain, but this is not always true. Explain this phenomenon from the point of view of conditioning, making specific reference to the results of Theorem 18.1, [1].

Problem 3.18 (Exercise 18.3, [1]). *Suppose you look across Lake Cayuga at a light from a house on the other side. If the lake surface is rippled, the reflected light appears as a long vertical streak. The same effect appears with taillights o the car ahead of you on a rainy road, or even with reflections of hallway lights on a shiny waxed floor. It is a real effect, not an optical illusion, and the explanation is a matter of geometry.*

1. *Derive a quantitative theory explaining this phenomenon. Specifically, suppose you and the house across the lake are each fifty meters above the surface, and the lake is one kilometer wide. What is the length-to-width ratio of the streak as it appears in your visual fields?*

2. *Describe a connection between this problem and one of the geometrical arguments of Lecture 18, [1].*

Problem 3.19 (Exercise 18.4, [1]). *Explain why, as remarked after Theorem 18.1, [1], the condition number of y with respect to perturbations in A becomes 0 in the case $m = n$.*

3.8 Lecture 19: Stability of Least Squares Algorithms

Problem 3.20 (Exercise 19.1, [1]). *Given $A \in \mathbb{C}^{m\times m}$ of rank n and $b \in \mathbb{C}^m$, consider the block 2×2 system of equations*

$$\begin{bmatrix} I & A \\ A^* & 0 \end{bmatrix} \begin{bmatrix} r \\ x \end{bmatrix} = \begin{bmatrix} b \\ 0 \end{bmatrix} \tag{3.25}$$

where I is the $m \times m$ identity. Show that this system has a unique solution $(r, x)^T$, and that the vectors r and x are the residual and the solution of the least squares problem

Given $A \in \mathbb{C}^{m\times m}$ of full rank, $m \geq n, b \in \mathbb{C}^m$,
find $x \in \mathbb{C}^m$ such that $\|b - Ax\|$ is minimized.

Problem 3.21 (Exercise 19.2, [1]). *Heree is a stripped-down version of one of* MATLAB*'s built-in m-files.*

```
[U,S,V] = svd(A);
S = diag(S);
tol = max(size(A))*S(1)*eps;
r = sum(S > tol);
S = diag(ones(r,1)./S(1:r));
X = V(:,1:r)*S*U(:,1:r)';
```

What does this program compute?

Chapter 4

Systems of Equations

4.1 Lecture 20: Gaussian Elimination

Problem 4.1 (Exercise 20.1, [1]). *Let $A \in \mathbb{C}^{m \times m}$ be nonsingular. Show that A has an LU factorization if and only if for each k with $1 \leq k \leq m$, the upper-left $k \times k$ block $A_{1:k,1:k}$ is nonsingular. (Hint: The row operations of Gaussian elimination leave the determinants $\det(A_{1:k,1:k})$ unchanged.) Prove that this LU factorization is unique.*

Solution.

1. (⇒) If A has an LU factorization, ie.,
$$A = LU \tag{4.1}$$
then all the diagonal elements of U are nonzero since A is nonsingular.
The factorization (4.1) also implies that
$$A_{1:k,1:k} = L_{1:k,1:k} U_{1:k,1:k} \tag{4.2}$$
Combining (4.2) with two facts that $U_{1:k,1:k}$ is nonsingular (U is triangular and all its diagonal elements are nonzero) and $L_{1:k,1:k}$ is unit lower-triangular, yields that the upper-left $k \times k$ block $A_{1:k,1:k}$ is nonsingular for each k with $1 \leq k \leq m$.

2. (⇐) Suppose that for each k with $1 \leq k \leq m$, the upper-left $k \times k$ block $A_{1:k,1:k}$ is nonsingular.
We recall Gaussian elimination in [1].

Algorithm 3.1. Gaussian Elimination without Pivoting.

```
- $U = A, L = I$
- for $k = 1$ to $m - 1$
-         for $j = k + 1$ to $m$
-               $l_{jk} = \frac{u_{jk}}{u_{kk}}$
-               $u_{j,k:m} = u_{j,k:m} - l_{j,k} u_{k,k:m}$
```

Combining the assumption above with the fact that the row operations of Gaussian elimination leave the determinants $\det(A_{1:k,1:k})$ unchanged, we deduce that u_{kk} in the above Algorithm is always nonzero. Hence the step $l_{jk} = \dfrac{u_{jk}}{u_{kk}}$ makes sense and the algorithm proceeds straightforward. Therefore, A has an LU factorization.

Uniqueness. Suppose that A has two LU factorizations,

$$A = L_1U_1 = L_2U_2 \tag{4.3}$$

where L_1, L_2 are unit lower-triangular matrices and U_1, U_2 are nonsingular upper-triangular matrices, which is deduced from nonsingularity of A. Using the nonsingularity of L_2 and U_1 for (2.144) yields

$$L_2^{-1}L_1 = U_2U_1^{-1} \tag{4.4}$$

We have the left-hand side of (4.4) is a lower-triangular matrix whose all diagonal elements equal to 1. Whereas the right-hand side of (4.4) is a upper-triangular matrix. This possily happens if and only if

$$L_2^{-1}L_1 = U_2U_1^{-1} = I \tag{4.5}$$

or equivalently,

$$L_1 = L_2 \tag{4.6}$$
$$U_1 = U_2 \tag{4.7}$$

Therefore, the LU factorization of A is unique. □

Problem 4.2 (Exercise 20.2, [1]). *Suppose $A \in \mathbb{C}^{m\times m}$ satisfies the condition of Exercise 20.1, [1], and is banded with bandwidth $2p+1$, i.e., $a_{ij} = 0$ for $|i-j| > p$. What can you say about the sparsity patterns of the factors L and U of A?*

Problem 4.3 (Exercise 20.3, [1]). *Suppose an $m \times m$ matrix A is written in the block form*

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \tag{4.8}$$

where A_{11} is an $n \times n$ and A_{22} is $(m-n)\times(m-n)$. Assume that A satisfies the condition of Exercise 20.1, [1].

1. *Verify the formula*

$$\begin{bmatrix} I & \\ -A_{21}A_{11}^{-1} & I \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ & A_{22} - A_{21}A_{11}^{-1}A_{12} \end{bmatrix} \tag{4.9}$$

   *for "elimination" of the block A_{21}. The matrix $A_{22} - A_{21}A_{11}^{-1}A_{12}$ is known as the **Schur complement** of A_{11} in A.*

2. *Suppose A_{21} is eliminated row by row by means of n steps of Gaussian elimination. Show that the bottom-right $(m-n)\times(m-)$ block of the result is again $A_{22} - A_{21}A_{11}^{-1}A_{12}$.*

Problem 4.4 (Exercise 20.4, [1]). *Like most of the algorithms in this book, Gaussian elimination involves a triply nested loop. In Algorithm 20.1, there are two explicit for loops, and the third loop is implicit in the vectors $u_{j,k:m}$ and $u_{k,k:m}$. Rewrite this algorithm with just one explicit for loop indexed by k. Inside this loop, U will be updated at each step by a certain rank-one outer product. This "outer product" form of Gaussian elimination may be a better starting point than Algorithm 20.1, [1], if one wants to optimize computer performance.*

Problem 4.5 (Exercise 20.5, [1]). *We have seen that Gaussian elimination yields a factorization $A = LU$, where L has ones one the diagonal but U does not. Describe at a high level the factorization that results if this process is varied in the following ways:*

1. *Elimination by columns from left to right, rather than by rows from top to bottom, so that A is made lower-triangular.*
2. *Gaussian elimination applied after a preliminary scaling of the columns of A by a diagonal matrix D. What form does a system $Ax = b$ take under this rescaling? Is it the equations or the unknowns that are rescaled by D?*
3. *Gaussian elimination carried further, so that after A (assumed nonsingular) is brought to upper-triangular form, addition column operations are carried out so that this upper-triangular matrix is made diagonal.*

4.2 Lecture 21: Pivoting

Problem 4.6 (Exercise 21.1, [1]). *Let A be the* 4×4 *matrix*

$$A = \begin{bmatrix} 2 & 1 & 1 & 0 \\ 4 & 3 & 3 & 1 \\ 8 & 7 & 9 & 5 \\ 6 & 7 & 9 & 8 \end{bmatrix} \tag{4.10}$$

considered in Lecture 21, [1], in this lecture and the previous one.

1. *Determine* $\det A$ *from*

$$\begin{bmatrix} 2 & 1 & 1 & 0 \\ 4 & 3 & 3 & 1 \\ 8 & 7 & 9 & 5 \\ 6 & 7 & 9 & 8 \end{bmatrix} = \begin{bmatrix} 1 & & & \\ 2 & 1 & & \\ 4 & 3 & 1 & \\ 3 & 4 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 & 0 \\ & 1 & 1 & 1 \\ & & 2 & 2 \\ & & & 2 \end{bmatrix} \tag{4.11}$$

2. *Determine* $\det A$ *from*

$$\begin{bmatrix} & & 1 & \\ & & & 1 \\ & 1 & & \\ 1 & & & \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 & 0 \\ 4 & 3 & 3 & 1 \\ 8 & 7 & 9 & 5 \\ 6 & 7 & 9 & 8 \end{bmatrix} \tag{4.12}$$

$$= \begin{bmatrix} 1 & & & \\ \frac{3}{4} & 1 & & \\ \frac{1}{2} & -\frac{2}{7} & 1 & \\ \frac{1}{4} & -\frac{3}{7} & \frac{1}{3} & 1 \end{bmatrix} \begin{bmatrix} 8 & 7 & 9 & 5 \\ & \frac{7}{4} & \frac{9}{4} & \frac{17}{4} \\ & & -\frac{6}{7} & -\frac{2}{7} \\ & & & \frac{2}{3} \end{bmatrix} \tag{4.13}$$

3. *Describe how Gaussian elimination with partial pivoting can be used to find the determinant of a general square matrix.*

Problem 4.7 (Exercise 21.2, [1]). *Suppose $A \in \mathbb{C}^{m\times m}$ is banded with bandwidth $2p+1$, as in Exercise 20.2, [1], and a factorization $PA = LU$ is computed by Gaussian elimination with partial pivoting. What can you say about the sparsity patterns of L and U?*

Problem 4.8 (Exercise 21.3, [1]). *Consider Gaussian elimination carried out with pivoting by columns instead of rows, leading to a factorization $AQ = LU$, where Q is a permutation matrix.*

1. *Show that if A is nonsingular, such a factorization always exists.*

2. *Show that if A is singular, such a factorization does not always exist.*

Problem 4.9 (Exercise 21.4, [1]). *Gaussian elimination can be used to compute the inverse A^{-1} of a nonsingular matrix $A \in \mathbb{C}^{m\times m}$, though it is rarely really necessary to do so.*

1. *Describe an algorithm for computing A^{-1} by solving m systems of equations, and show that its asymptotic operation count is $\dfrac{8m^3}{3}$ flops.*

2. *Describe a variant of your algorithm, taking advantage of sparsity, that reduces the operation count to $2m^3$ flops.*

3. *Suppose one wishes to solve n systems of equations $Ax_j = b_j$, or equivalently, a block system $AX = B$ with $B \in \mathbb{C}^{m\times m}$. What is the asymptotic operation count (a function of m and n) for doing this*

   (a) *directly from the LU factorization and*

   (b) *with a preliminary computation of A^{-1}?*

Problem 4.10 (Exercise 21.5, [1]). *Suppose $A \in \mathbb{C}^{m\times m}$ is hermitian, or in the real case, symmetric (but not necessarily positive definite).*

1. *Describe a strategy of* ***symmetric pivoting*** *to preserve the hermitian structure while still leading to a unit lower-triangular matrix with entries $|l_{ij}| \le 1$.*

2. *What is the form of the matrix factorization computed by your algorithm?*

3. *What is its asymptotic operation count?*

Problem 4.11 (Exercise 21.6, [1]). *Suppose $A \in \mathbb{C}^{m\times m}$ is* ***strictly column diagonally dominant****, which means that for each k,*

$$|a_{kk}| > \sum_{j\neq k} |a_{jk}| \tag{4.14}$$

Show that if Gaussian elimination with partial pivoting is applied to A, no row interchanges take place.

Problem 4.12 (Exercise 21.7, [1]). *In Lecture 20, [1], the "two strokes of luck" were explained by the use of the vectors e_k and l_k. Give an explanation based on these vectors for the "third stroke of luck" in the Lecture 21, [1].*

4.3 Lecture 22: Stability of Gaussian Elimination

Problem 4.13 (Exercise 22.1, [1]). *Show that for Gaussian elimination with partial pivoting applied to any matrix $A \in \mathbb{C}^{m\times m}$, the growth factor*

$$\rho = \frac{\max_{i,j} |u_{ij}|}{\max_{i,j} |a_{ij}|} \tag{4.15}$$

satisfies $\rho \leq 2^{m-1}$.

Problem 4.14 (Exercise 22.2, [1]). *Experiment with solving 60×60 systems of equations $Ax = b$ by Gaussian elimination with partial pivoting, with A having the form*

$$A = \begin{bmatrix} 1 & & & & 1 \\ -1 & 1 & & & 1 \\ -1 & -1 & 1 & & 1 \\ -1 & -1 & -1 & 1 & 1 \\ -1 & -1 & -1 & -1 & 1 \end{bmatrix} \tag{4.16}$$

Do you observe that the results are useless because of the growth factor of order 2^{60}? At your first attempt you may not observe this, because the integer entries of A may prevent any rounding errors from occurring. If so, find a way to modify your problem slightly so that the growth factor is the same or nearly so and catastrophic rounding errors really do take place.

Problem 4.15 (Exercise 22.3, [1]). *Reproduce the figures of Lecture 22, [1], approximately if not in full detail, but based on random matrices with entries uniformly distributed in $[-1, 1]$ rather than normally distributed. Do you see any significant differences?*

Problem 4.16 (Exercise 22.4, [1]).

1. *Suppose $PA = LU$ (LU factorization with partial pivoting) and $A = QR$ (QR factorization). Describe a relationship between the last row of L^{-1} and the last column of Q.*

2. *Show that if A is random in the sense of having independent, normally distributed entries, then its column spaces are randomly oriented, so that in particular, the last column of Q is a random unit vector.*

3. *Combine the results of (1) and (2) to make a statement about the final row of L^{-1} in Gaussin elimination applied to a random matrix A.*

4.4 Lecture 23: Cholesky Factorization

Problem 4.17 (Exercise 23.1, [1]). *Let A be a nonsingular square matrix and let $A = QR$ and $A^*A = U^*U$ be QR and Cholesky factorizations, respectively, with the usual normalizations $r_{jj}, u_{jj} > 0$. Is it true or false that $R = U$?*

Problem 4.18 (Exercise 23.2, [1]). *Using the proof of Theorem 16.2, [1], as a guide, derive Theorem 23.3, [1], from Theorem 23.2 and Theorem 17.1, [1].*

Problem 4.19 (Exercise 23.3, [1]). ***Reverse Software Engineering of "\\".*** *The following* MATLAB *session records a sequence of tests of the elapsed times for various computations on a workstation manufactured in 1991. For each part, try to explain*

1. *Why was this experiment carried out?*
2. *Why did the result came out as it did? Your answer should refer to formulas from the text for flop counts. The* MATLAB *queries* ***`help chol`*** *and* ***`help slash`*** *may help in your detective work.*

(a)
```
m = 200; Z = randn(m,m);
A = Z'*Z; b = randn(m,l);
tic; x = A\b; toe;
    elapsed_time = 1.0368
```

(b)
```
tic; x = A\b; toe;
    elapsed_time = 1.0303
```

(c)
```
A2 = A; A2(m,l) = A2(m,l)/2;
tic; x = A2\b; toe;
    elapsed_time = 2.0361
```

(d)
```
I = eye(m,m); emin = min(eig(A));
A3 = A - .9*emin*I;
tic; x = A3\b; toe;
    elapsed_time = 1.0362
```

(e)
```
A4 = A - l.l*emin*I;
tic; x = A4\b; toe;
    elapsed_time = 2.9624
```

(f)
```
A5 = triu(A);
tic; x = A5\b; toe;
    elapsed_time = 0.1261
```

(g)
```
A6 = A5; A6(m,l) = A5(l,m);
tic; x = A6\b; toe;
    elapsed-time = 2.0012
```

Chapter 5

Eigenvalues

5.1 Lecture 24: Eigenvalue Problems

Problem 5.1 (Exercise 24.1, [1]). *For each of the following statements, prove that it is true or give an example to show it is false. Throughout, $A \in \mathbb{C}^{m\times m}$ unless otherwise indicated, and "ew" stands for eigenvalue. (This comes from the German "Eigenwert". The corresponding abbreviation for eigenvector is "ev", from "Eigenvektor".)*

1. *If λ is an ew of A and $\mu \in \mathbb{C}$, then $\lambda - \mu$ is an ew of $A - \mu I$.*
2. *If A is real and λ is an ew of A, then so is $-\lambda$.*
3. *If A is real and λ is an ew of A, then so is $\bar{\lambda}$.*
4. *If λ is an ew of A and A is nonsingular, then λ^{-1} is an ew of A^{-1}.*
5. *If all the ew's of A are zero, then $A = 0$.*
6. *If A is hermitian and λ is an ew of A, then $|\lambda|$ is a singular value of A.*
7. *If A is diagonalizable and all its ew's are equal, then A is diagonal.*

Solution.

1. *Correct.* Since λ is an eigenvalue of A, there exists an nonzero vector v such that

$$Av = \lambda v \tag{5.1}$$

Hence,

$$\begin{aligned}(A - \mu I)\, v &= Av - \mu v &(5.2)\\ &= (\lambda - \mu)\, v &(5.3)\end{aligned}$$

Therefore, $\lambda - \mu$ is an eigenvalue of $A - \mu I$.

2. *Incorrect.* The easiest counter-examples are the unit matrices I_n.

3. *Correct.* Consider the characteristic polynomial of A

$$p_A(x) = \sum_{i=0}^{n} a_i x^i \tag{5.4}$$

Since λ is an eigenvalue of A, we have

$$p_A(\lambda) = 0 \tag{5.5}$$

Taking the complex conjugate of both sides of (5.5) yields

$$\begin{aligned} 0 &= \overline{p_A(\lambda)} \qquad (5.6) \\ &= \overline{\sum_{i=0}^{n} a_i \lambda^i} \qquad (5.7) \\ &= \sum_{i=0}^{n} a_i \bar{\lambda}^i \qquad (5.8) \\ &= p_A(\bar{\lambda}) \qquad (5.9) \end{aligned}$$

where we have used the assumption that A is real. Therefore, $\bar{\lambda}$ is also an eigenvalue of A.

4. *Correct.* Since λ is an eigenvalue of A, there exists a nonzero vector v such that

$$Av = \lambda v \tag{5.10}$$

Since A is nonsingular, (5.10) is equivalent to

$$A^{-1} v = \lambda^{-1} v \tag{5.11}$$

Therefore, λ^{-1} is an eigenvalue of A^{-1}.

5. *Incorrect.* The easiest counter-examples are nonzero strictly upper-triangular matrices (or nonzero strictly lower-triangular matrices).

6. *Correct.* Theorem 5.5, [1].

7. *Correct.* Since A is diagonalizable, A can be represented as

$$A = P \sum P^{-1} \tag{5.12}$$

where $\sum$ is the diagonal matrix whose elements are eigenvalues of A. By the assumption that all eigenvalues of A are equal, we deduce that

$$\sum = \lambda I \tag{5.13}$$

where λ's are eigenvalues of A.
Combining (5.12) and (5.13) yields

$$\begin{aligned} A &= cPIP^{-1} \qquad (5.14) \\ &= cI \qquad (5.15) \end{aligned}$$

i.e., A is diagonal.

Done. □

Problem 5.2 (Exercise 24.2, [1]). *Here is Gerschgorin's theorem, which holds for any $m \times m$ matrix A, symmetric or nonsymmetric. "Every eigenvalues of A lies in at least one of the m circular disks in the complex plane with centers a_{ii} and radii $\sum_{j \neq i} |a_{ij}|$. Moreover, if n of these disks form a connected domain that is disjoint from the other $m - n$ disks, then there are precisely n eigenvalues of A within this domain.*

1. *Prove the first part of Gerschgorin's theorem. (Hint: Let λ be any eigenvalue of A, and x a corresponding eigenvector with largest entry 1.*
2. *Prove the second part. (Hint: Deform A to a diagonal matrix and use the fact that the eigenvalues of a matrix are continuous functions of its entries.*
3. *Give estimates based on Gerschgorin's theorem for the eigenvalues of*

$$A = \begin{pmatrix} 8 & 1 & 0 \\ 1 & 4 & \varepsilon \\ 0 & \varepsilon & 1 \end{pmatrix}, \quad |\varepsilon| < 1 \tag{5.16}$$

4. *Find a way to establish the tighter bound $|\lambda_3 - 1| \leq \varepsilon^2$ on the smallest eigenvalue of A. (Hint: Consider diagonal similarity transformations.)*

Solution. First of all, we state the Gerschgorin's theorem and its generalization and we then prove them in order. The following subsection includes solution of Problem 2.1 and 2.2. Problem 2.3 and 2.4 will be proved later.

5.1.1 Gershgorin Circle Theorem

See [2].

In mathematics, the *Gershgorin circle theorem* may be used to bound the spectrum of a square matrix. It was first published by the Soviet mathematican Semyon Aronovich Gershgorin in 1931. Let A be a complex $n \times n$ matrix, with entries a_{ij}. For $i = 1, 2, \ldots, n$ let

$$R_i = \sum_{j \neq i} |a_{ij}| \tag{5.17}$$

be the sum of the absolute values of the non-diagonal entries in ith row. Let $D(a_{ii}, R_i)$ be the closed disc centered at a_{ii} with radius R_i. Such a disc is called a *Gershgorin disc*.

Theorem 5.2.1. *Every eigenvalue of A lies within at least one of the Gershgorin discs $D(a_{ii}, R_i)$.*

Proof. Let λ be an eigenvalue of A and let $x = (x_j)$ be a corresponding eigenvector. Let $i \in \{1, 2, \ldots, n\}$ be chosen so that

$$|x_i| = \max_j |x_j| \tag{5.18}$$

That is to say, choose i so that x_i is the largest number in absolute value in the the vector x. Then $|x_i| > 0$, otherwise $x = 0$. Since x is an eigenvector, $Ax = \lambda x$, and thus

$$\sum_j a_{ij} x_j = \lambda x_i, \quad i = 1, 2, \ldots, n \tag{5.19}$$

So, splitting the sum, we get

$$\sum_{j \neq i} a_{ij} x_j = \lambda x_i - a_{ii} x_i \tag{5.20}$$

We may then divide both sides by $x_i \neq 0$ and take the absolute value to obtain

$$\begin{aligned} |\lambda - a_{ii}| &= \left| \frac{\sum\limits_{j \neq i} a_{ij} x_j}{x_i} \right| && (5.21) \\ &\leq \sum_{j \neq i} \left| \frac{a_{ij} x_j}{x_i} \right| && (5.22) \\ &\leq \sum_{j \neq i} |a_{ij}| && (5.23) \\ &= R_i && (5.24) \end{aligned}$$

where the last inequality is valid because

$$\left| \frac{x_j}{x_i} \right| \leq 1, \forall j \neq i \tag{5.25}$$

Done. □

Corollary 5.2.2. *The eigenvalues of A must also lie within the Gershgorin discs C_j corresponding to the columns of A.*

Proof. Apply Theorem 5.2.1 to A^T. □

For a diagonal matrix, the Gershgorin discs coincide with the spectrum. Conversely, if the Gershgorin discs coincide with the spectrum, the matrix is diagonal.

Theorem 5.2.3 (Strengthening of Gershgorin Circle Theorem). *If the union of k discs is disjoint from the union of the other $n - k$ discs then the former union contains exactly k and the latter $n - k$ eigenvalues of A.*

Proof. Let D be the diagonal matrix with entries equal to the diagonal entries of A and let

$$B(t) = (1 - t) D + tA \tag{5.26}$$

We will use the fact that the eigenvalues are continuous in t, and show that if any eigenvalue moves from one of the unions to the other, then it must be outside all the discs for some t, which is a contradiction.

The statement is true for $D + B(0)$. The diagonal entries of $B(t)$ are equal to that of A, thus the centers of the Gershgorin circles are the same, however their radii are t times that of A. Therefore the union of the corresponding k discs of $B(t)$ is disjoint from the union of the remaining $n-k$ for all $t \in [0,1]$. The discs are closed, so the distance of the two unions for A is $d > 0$. The distance for $B(t)$ is a decreasing function of $t \in [0,1]$, so it is always at least d. Since the eigenvalues of $B(t)$ are continuous functions of t, for any eigenvalue $\lambda(t)$ of $B(t)$ in the union of the k discs its distance $d(t)$ from the union of the other $n-k$ discs is also continuous. Obviously, $d(0) \geq d$, and assume $\lambda(1)$ lies in the union of the $n-k$ discs. Then $d(1) = 0$, so there exists $0 < t_0 < 1$ such that $0 < d(t_0) < d$.

But this means $\lambda(t_0)$ lies outside the Gershgorin discs, which is impossible. Therefore $\lambda(1)$ lies in the union of the k discs, and the theorem is proven. □

5.1.2 An Applications of Gerschgorin's Theorem

Applying Gerschgorin's theorem to the matrix A defined by (5.16), where Gerschgorin's discs are defined by

$$D_1 = D(8,1) \tag{5.27}$$
$$D_2 = D(4, 1 + |\varepsilon|) \tag{5.28}$$
$$D_3 = D(1, |\varepsilon|) \tag{5.29}$$

yields that every eigenvalues of A must satisfy at least one of the following three inequalities

$$|\lambda - 8| \leq 1 \tag{5.30}$$
$$|\lambda - 4| \leq 1 + |\varepsilon| \tag{5.31}$$
$$|\lambda - 1| \leq |\varepsilon| \tag{5.32}$$

We first notice that D_1, D_2, D_3 are pairwise separated due to the assumption $|\varepsilon| < 1$. Hence, we can apply Theorem 5 for this matrix A and deduce that each $D_i,\ i = 1,2,3$ contains exactly one eigenvalue of A, i.e., with suitable reindexing, we have

$$\lambda_i \in D_i, \ \ i = 1,2,3 \tag{5.33}$$

In particular, we only obtain

$$|\lambda_3 - 1| \leq |\varepsilon| \tag{5.34}$$

This inequality is not sharp. We can investigate to obtain the inequality

$$|\lambda_3 - 1| \leq \varepsilon^2 \tag{5.35}$$

****

Problem 5.3 (Exercise 24.3, [1]). *Let A be a 10×10 random matrix with entries from the standard normal distribution, minus twice the identity. Write a program to plot $\left\|e^{tA}\right\|_2$ against t for $0 \leq t \leq 20$ on a log scale, comparing the result to the straight line $e^{t\alpha(A)}$, where $\alpha(A) = \max_j Re(\lambda_j)$ is the **spectral***

abscissa *of A. Run the program for ten random matrices A and comment on the results. What property of a matrix leads to a $\left\|e^{tA}\right\|_2$ curve that remains oscillatory as $t \to \infty$?*

Problem 5.4 (Exercise 24.4, [1]). *For an arbitrary $A \in \mathbb{C}^{m\times m}$ and norm $\|\cdot\|$, prove using Theorem 24.9, [1].*

1.
$$\lim_{n\to\infty} \|A^n\| = 0 \Leftrightarrow \rho(A) < 1 \tag{5.36}$$
where ρ is the spectral radius (Exercise 3.2, [1]).

2.
$$\lim_{t\to\infty} \left\|e^{tA}\right\| = 0 \Leftrightarrow \alpha(A) < 0 \tag{5.37}$$
where α is the spectral abscissa.

5.2 Lecture 25: Overview of Eigenvalue Algorithms

Problem 5.5 (Exercise 25.1, [1]).

1. *Let $A \in \mathbb{C}^{m\times m}$ be tridiagonal and hermitian, with all its sub- and super-diagonal entries nonzero. Prove that the eigenvalues of A are distinct. (Hint: Show that for any $\lambda \in \mathbb{C}, A - \lambda I$ has rank at least $m-1$.)*

2. *On the other hand, let A be upper-Hessenberg, with all its subdiagonal entries nonzero. Give an example that shows that the eigenvalues of A are not necessarily distinct.*

Problem 5.6 (Exercise 25.2, [1]). *Let $e_1, e_2, e_3, \ldots$ be a sequence of non-negative numbers representing errors inn some iterative process that converge to zero, and suppose there are a constant C and an exponent α such that for all sufficiently large k, $e_{k+1} \leq C(e_k)^\alpha$. Various algorithms for "Phase 2" of an eigenvalue calculation exhibit* ***cubic convergence*** *($\alpha = 3$),* ***quadratic convergence*** *($\alpha = 2$), or* ***linear convergence*** *($\alpha = 1$ with $C < 1$), which is also, perhaps confusingly, known as* ***geometric convergence.***

1. *Suppose we want an answer of accuracy $O(\epsilon_{machine})$. Assuming the amount of work for each step is $O(1)$, show that the total work requirement in the case of linear convergence is $O(\log(\epsilon_{machine}))$. How does the constant C enter into your work estimate?*

2. *Show that in the case of superlinear convergence, i.e., $\alpha > 1$, the work requirement becomes $O(\log(|\log(\epsilon_{machine})|))$. (Hint: The problem may be simplified by defining a new error measure $f_k = C^{\frac{1}{\alpha-1}} e_k$.) How does the exponent α enter into your work estimate?*

Problem 5.7 (Exercise 25.3, [1]). *Suppose we have a* 3×3 *matrix and wish to introduce zeros by left and/or right-multiplications by unitary matrices* Q_j *such as Householder reflectors or Givens rotations. Consider the following three matrix structures.*

$$\begin{bmatrix} \times & \times & 0 \\ 0 & \times & \times \\ 0 & 0 & \times \end{bmatrix}, \begin{bmatrix} \times & \times & 0 \\ \times & 0 & \times \\ 0 & \times & \times \end{bmatrix}, \begin{bmatrix} \times & \times & 0 \\ 0 & 0 & \times \\ 0 & 0 & \times \end{bmatrix} \tag{5.38}$$

For each one, decide which of the following situations holds, and justify your claim.

1. *Can be obtained by a sequence of left-multiplications by matrices* Q_j.
2. *Not (1), but can be obtained by a sequence of left- and right-multiplications by matrices* Q_j.
3. *Cannot be obtained by any sequence of left- and right-multiplications by matrices* Q_j.

5.3 Lecture 26: Reduction to Hessenberg or Tridiagonal Form

Problem 5.8 (Exercise 26.1, [1]). *Theorem 26.1, [1], and its successors in Lecture 27, [1], show that we can compute eigenvalues* $\left\{\tilde{\lambda}_k\right\}$ *of* A *numerically that are the exact eigenvalues of a matrix* $A + \delta A$ *with* $\frac{\|\delta A\|}{\|A\|} = O\left(\epsilon_{machine}\right)$. *Does this mean they are close to the exact eigenvalues* $\{\lambda_k\}$ *of* A? *This is a question of eigenvalue perturbation theory.*

One can approach such problems geometrically as follows. Given $A \in \mathbb{C}^{m\times m}$ *with spectrum* $\Lambda(A) \subseteq \mathbb{C}$ *and* $\epsilon > 0$, *define the 2-norm* ***ε-pseudospectrum of*** A, $\Lambda_\epsilon(A)$, *to be the set of numbers* $z \in \mathbb{C}$ *satisfying any of the following conditions*

1. z *is an eigenvalue of* $A + \delta A$ *for some* δA *with* $\|\delta A\|_2 \leq \varepsilon$.
2. *There exists a vector* $u \in \mathbb{C}^m$ *with* $\|(A - zI)u\|_2 \leq \epsilon$ *and* $\|u\|_2 = 1$.
3. $\sigma_m(zI - A) \leq \varepsilon$.
4. $\left\|(zI - A)^{-1}\right\|_2 \geq \frac{1}{\epsilon}$.

The matrix $(zI - A)^{-1}$ *in (4) is known as the* ***resolvent*** *of* A *at* z; *if* z *is an eigenvalue of* A, *we use the convention* $\left\|(zI - A)^{-1}\right\|_2 = \infty$. *In (3),* σ_m *denotes the smallest singular value.*

Prove that conditions (1)-(4) are equivalent.

Problem 5.9 (Exercise 26.2, [1]). *Let* A *be the* 32×32 *matrix with* -1 *on the main diagonal, 1 on the first and second superdiagonals, and 0 elsewhere.*

1. *Using an SVD algorithm built in to* MATLAB *or another software system, together with contour plotting software, generate a plot of the boundaries of the 2-norm* ϵ*-pseudospectra of* A *for* $\epsilon = 10^{-1}, 10^{-2}, \ldots, 10^{-8}$.

2. *Produce a semilogy plot of $\left\|e^{tA}\right\|_2$ against t for $0 \leq t \leq 50$. What is the initial growth rate of the curve before the eventual decay sets in? Can you relate this to your plot of pseudospectra? (Compare Exercise 24.3, [1]).*

Problem 5.10 (Exercise 26.3, [1]). *One of the best known results of eigenvalue perturbation theory is the* ***Bauer-Fike theorem,***

Theorem 5.10.1 (Bauer-Fike). *Suppose $A \in \mathbb{C}^{m\times m}$ is diagonalizable with $A = V\Lambda V^{-1}$, and let $\delta A \in \mathbb{C}^{m\times m}$ be arbitrary. Then every eigenvalue of $A+\delta A$ lies in at least one of the m circular disks in the complex plane of radius $\kappa\left(V\right)\left\|\delta A\right\|_2$ centered at the eigenvalues of A, where κ is the 2-norm condition number. (Compare Exercise 24.2, [1]).)*

1. *Prove the Bauer-Fike theorem by using the equivalence of conditions (1) and (4) of Exercise 26.1, [1].*

2. *Suppose A is normal. Show that for each eigenvalue $\tilde{\lambda}_j$ of $A+\delta A$, there is an eigenvalue λ_j of A such that*

$$\left|\tilde{\lambda}_j - \lambda_j\right| \leq \left\|\delta A\right\|_2 \tag{5.39}$$

5.4 Lecture 27: Rayleigh Quotient, Inverse Iteration

Problem 5.11 (Exercise 27.1, [1]). *Let $A \in \mathbb{C}^{m\times m}$ be given, not necessarily hermitian. Show that a number $z \in \mathbb{C}$ is a Rayleigh quotient of A if and only if it is a diagonal entry of Q^*AQ for some unitary matrix Q. Thus Rayleigh quotient are just diagonal entries of matrices, once you transform orthogonally to the right coordinate system.*

Problem 5.12 (Exercise 27.2, [1]). *Again let $A \in \mathbb{C}^{m\times m}$ be arbitrary. The set of all Rayleigh quotients of A, corresponding to all nonzero vectors $x \in \mathbb{C}^m$, is known as the* ***field of values*** *or* ***numerical range*** *of A, a subset of the complex plane denoted by $W\left(A\right)$.*

1. *Show that $W\left(A\right)$ contains the convex hull of the eigenvalues of A.*

2. *Show that if A is normal,then $W\left(A\right)$ is equal to the convex hull of the eigenvalues of A.*

Problem 5.13 (Exercise 27.3, [1]). *Show that for a nonhermitian matrix $A \in \mathbb{C}^{m\times m}$, the Rayleigh quotient $r\left(x\right)$ gives an eigenvalue estimate whose accuracy is generally linear, not quadratic. Explain what convergence rate this suggests for the Rayleigh quotient iteration applied to nonhermitian matrices.*

Problem 5.14 (Exercise 27.4, [1]). *Every real symmetric square matrix can be orthogonally diagonalized, and the developments of Lecture 27, [1], are invariant under orthogonal changes of coordinates. Thus it would have been sufficient to carry out each derivation of this lecture under the assumption that A*

is a diagonal matrix with entries ordered by decreasing absolute value. Making this assumption, describe the form taken by

$$\begin{aligned} v^{(k)} &= c_k A^k v^{(0)} && (5.40)\\ &= c_k \left(a_1\lambda_1^k q_1 + a_2\lambda_2^k q_2 + \cdots + a_m\lambda_m^k q_m\right) && (5.41)\\ &= c_k\lambda_1^k \left(a_1 q_1 + a_2\left(\frac{\lambda_2}{\lambda_1}\right)^k q_2 + \cdots + a_m\left(\frac{\lambda_m}{\lambda_1}\right)^k q_m\right) && (5.42)\end{aligned}$$

and

$$\left\| v^{(k)} - (\pm q_1)\right\| = O\left(\left|\frac{\lambda_2}{\lambda_1}\right|^k\right) \tag{5.43}$$

$$\left|\lambda^{(k)} - \lambda_1\right| = O\left(\left|\frac{\lambda_2}{\lambda_1}\right|^{2k}\right) \tag{5.44}$$

and Algorithm 27.3, [1].

Problem 5.15 (Exercise 27.5, [1]). *As mentioned in the text, inverse iteration depends on the solution of a system of equations that may be exceeding ill-conditioned, with condition number on the order of $\epsilon_{machine}^{-1}$. We know that it is impossible in general to solve ill-conditioned systems accurately. Is this not a fatal flaw in the algorithm?*

Show as follows that the answer is no - that ill-conditioning is not a problem in inverse iteration. Suppose A is a real symmetric matrix with one eigenvalue much smaller than the others in absolute value (without loss of generality, we are taking $\mu = 0$). Suppose v is a vector with components in the directions of all the eigenvalues $q_1, \ldots, q_m$ of A, and suppose $Aw = v$ is solved backward stably, yielding a computed vector $\tilde{w}$. Making use of the calculation on [1] - p.95, show that although $\tilde{w}$ may be far from w, $\frac{\tilde{w}}{\|\tilde{w}\|}$ will not be far from $\frac{w}{\|w\|}$.

Problem 5.16 (Exercise 27.6, [1]). *What happens to Figure 27.1 if two of the eigenvalues of A are equal?*

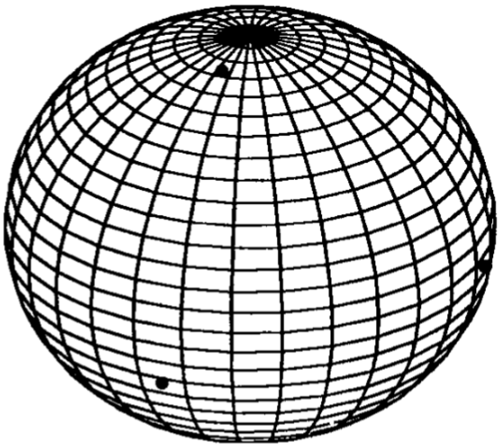


Figure 5.1: *The Rayleigh quotient $r(x)$ is a continuous function on the unit sphere $\|x\| = 1$ in $\mathbb{R}^n$, and the stationary points of $r(x)$ are the normalized eigenvectors of A. In this example with $m = 3$, there are three orthogonal stationary points (as well as their antipodes).*

5.5 Lecture 28: QR Algorithm without Shifts

Problem 5.17 (Exercise 28.1, [1]). *What happens if you apply the unshifted QR algorithm to an orthogonal matrix? Figure out the answer, and then explain how it relates to Theorem 28.4, [1].*

Problem 5.18 (Exercise 28.2, [1]). *The preliminary reduction to tridiagonal form would be of little use if the steps of the QR algorithm did not preserve this structure. Fortunately, they do.*

1. *In the QR factorization $A = QR$ of a tridiagonal matrix A, which entries of R are in general nonzero? Which entries of Q? (In practice we do not form Q explicitly.)*
2. *Show that the tridiagonal structure is recovered when the product RQ is formed.*
3. *Explain how Givens rotations or 2×2 Householder reflections can be used in the computations of the QR factorization of a tridiagonal matrix, reducing the operation count far below what would be required for a full matrix.*

Problem 5.19 (Exercise 28.3, [1]). *A real symmetric matrix A has an eigenvalue 1 of multiplicity 8, while all the rest of the eigenvalues are ≤ 0.1 in absolute value. Describe an algorithm for finding an orthonormal basis of the 8-dimensional eigenspace corresponding to the dominant eigenvalue.*

Problem 5.20 (Exercise 28.4, [1]). *Consider one step of Algorithm 28.1, applied to a tridiagonal symmetric matrix $A \in \mathbb{R}^{m \times m}$.*

1. *If only eigenvalues are desired, then only $A^{(k)}$ is needed at step k, not $Q^{(k)}$. Determine how many flops are required to get from $A^{(k-1)}$ to $A^{(k)}$ using standard methods described in [1].*

2. *If all the eigenvectors are desired, then the matrix*

$$\bar{Q}^{(k)} = Q^{(1)}Q^{(2)} \cdots Q^{(k)} \tag{5.45}$$

will need to be accumulated too. Determine how many flops are now required to get from step $k-1$ to step k.

5.6 Lecture 29: QR Algorithm with Shifts

Problem 5.21 (Exercise 29.1, [1]). *This five-part problem asks you to put together a* MATLAB *program that finds all the eigenvalues of a real symmetric matrix, using only elementary building blocks. It is not necessary to achieve optimal constant factors by exploiting symmetry or zero structure optimally. It is possible to solve the whole problem by a program about fifty lines long.*

1. *Write a function `T = tridiag(A)` that reduces a real symmetric $m \times m$ matrix to tridiagonal form by orthogonal similarity transformations. Your program should use only elementary* MATLAB *operations - not the built-in function `hess`, for example. Your output matrix T should be symmetric and tridiagonal up to rounding errors. If you like, add a line that forces T at the end to be exactly symmetric and tridiagonal. For an example, apply your program to `A = hilb(4)`.*

2. *Write a function `Tnew = qralg(T)` that runs the unshifted QR algorithm on a real tridiagonal matrix T. For the QR factorization at each step, use programs `[W,R] = house(A)` and `Q = formQ(W)` of Exercise 10.2, [1], if available, or* MATLAB*'s command `qr`, or, for greater efficiency, a new code based on Givens rotations or 2×2 Householder reflections rather than $m \times m$ operations. Again, you may wish to enforce symmetry and tridiagonally at each step. Your program should stop and return the current tridiagonal matrix T as `Tnew` when the $m, m-1$ element satisfies $|t_{m,m-1}| < 10^{-12}$ (hardy an industrial strength convergence criterion!). Again, apply your program to `A = hilb(4)`.*

3. *Write a driver program which (1) calls `qralg` to get one eigenvalue, (3) calles `qralq` with a smaller matrix to get another eigenvalue, and so on until all of the eigenvalues of A are determined. Set things up so that the values of $|t_{m,m-1}|$ at every QR iteration are stored in a vector and so that at the end, your program generates a semilogy plot of these values as a function of the number of Qr factorizations. (Here m will step from `length(A)` to `length(A)-1` and so on down to 3 and finally 2 as the deflation proceeds, and the plot will be correspondingly sawtoothed.) Run your program for `A = hilb(4)`. The output should be a set of eigenvalues and a "sawtooth plot".*

4. *Modify `qralg` so that it uses the Wilkinson shift at each step. Turn in the new sawtooth plot for the same example.*

5. *Return your program for the matrix* `A = diag(15:-1:1) + ones(15,15)` *and generate two sawtooth plots corresponding to shift and no shift. Discuss the rates of convergence observed here and for the earlier matrix. Is the convergence linear, superlinear, quadratic, cubic, ...? Is it meaningful to speak of a certain "number of QR iterations per eigenvalue?"*

5.7 Lecture 30: Other Eigenvalue Algorithms

Problem 5.22 (Exercise 30.1, [1]). *Derive the formula*

$$\tan(2\theta) = \frac{2d}{b-a} \tag{5.46}$$

and give a precise geometric interpretation of the transformation

$$J^T \begin{bmatrix} a & d \\ d & b \end{bmatrix} J = \begin{bmatrix} \neq 0 & 0 \\ 0 & \neq 0 \end{bmatrix} \tag{5.47}$$

based on this choice of θ.

Problem 5.23 (Exercise 30.2, [1]). *How many flops are required for one step*

$$J^T \begin{bmatrix} a & d \\ d & b \end{bmatrix} J = \begin{bmatrix} \neq 0 & 0 \\ 0 & \neq 0 \end{bmatrix} \tag{5.48}$$

of the Jacobi algorithm? How many flops for $\frac{m(m-1)}{2}$ such steps, i.e., one sweep? How does the operation count for one sweep compare with the total operation count for tridiagonalizing a real symmetric matrix and finding its eigenvalues by the QR algorithm?

Problem 5.24 (Exercise 30.3, [1]). *Show that if the largest off-diagonal entry is annihilated at each step of the Jacobi algorithm, then the sum of the squares of the off-diagonal entries decreases by at least the factor $1 - \frac{2}{m^2-m}$ at each step.*

Problem 5.25 (Exercise 30.4, [1]). *Suppose m is even and your computer has $\dfrac{m}{2}$ processors. Explain how $\dfrac{m}{2}$ transformations*

$$J^T \begin{bmatrix} a & d \\ d & b \end{bmatrix} J = \begin{bmatrix} \neq 0 & 0 \\ 0 & \neq 0 \end{bmatrix} \tag{5.49}$$

can be carried out in parallel if they involve the disjoint row/column pairs $(1,2), (3,4), (5,6), \ldots, (m-1,m)$.

Problem 5.26 (Exercise 30.5, [1]). *Write a program to find the eigenvalues of an $m \times m$ real symmetric matrix by the Jacobi algorithm with the standard row-wise ordering, plotting the sum of the squares of the off-diagonal entries on a log scale as a function of the number of sweeps. Apply your program to random matrices of dimensions 20, 40 and 80.*

Problem 5.27 (Exercise 30.6, [1]). *How many eigenvalues does*

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & 3 \end{bmatrix} \tag{5.50}$$

have in the interval $[1, 2]$*? Work out the answer on paper by bisection, making use of the recurrence*

$$p^{(k)}(x) = (a_k - x)\, p^{(k-1)}(x) - b_{k-1}^2 p^{(k-2)}(x) \tag{5.51}$$

Problem 5.28 (Exercise 30.7, [1]). *Construct a random real symmetric tridiagonal matrix* T *of dimension 100 and compute its eigenvalue decomposition,* $T = QDQ^T$*. Plot a few of the eigenvectors on a log scale (the absolute values of a few columns of* Q*) and observe the phenomenon of localization. What proportion of the 10000 entries of* Q *are greater than* 10^{-10} *in magnitude? What is the answer if instead of a random matrix,* T *is the discrete Laplacian with entries* $1, -2, 1$*?*

5.8 Lecture 31: Computing the SVD

Problem 5.29 (Exercise 31.1, [1]).

1. *Show that, as claimed in the text and illustrated in Figure,*

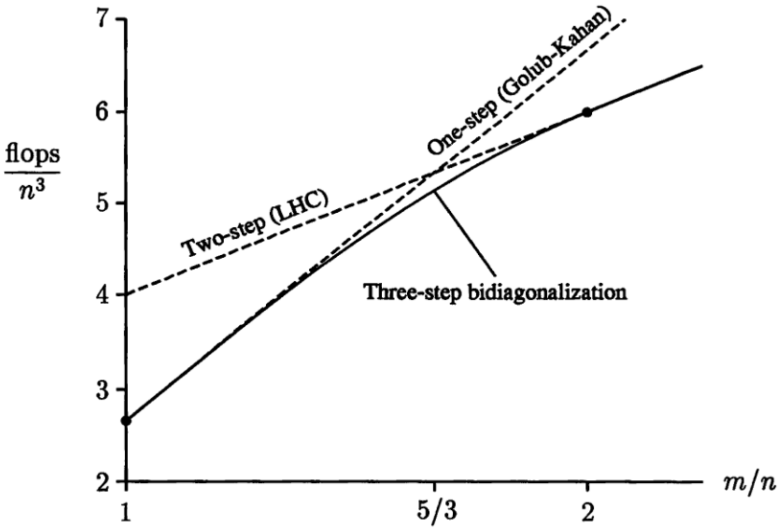


Figure 5.2: *Operation counts for three bidiagonalization algorithms applied to* $m \times n$ *matrices,. Three-step bidiagonalization provides a pleasingly smooth interpolant between the other two methods, though the improvement is hardly large.*

the crossover aspect ratio at which LHC bidiagonalization begins to beat Golub-Kahan bidiagonalization is $\frac{m}{n} = \frac{5}{3}$*?*

Problem 5.30 (Exercise 31.2, [1]). *Show that in three-step bidiagonalization, the optimal point at which to perform the QR factorization is when the matrix reaches an aspect ratio of 2.*

Problem 5.31 (Exercise 31.3, [1]). *Show that if the entries on both principal diagonals of a bidiagonal matrix are all nonzero, then the singular values of the matrix are distinct. (See Exercise 25.1, [1]).)*

Problem 5.32 (Exercise 31.4, [1]). *Let A be the $m \times n$ upper-triangular matrix with 0.1 on the main diagonal and 1 everywhere above the diagonal. Write a program to compute the smallest singular value of A in two ways: by calling a standard SVD software, and by forming A^*A and computing the square roots of its smallest eigenvalue. Run your program for $1 \leq m \leq 30$ and plot the results as two curves on a log scale. Do the results conform to our general discussion of these algorithms?*

Problem 5.33 (Exercise 31.5, [1]). *Let A be an $m \times n$ matrix whose entries are independent samples from $N(0,1)$, the normal distribution of mean 0, variance 1 (compare Exercise 12.3, [1]). Let B be a bidiagonal matrix*

$$B = \begin{bmatrix} x_m & y_{n-1} & & & \\ & x_{m-1} & y_{n-2} & & \\ & & \ddots & \ddots & \\ & & & x_{m-(n-2)} & y_1 \\ & & & & x_{m-(n-1)} \end{bmatrix} \tag{5.52}$$

where each x or y is the positive square root of an independent sample from the χ^2 distribution with degree equal to the attached subscript. (The χ^2 distribution of degree k is equal to the distribution of the sum of squares of k independent variables from $N(0,1)$.)

1. *Show that the distributions of the singular values of A and B are the same.*
2. *Verify this result by an experiment. Specifically, take $m = 100$ and $n = 50$, construct random matrices A and B as indicated, and plot the singular values of A against those of B.*

Chapter 6

Iterative Methods

6.1 Lecture 32: Overview of Iterative Methods

Problem 6.1 (Exercise 32.1, [1]). *An elliptic partial differential equation in three dimensions is discretized by a boundary element method. The result is a large dense linear system of equations in which each equation corresponds to a triangular surface element on a large sphere. To improve the accuracy, one must make the triangles smaller and thereby increase the number of equation, but the error shrinks only linearly in proportion to h, the diameter of the largest triangle.*

A value of h is chosen, the system is solved by Gaussian elimination, and a solution accurate to two digits is obtained in one minute of computer time. It is decided that three digits of accuracy are needed. Assuming storage is not a constraint, approximately how much time will be required for the new computation on the same computer?

Problem 6.2 (Exercise 32.2, [1]). *Consider the block matrix product*

$$\begin{bmatrix} W & X \\ Y & Z \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} E & F \\ G & H \end{bmatrix} \tag{6.1}$$

where, for simplicity, all the matrices $A, B, C, D, E, F, G, H, , W, X, Y, Z$ are assumed to be square and of the same dimension.

1. *Given A, B, C, D, E, F, G, H, how may (i) matrix additions and (ii) matrix multiplications does it take to compute W, X, Y, Z by the obvious algorithm?*

2. *Strassen showed that W, X, Y, Z can also be computed by the formulas*

$$P_1 = (A + D)(E + H) \tag{6.2}$$
$$P_2 = (C + D)E \tag{6.3}$$
$$P_3 = A(F - H) \tag{6.4}$$
$$P_4 = D(G - E) \tag{6.5}$$
$$P_5 = (A + B)H \tag{6.6}$$
$$P_6 = (C - A)(E + F) \tag{6.7}$$

$$P_7 = (B - D)(G + H) \tag{6.8}$$
$$W = P_1 + P_4 - P_5 + P_7 \tag{6.9}$$
$$X = P_3 + P_5 \tag{6.10}$$
$$Y = P_2 + P_4 \tag{6.11}$$
$$Z = P_1 + P_3 - P_2 + P_6 \tag{6.12}$$

How many (i) matrix additions or subtractions and (ii) matrix multiplications are involved now?

3. *Show that by applying Strassen's formulas recursively, one can obtain an algorithm for multiplying matrices of dimension $m = 2^k$ with an operation count $O\left(m^{\log_2 7}\right)$ as $m \to \infty$.*

4. *Write a recursive program that implements this idea, and give numerical evidence that your program works.*

6.2 Lecture 33: The Arnoldi Iteration

Problem 6.3 (Exercise 33.1, [1]). *Let $A \in \mathbb{C}^{m\times m}$ and $b \in \mathbb{C}^m$ be arbitrary. Show that any $x \in K_n$ is equal to $p(A)\,b$ for some polynomial p of degree $\leq n-1$.*

Problem 6.4 (Exercise 33.2, [1]). *Suppose Algorithm 33.1, [1], is executed for a particular A and b until at some step n, an entry $h_{n+1,n} = 0$ is encountered.*

1. *Show how*

$$AQ_n = Q_{n+1}\tilde{H}_n \tag{6.13}$$

can be simplified in this case. What does this simply about the structure of a full $m \times m$ Hessenberg reduction $A = QHQ^*$ of A?

2. *Show that $\mathcal{K}_n$ is an invariant subspace of A, i.e., $A\mathcal{K}_n \subseteq \mathcal{K}_n$.*

3. *Show that if the Krylov subspaces of A generated by b are defined by $K_k = \left\langle b, Ab, \ldots, A^{k-1}b\right\rangle$ as in*

$$\mathcal{K}_n = \left\langle b, Ab, \ldots, A^{n-1}b\right\rangle \tag{6.14}$$
$$= \langle q_1, q_2, \ldots, q_n\rangle \subseteq C^m \tag{6.15}$$

then

$$\mathcal{K}_n = \mathcal{K}_{n+1} = \mathcal{K}_{n+2} = \cdots \tag{6.16}$$

4. *Show that each eigenvalue of H_n is an eigenvalue of A.*

5. *Show that if A is nonsingular, then the solution x to the system of equations $Ax = b$ lies in $\mathcal{K}_n$.*

The appearance of an entry $h_{n+1,n} = 0$ is called a ***breakdown*** *of the Arnoldi iteration, but it is a breakdown of a benign sort. For applications in computing eigenvalues (Lecture 34, [1]) or solving systems of equations (Lecture 35, [1]), because of (4) and (5), a breakdown usually means that convergence has occurred and the iteration can be terminated. Alternatively, a new orthonormal vector q_{n+1} could be selected at random and the iteration then continued.*

Problem 6.5 (Exercise 33.3, [1]).

1. *Suppose Algorithm 33.1, [1], is executed for a particular A and b and runs to completion ($n = m$), with no breakdown of the kind described in the last exercise. Show that this implies that the minimal polynomial of A is of degree m.*

2. *Conversely, suppose that the minimal polynomial of A is of degree m. Show that this does not imply that for a particular choice of b, Algorithm 33.1, [1], will necessarily run to completion.*

3. *Explain why the result of (1) does not contradict Exercise 25.1(2), [1].*

6.3 Lecture 34: How Arnoldi Locates Eigenvalues

Problem 6.6 (Exercise 34.1, [1]). *Given $A \in \mathbb{C}^{m\times m}, b \in \mathbb{C}^m$ and $p \in P^*$, suppose we want to compute $p(A)\, b$. A natural place to start is with Horner's rule, which can be written*

$$p(z) = c_0 + z\,(c_1 + z\,(c_2 + \cdots + z\,(c_{m-1} + z)\cdots)) \tag{6.17}$$

1. *Write a **for** loop based on* (6.17) *(on paper, not on a computer) that computes $p(A)$ and then applies the result to b. Determine the number of flops required by this algorithm, to leading order.*

2. *Write another **for** loop for computing $p(A)\, b$, a far better one, in which b is introduced into the process at the beginning. Again determine the number of flops required to leading order.*

3. *In introductory numerical analysis texts, Horner's rule is recommended for evaluation of polynomials because it is faster than the obvious method of computing powers z^k, multiplying by coefficients c_k, and adding. Show that for computing $p(A)$ or $p(A)\, b$, by contrast, Horner's rule is not significantly faster than the obvious method.*

Problem 6.7 (Exercise 34.2, [1]). *We have seen that the Arnoldi polynomial p^n minimizes $\|p^n(A)\, n\|$. Another polynomial that might give clearer information about eigenvalues would be the* ***ideal Arnoldi polynomial****, also known as the* ***Chebyshev polynomial of*** *A, defined as the unique $p^* \in P^*$ that minimizes $\|p^n(A)\| s$. (This polynomial is not used in practice, because there is no fast way to compute it.)*

1. *Prove that p^* exists.*

2. *Prove that provided $p^*(A) \neq 0, p^*$ is unique. (Hint: Suppose p_1 and p_2 are distinct ideal Arnoldi polynomials for given A and n, set $p = \frac{p_1 + p_2}{2}$, and consider the singular vectors of $p(A)$ corresponding to the maximal singular value. This is a hard problem.*

Problem 6.8 (Exercise 34.3, [1]). *Let A be the $N \times N$ bidiagonal matrix with $a_{k,k+1} = a_{k,k} = \frac{1}{\sqrt{k}}, N = 64$. (In the limit $N \to \infty$, A becomes a non-self-adjoint compact operator.)*

1. *Produce a plot showing the spectrum $\Lambda(A)$ and the boundaries of the ϵ-pseudospectra $\Lambda_\epsilon(A)$ (Exercises 26.1 and 26.2, [1]) for*

$$\epsilon = 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4} \tag{6.18}$$

2. *Starting from a random initial vector, run the Arnoldi iteration and compute Ritz values at steps $n = 1, 2, \ldots, 30$. Produce plots to indicate the rates of convergence to eigenvalues of A, and comment on your results.*

3. *The Arnoldi iteration can also be used to approximate pseudospectra of A by those of H_n or $\tilde{H}_n$. (In the latter case, the boundary of $\Lambda_\epsilon\left(\tilde{H}_n\right)$ is defined by the condition $\sigma_{\min}\left(zI - \tilde{H}_n\right) = \epsilon$, or equivalently $\left\|\left(zI - \tilde{H}_n\right)^+\right\| = \frac{1}{\epsilon}$, where I is a rectangular version of the identity.) Experiment with this idea by plotting the ϵ-pseudospectra of $\tilde{H}_n$ for $n = 5, 10, 15, 20$. How closely do they match the corresponding pseudospectra of A?*

6.4 Lecture 35: GMRES

Problem 6.9 (Exercise 35.1, [1]). *Show that if $S \subseteq \mathbb{C}$ contains infinitely many points, then*

$$\|p\|_S = \sup_{z \in S} |p(z)| \tag{6.19}$$

defines a norm on the vector space of all polynomials with complex coefficients. Explain what goes wrong if S has only finitely many points.

Problem 6.10 (Exercise 35.2, [1]).

1. *Let $S \subseteq \mathbb{C}$ be a set whose convex hull contains 0 in its interior. That is, S is contained in no half-plane disjoint from the origin. Show that there is no $p \in P_1$ (i.e., no polynomial p of degree 1 with $p(0) = 1$) such that $\|p\|_S < 1$.*

2. *Let A be a matrix, not necessarily normal, whose spectrum $\Lambda(A)$ has the property (1). Show that there is no $p \in P_1$ such that $\|p(A)\| < 1$.*

3. *Though the convergence in Figure 35.5 is slow, it is clear that $\|r_1\| < \|r_0\|$. Explain why this does not contradict the result of (2). Describe what kind of polynomial $p_1 \in P_1$ GMRES has probably found to achieve $\|r_1\| < \|r_0\|$.*

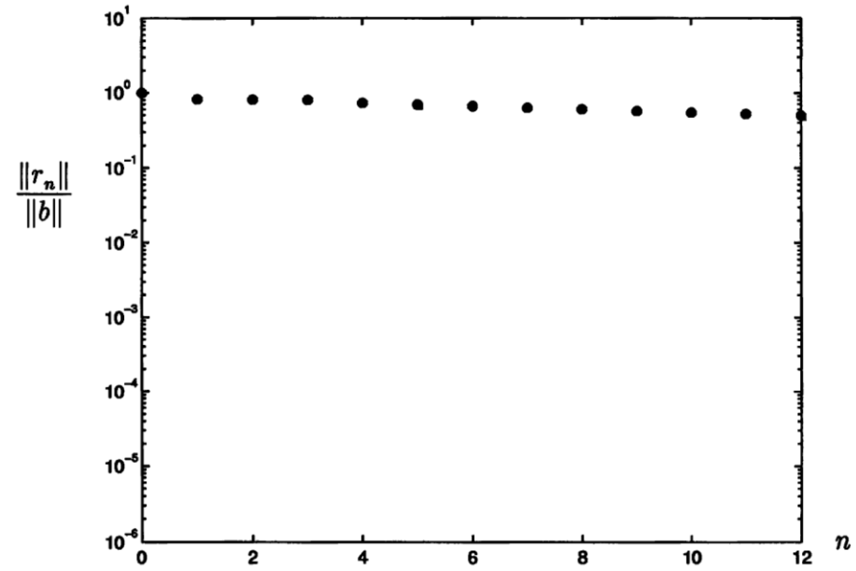


Figure 6.1: *GMRES convergence curve for the matrix of of Figure 35.4. The convergence has slowed down greatly. When an iterative method stagnates like this, it is time to look for a better preconditioner.*

Problem 6.11 (Exercise 35.3, [1]). *The recurrence*

$$x_{n+1} = x_n + \alpha r_n \tag{6.20}$$
$$= x_n + \alpha (b - Ax_n) \tag{6.21}$$

where α is a scalar constant, is known as a ***Richardson iteration****.*

1. *What polynomial $p(A)$ at step n does this correspond to?*
2. *What choice of α would you recommend for the matrix A of Figure 35.2*

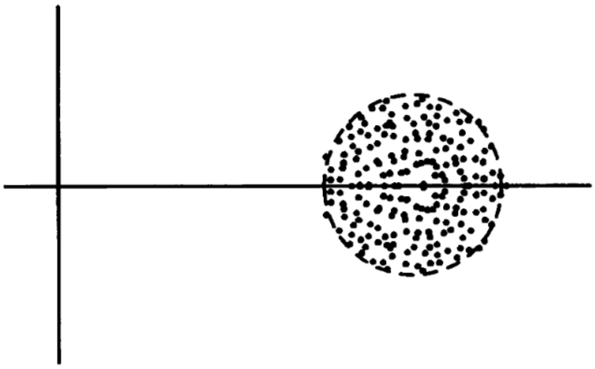


Figure 6.2: *Eigenvalues of the* 200×200 *matrix* A *of* `m = 200; A = 2*eye(m) + 0.5*randn(m)/sqrt(m)`. *The dashed curve is the circle of radius* $\frac{1}{2}$ *with center* $z = 2$ *in* $\mathbb{C}$. *The eigenvalues are approximately uniformly distributed within this disk.*

and what would you expect to be the corresponding convergence rate?

3. *Same question for the matrix of Figure 35.4.*

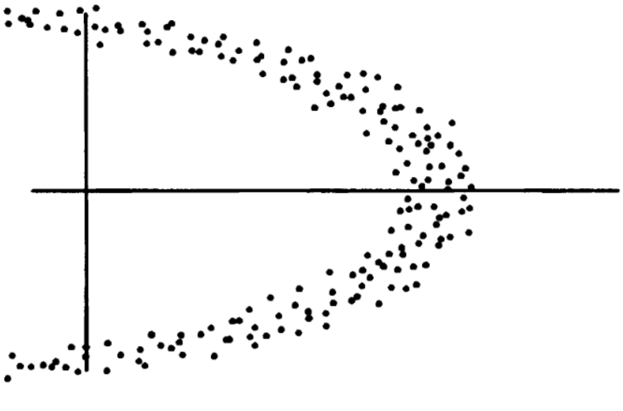


Figure 6.3: *Eigenvalues of the* 200×200 *matrix, like that of* `m = 200; A = 2*eye(m) + 0.5*randn(m)/sqrt(m)` *except with a modified diagonal. Now the eigenvalues surround the origin on one side.*

Problem 6.12 (Exercise 35.4, [1]).

1. *Describe an* $O\left(n^2\right)$ *algorithm based on QR factorization by Givens rotations (Exercise 10.4, [1]) for solving the least squares problem of Algorithm 35.1, [1].*

2. *Show how the operation count can be improved to* $O\left(n\right)$, *as mentioned on [1] - p.268, if the problem for step* $n-1$ *has already been solved.*

Problem 6.13 (Exercise 35.5, [1]). *Our statement of the GMRES algorithm (Algorithm 35.1, [1]) begins with the initial guess $x_0 = 0, r_0 = b$. (The same applies to CG and BCG, Algorithm 38.1 and 39.1, [1].) Show that if one wishes to start with an arbitrary initial guess x_0, this can be accomplished by an easy modification of the right-hand side b.*

Problem 6.14 (Exercise 35.6, [1]). *For larger values of n, the cost of GMRES in operations and storage may be prohibitive. In such circumstances a method called k-**step restarted GMRES or GMRES(k)** is often employed, in which, after k steps, the GMRES iteration is started anew with the current vector x_k as an intial guess.*

1. *Compare the asymptotic operation counts and storage requirements of GMRES and GMRES(k), for fixed k and increasing n.*

2. *Describe an example in which GMRES(k) can be expected to converge in nearly as few iterations as GMRES (hence much faster in operation count).*

3. *Describe another example in which GMRES(k) can be expected to fail to converge, whereas GMRES succeeds.*

6.5 Lecture 36: The Lanczos Iteration

Problem 6.15 (Exercise 36.1, [1]). *In Lecture 27, [1], it was pointed out that the eigenvalues of a symmetric matrix $A \in \mathbb{R}^{m\times m}$ are the stationary values of the Rayleigh quotient*

$$r\left(x\right) = \frac{x^T A x}{x^T x}, \forall x \in \mathbb{R}^m \tag{6.22}$$

Show that the Ritz values at step n of the Lanczos iteration are the stationary values of $r\left(x\right)$ if x is restricted to $\mathcal{K}_n$.

Problem 6.16 (Exercise 36.2, [1]). *Consider a polynomial $p \in P^n$, i.e., $p\left(z\right) = \prod_{k=1}^{n} \left(z - z_k\right)$ for some $z_k \in \mathbb{C}$.*

1. *Write $\log\left|p\left(z\right)\right|$ as a sum of n terms corresponding to the points z_k.*

2. *Explain why the term involving z_k can be interpreted as the potential corresponding to a negative unit point charge located at z_k, if charges repel in inverse proportion to their separation. Thus $\log\left|p\left(z\right)\right|$ can be viewed as the potential at z induced by n point charges.*

3. *Replacing each charge -1 by $-\frac{1}{n}$ and taking the limit $n \to \infty$, we get a continuous charge density distribution $\mu\left(\zeta\right)$ with integral -1, which we can expect to be related to the limiting density of zeros of polynomials $p \in P^n$ as as $n \to \infty$. Write an integral representing the potential $\varphi\left(z\right)$ corresponding to $\mu\left(\zeta\right)$, and explain its connection to $\left|p\left(z\right)\right|$.*

4. *Let S be a closed, bounded subset of $\mathbb{C}$ with no isolated points. Suppose we seek a distribution $\mu(z)$ with support in S that minimizes* $\max_{z\in S}\varphi(z)$. *Give an argument (not rigorous) for why such a $\mu(z)$ should satisfy $\varphi(z) =$ constant throughout S. Explain why this means that the "charges" are in equilibrium, experiencing no net forces. In other words, S is like a 2D electrical conductor on which a quantity -1 of charge has distributed itself freely. Except for an additive constant, $\varphi(z)$ is the* ***Green's function*** *for S.*

Problem 6.17 (Exercise 36.3, [1]). *Let A be the* 1000×1000 *symmetric matrix whose entries are all zero except for $a_{ij} = \sqrt{i}$ on the diagonal, $a_{ij} = 1$ on the sub- and superdiagonals, and $a_{ij} = 1$ on the 100the sub- and superdiagonals, i.e., for $|i - j| = 100$. Determine the smallest eigenvalue of A to six digits of accuracy by the Lanczos iteration.*

Problem 6.18 (Exercise 36.4, [1]). *As a special case of the Arnoldi lemniscates of Lecture 34, [1], "Lanczos lemniscates" can be employed to illustrate the convergence of the Lanczos iteration. Find a way to modify your program of Exercise 36.3, [1], to plot the Lanczos lemniscates at each step. Your method need not be elegant, efficient, or numerically robust. Produce plots of Lanczos lemniscates at steps $n = 1, 2, \ldots, 12$ for the example of Figure 36.2 and for an example of your own choosing.*

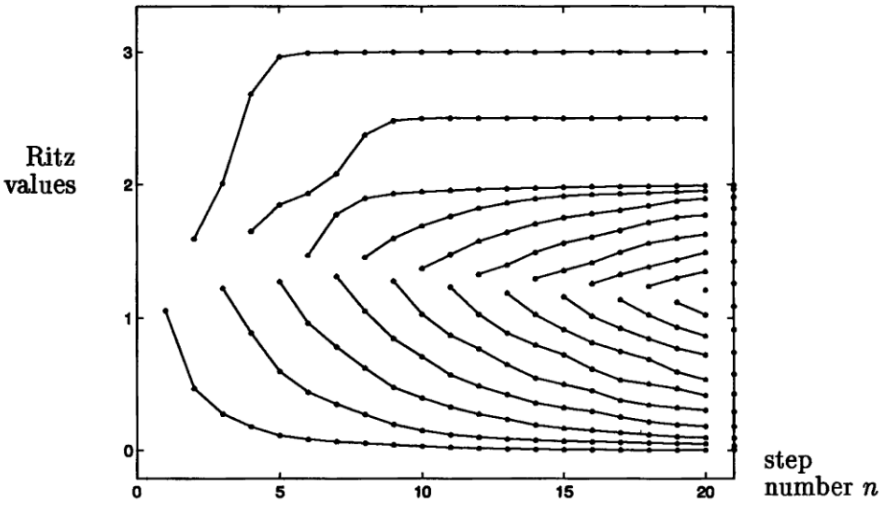


Figure 6.4: *Ritz values for the first 20 steps of the Lanczos iteration applied to the same matrix. The convergence to the eigenvalues* 2.5 *and* 3.0 *is geometric. Little useful convergence to individual eigenvalues occurs in the* $[0, 2]$ *part of the spectrum. Instead, the Ritz values in* $[0, 2]$ *approximate Chebyshev points in that interval, marked by dots on the right-hand boundary.*

6.6 Lecture 37: From Lanczos to Gauss Quadrature

Problem 6.19 (Exercise 37.1, [1]). *The standard recurrence relation for Legendre polynomials is*

$$P_n(x) = \frac{2n-1}{n} x P_{n-1}(x) - \frac{n-1}{n} P_{n-2}(x) \tag{6.23}$$

with initial values $P_0(x) = 1, P_1(x) = x$.

1. *Confirm that* (6.23) *gives the polynomials* $P_2(x)$ *and* $P_3(x)$ *of*

$$P_0(x) = 1 \tag{6.24}$$

$$P_1(x) = x \tag{6.25}$$

$$P_2(x) = \frac{3}{2}x^2 - \frac{1}{2} \tag{6.26}$$

$$P_3(x) = \frac{5}{2}x^3 - \frac{3}{2}x \tag{6.27}$$

2. *Since* $\{P_n(x)\}$ *and* $\{q_{n+1}(x)\}$ *are normalized differently,* (6.23) *is not the same as the recurrence*

$$x q_n(x) = \beta_{n-1} q_{n-1}(x) + \alpha_n q_n(x) + \beta_n q_{n+1}(x) \tag{6.28}$$

with the coefficients

$$\beta_n = \frac{1}{2}\left(1 - \frac{1}{4n^2}\right)^{-\frac{1}{2}} \tag{6.29}$$

Write down the two tridiagonal matrices corresponding to these formulas, and derive the relationship between them.

3. *Use the result of (2) to determine a formula for* $q_{n+1}(1)$*, or equivalently, for* $\|P_n\|$.

Problem 6.20 (Exercise 37.2, [1]). *Show based on the definition of orthogonality that* $q_{n+1}(x)$ *has* n *distinct zeros, all contained in the open interval* $(-1, 1)$*. (The fact that they are distinct also follows from Exercise 25.1, [1], but here, use a direct argument.*

Problem 6.21 (Exercise 37.3, [1]). *The problem of interpolating* n *data values* $\{y_j\}$ *in* n *distinct data points* $\{x_j\}$ *by a polynomial of degree* $\leq n-1$ *was expressed in*

$$\begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{m-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{m-1} \\ 1 & x_3 & x_3^2 & \cdots & x_3^{m-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_m & x_m^2 & \cdots & x_m^{m-1} \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ \vdots \\ c_m \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} \tag{6.30}$$

as a square Vandermonde linear system of equations.

1. *Prove that this Vandermonde matrix is nonsingular by arguing that if the interpolation problem has a solution, it must be unique.*

2. *Write down the analogous system of equations implicit in Theorem 37.2, [1]. Using the result of (1), prove this theorem.*

Problem 6.22 (Exercise 37.4, [1]).

1. *Write a six-line* MATLAB *program that computes the nodes and weights for the n-point Gauss-Legendre quadrature formula and applies these numbers to compute the approximate integral of the function f.*

2. *Taking $f(x) = e^x$ and $n = 4$, confirm the example in the text. Then plot $|I(e^x) - I_n(e^x)|$ on a log scale for $n = 1, 2, \ldots, 40$ and comment on the results.*

3. *Produce a similar plot for $f(x) = e^{|x|}$, and comment.*

Problem 6.23 (Exercise 37.5, [1]). *The program of Exercise 37.4, [1], computes the zeros of Legendre polynomials, also known as **Legendre points** in $[-1, 1]$. The zeros of Chebyshev polynomials, **Chebyshev points**, are given by the explicit formula*

$$x_j = \cos\theta_j \tag{6.31}$$

$$\theta_j = \frac{\left(j - \frac{1}{2}\right)\pi}{m}, 1 \leq j \leq m \tag{6.32}$$

Perform a sequence of calculations to generate numbers and plots illustrating as elegantly as you can that in the limit $n \to \infty$, both Legendre and Chebyshev points approach the limiting density distribution

$$\mu(x) = \frac{1}{\pi}\left(1 - x^2\right)^{-\frac{1}{2}} \tag{6.33}$$

(in the notation of Exercise 36.2, [1]). Produce further plots and numbers to explore the question: how close are Legendre points to Chebyshev points for various values of n?

6.7 Lecture 38: Conjugate Gradients

Problem 6.24 (Exercise 38.1, [1]). *Based on the condition numbers κ reported in the text, determine the rate of convergence predicted by Theorem 38.5, [1], for the matrices A of Figure 38.1*

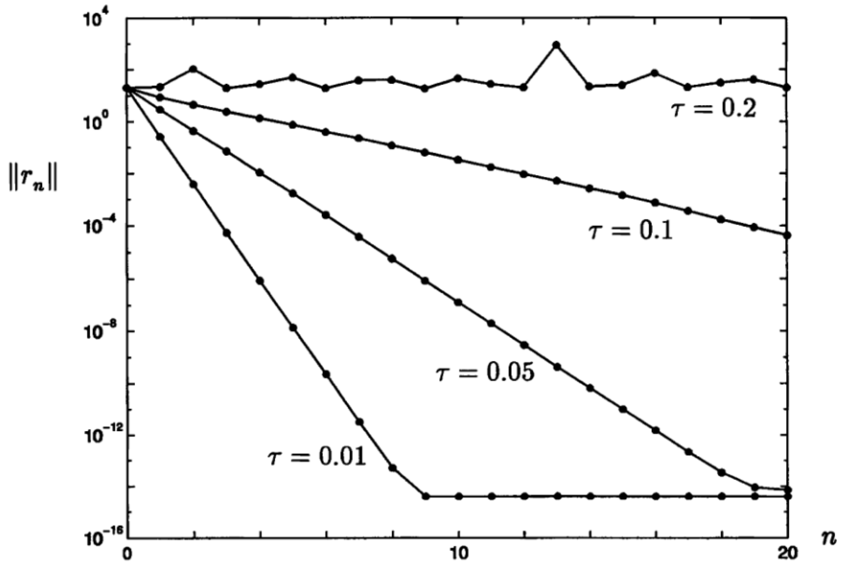


Figure 6.5: *CG convergence curves for the* 500×500 *sparse matrices* A *described in the text. For* $\tau = 0.01$*, the system is solved about 700 times faster by CG than by Cholesky factorization. For* $\tau = 0.2$*, the matrix is not positive definite and there is no convergence.*

with $\tau = 0.01, 0.05, 0.1$*. Draw lines on a copy of Figure 38.1 indicating how closely these predictions match the actual convergence rates.*

Problem 6.25 (Exercise 38.2, [1]). *Suppose* A *is a real symmetric* 805×805 *matrix with eigenvalues* $1.00, 1.01, 1.02, \ldots, 8.98, 8.99, 9.00$ *and also* $10, 12, 16, 24$*. How many steps of the conjugate gradient iteration must you take to be sure of reducing the initial error* $\|e_0\|_A$ *by a factor of* 10^6*?*

Problem 6.26 (Exercise 38.3, [1]). *The conjugate gradient is applied to a symmetric positive definite matrix* A *with the result* $\|e_0\|_A = 1, \|e_{10}\|_A = 2 \times 2^{-10}$*. Based solely on this data,*

1. *What bound can you give on* $\kappa(A)$*?*
2. *What bound can you give on* $\|e_{20}\|_A$*?*

Problem 6.27 (Exercise 38.4, [1]). *Suppose* A *is a dense symmetric positive definite* 1000×1000 *matrix with* $\kappa(A) = 100$*. Estimate roughly how many flops are required to solve* $Ax = b$ *to ten-digit accuracy by*

1. *Cholesky factorization.*
2. *Richardson iteration with the optimal parameter* α *(Exercise 35.3, [1]).*
3. *CG.*

Problem 6.28 (Exercise 38.5, [1]). *We have described CG as an iterative minimization of the function $\varphi(x)$ of*

$$\varphi(x) = \frac{1}{2}x^T Ax - x^T b \tag{6.34}$$

Another way to minimize the same function - far slower, in general - is by the method of ***steepest descent.***

1. *Derive formula $\nabla\varphi(x) = -r$ for the gradient of $\varphi(x)$. Thus the steepest descent iteration corresponds to the choice $p_n = r_n$ instead of $p_n = r_n + \beta_n p_{n-1}$ in Algorithm 38.1, [1].*
2. *Determine the formula for the optimal step length α_n of the steepest descent iteration.*
3. *Write down the full steepest descent iteration. There are three operations inside the main loop.*

Problem 6.29 (Exercise 38.6, [1]). *Let A be the 100×100 tridiagonal symmetric matrix with $1, 2, \ldots, 100$ on the diagonal and 1 on the sub- and superdiagonals, and set $b = (1, 1, \ldots, 1)^T$. Write a program that takes 100 steps of the CG and also the steepest descent iteration to approximately solve $Ax = b$. Produce a plot with four curves on it: the computed residual norms $\|r_n\|_2$ for steepest descent, and the estimate $2\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^n$ of Theorem 38.5, [1]. Comment on your results.*

6.8 Lecture 39: Biorthogonalization Methods

Problem 6.30 (Exercise 39.1, [1]). *Consider a problem $Ax = b$ for the matrix*

$$A = \begin{bmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & 0 & 1 & \\ & & & 0 & 1 \\ 1 & & & & 0 \end{bmatrix} \tag{6.35}$$

of dimension m.

1. *Show that the singular values are all 1 and that this implies that CGN converges in one step.*
2. *Show that the eigenvalues are the mth roots of unity and that this implies that GMRES requires m steps to converge for general b.*
3. *This matrix A has so much structure that one does not need to consider eigenvalues or singular values to understand its convergence behavior. In particular, explain by an elementary argument why GMRES takes m steps to converge for the right-hand side $b = (1, 0, 0, \ldots, 0)^T$.*

Problem 6.31 (Exercise 39.2, [1]). *As a converse to Exercise 39.1, [1], devise an example of a matrix of arbitrary dimension m with almost the opposite property: GMRES converges in two steps, but CGN requires m steps.*

Problem 6.32 (Exercise 39.3, [1]).

1. *If A is hermitian and s_0 is chosen appropriately, Algorithm 39.1 reduces to Algorithm 38.1, [1]. Confirm this statement and determine the appropriate s_0.*

2. *Suppose A is a complex matrix that is symmetric but not hermitian. Show that with a different choice of s_0, Algorithm 39.1, [1], again reduces to an iteration involving just one three-term recurrence.*

Problem 6.33 (Exercise 39.4, [1]). *Figure 39.2 illustrated that if the convergence curve for a biorthogonalization method has spikes in it, this may effect the attainable accuracy in floating point arithmetic. Without trying to be rigorous, explain why this is so, and comment on the analogy with growth factors in Gaussian elimination (Lecture 22, [1]).*

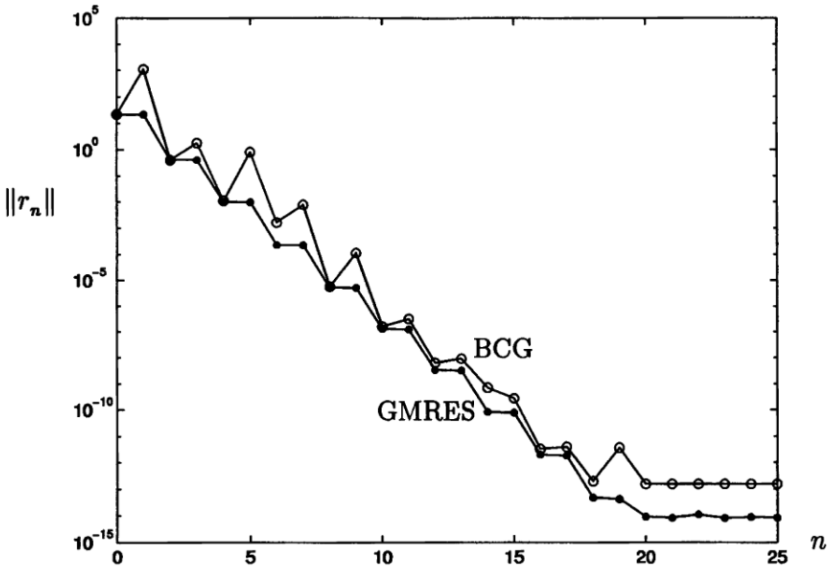


Figure 6.6: *Comparison of GMRES and BCG for the* 500×500 *matrix labeled* $\tau = 0.01$ *in Figure 38.1, but with the signs of the entries randomized.*

Problem 6.34 (Exercise 39.5, [1]). *Which of CG, GMRES, CGN, or BCG would you expect to be most effective for the following $m \times m$ problems $Ax = b$, and why?*

1. *A dense nonhermitian matrix with $m = 10^4$, all but three of whose eigenvalues are approximately equal to -1.*

2. *The same, but with all but three of the eigenvalues scattered about the region $-10 \leq Real(\lambda) \leq 10, -1 \leq Imag(\lambda) \leq 1$.*

3. *A sparse nonhermitian matrix with* $m = 10^6$ *but only* 10^7 *nonzero entries, with eigenvalues as in (1).*

4. *A sparse hermitian matrix with* $m = 10^5$ *whose eigenvalues are scattered through the interval* $[1, 100]$.

5. *The same, except for outlying eigenvalues at* 0.01 *and* $10,000$.

6. *The same, but with additional outliers at* $-1, -10,$ *and* -100.

7. *A sparse, normal matrix with* $m = 10^5$ *whose eigenvalues are complex numbers scattered about the annulus* $1 \leq |\lambda| \leq 2$.

6.9 Lecture 40: Preconditioning

Problem 6.35 (Exercise 40.1, [1]). *Suppose* $A = M - N$, *where* M *is nonsingular. Suppose* $\left\|I - M^{-1}N\right\|_2 \leq \frac{1}{2}$, *and* M *is used as a preconditioner as in*

$$M^{-1}Ax = M^{-1}b \tag{6.36}$$

1. *Show that if GMRES is applied to this preconditioned problem, then the residual norm is guaranteed to be six orders of magnitude smaller, or better, after twenty steps.*

2. *How many steps of CGN are needed for the same guarantee?*

Problem 6.36 (Exercise 40.2, [1]). *Show that if a matrix* A *and a preconditioner* M *are hermitian positive definite, then the same CG convergence rate is obtained whether* M *is used as a left preconditioner or a right preconditioner. Explain why this result does not hold for nonhermitian matrices and iterations such as GMRES, CGN, or BCG.*

The End

Bibliography

[1] Lloyd N. Trefethen, David Bau III, *Numerical Linear Algebra*, SIAM Society for Industrial and Applied Mathematics, 1997.

[2] `https://en.wikipedia.org/wiki/Gershgorin_circle_theorem`